Genetic analysis of macular telangiectasia

Nancy Lynn Parmalee

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

©2012

Nancy Parmalee

All rights reserved

Abstract

GENETIC ANALYSIS OF MACULAR TELANGIECTASIA

NANCY LYNN PARMALEE

Macular telangiectasia type 2, or MacTel, is an adult onset retinal disease that causes progressive loss of central vision, usually beginning between the 5th and 7th decades of life. Macular degenerative diseases comprise a large portion of the blinding diseases affecting people over the age of 50. The collective burden of vision loss and low vision includes the impact on patients in loss of independence and decline of self-reported quality of life, as well as significant resource allocation in medical and supportive care. There are no cures for macular diseases; currently available treatments may slow the progression in some cases, but patients remain at risk of losing their eyesight and the independence that vision affords. While macular telangiectasia is not as prevalent as other macular degenerative diseases, such as age-related macular degeneration, elucidating the causes of MacTel may reveal mechanisms that are common to other macular diseases, providing clues to guide research into possible treatment targets.

In this work, I have assembled the first cohort of MacTel patients and their family members in collaboration with the MacTel Project, an international consortium of researchers and clinicians dedicated to studying this disease. I have identified and analyzed families with multiple affected relatives in which MacTel appears to be an inherited disease. The goal of this work has been to

identify a genetic cause for the disease in these families. To that end, I used genotyping data from approximately 900,000 single nucleotide polymorphism per individual to create sets of genetic markers for linkage analysis. I analyzed these markers using two-point and multipoint linkage algorithms to search for a region of the genome that was inherited in conjunction with the phenotype more often than would be expected by chance alone, with the goal of finding a causative genetic variant. This analysis resulted in the identification of a region of chromosome 1 spanning approximately 15 million bases that is significantly linked to the MacTel phenotype.

I analyzed recombination breakpoints in this region, as well as across the genome, to map chromosomal segments inherited identical by descent in siblings that express the MacTel phenotype. This approach provided a profile for each set of siblings of regions of the genome incompatible with inherited genetic disease by virtue of the fact that no alleles were shared identical by descent. I then compared these regions across families under the hypothesis that if MacTel is caused by the same gene in all families, the chromosomal region harboring a causative variant would fail to be excluded in any family. This analysis excluded much of the genome and refined the boundaries of the linkage interval on chromosome 1, but did not appreciably narrow the region of interest.

In parallel with analysis of linkage in families, I sequenced 40 genes of interest as possible functional candidates, using Sanger sequencing in the probands of families with multiple affected relatives. I selected candidate genes based on gene function and expression, choosing

genes with a plausible connection to the phenotype based on information in the literature. Thirteen of these genes were selected based on their location in the region of the maximum linkage score. I was unable to identify any variant that was consistent with being a causative variant for inherited MacTel.

I performed a broader search using exome sequencing to investigate all protein coding regions in the genome in four affected individuals—two affected sisters, and two unrelated family probands. The goal of this effort was to identify a gene with rare or unknown missense variants in all four of the affected individuals. The expectation was that a causative variant would be inherited identical by descent in the affected sisters, and either the same variant, or a different variant in the same gene would be present in the unrelated probands. At the time this experiment was performed, exome sequencing had only just become broadly available to researchers, and, as I learned, was not free of technical limitations. The data failed to provide adequate coverage of the exome to evaluate the intended hypothesis, but did provide useful lessons in the promise and the hazards of high throughput sequencing, which I will discuss in detail.

In summary, this work represents the first in depth genetic analysis of a cohort of MacTel families with inherited disease. I have identified a region of chromosome 1 linked to the disease, providing the first evidence for a MacTel susceptibility locus. This work is ongoing, in the hopes of identifying a causative gene for MacTel that will shed light on the underlying

mechanism of the disease, and potentially guide the effort to develop effective treatments for the patients who suffer from this disease.

Table of Contents

Table of Contents i
List of Figures iv
List of Tables v
Acknowledgements vi
Dedication ix

Chapter One – Introduction 1

Macular telangiectasia type 2 and the MacTel Project 1
The architecture of the retina 4
The MacTel phenotype 7
The hypothesis of MacTel as an inherited disease 12
Summary 12

Chapter Two - Description of Sample and Introduction to Genetic Methods 14

Recruitment of MacTel patients and relatives 14
Genetic Approaches 20
Candidate gene screening 21
Linkage Analysis 22
Association 25
Exome Sequencing 28
Analysis of risk haplotypes 29

Introduction to Chapter Three 31

Chapter Three – Analysis of candidate genes for macular telangiectasia type 2 32

Abstract 32
Introduction 32
Methods 33
Study population 33
Diagnosis of MacTel 33
Sequencing and genotyping 33
Results 35
Genes involved in angiogenesis 35

Genes involved in macular pigment transport 35
Genes identified from expression arrays 35
Genes identified from suggestive linkage regions 35
Genes involved in diseases with related phenotypes 35
Discussion 37
References 37
Summary of Relevance 45

INTRODUCTION TO CHAPTER FOUR 46

CHAPTER FOUR – MACULAR TELANGIECTASIA TYPE 2: LINKAGE TO 1Q41-42 AND ANALYSIS OF THE LOCUS 47

Introduction 48
Results 51
Study population 51
Multipoint linkage analysis 52
Recomination mapping by determination of identical by descent sharing 53
Comparison of identical by descent to linkage exclusion 53
Gene sequencing 54
Discussion 56
Materials and Methods 59
Study cohort 59
Genotyping and marker selection 60
Linkage analysis 60
Candidate gene sequencing 62
Summary of Relevance 69

CHAPTER FIVE – WHOLE EXOME SEQUENCING, ADDITION OF FAMILIES TO THE LINKAGE COHORT, AND HAPLOTYPE ANALYSIS 70

Whole exome analysis of four MacTel patients 71
Introduction 71
Methods 74
Results 75
Discussion 79
Addition of families to the linkage cohort and indications of genetic heterogeneity 82
Introduction 82
Methods 83

RESULTS 85
DISCUSSION 92
HAPLOTYPING OF AFFECTED SIBLINGS 97
INTRODUCTION 97
METHODS 98
RESULTS 100
DISCUSSION 102

CHAPTER SIX – SUMMARY AND CONCLUSIONS 106

SUMMARY OF RESULTS 106
FUTURE DIRECTIONS 111

APPENDIX ONE – A MATHEMATICAL EXPLORATION OF MODEL PARAMETERS 114

REFERENCES 124

BIBLIOGRAPHY 134

RESEARCH ARTICLES 134
REVIEW ARTICLES 135
ABSTRACTS AND PRESENTATIONS 135

List of Figures

Figure 1. Anatomy of the eye 5

Figure 2. Vertical section of the human retina 6

Figure 3. The Amsler grid 8

Figure 4. Retinal images from patients evaluated for MacTel 17

Figure 5. Segregation of the p.Pro33Ser/p.Pro168Ser Frizzled-4 variant 31

Figure 6. Pedigrees of four families with multiple relatives affected with MacTel 60

Figure 7. Genome-wide linkage scan in 23 MacTel families 61

Figure 8. LOD, HLOD, and NPL scores for choromosome 1 62

Figure 9. Identical by descent allele sharing on chromosome 1 in 8 families 63

Figure 10. Base coverage in whole exome sequencing 73

Figure 11. Genome-wide LOD and HLOD scores after adding families 85

Figure 12. Allele sharing identical by descent in 7 additional families 87

Figure 13. Family contributions to LOD scores at chromosome 1q41-42 88

Figure 14. Haplotypes in selected families on chromosome 1 97

Figure 15. Comparison of genotypes in two families on chromosome 1 98

Figure 16. Population linkage disequilibrium at 1q41-42 100

List of Tables

Table 1a. Summary of the MacTel cohort by enrollment center 15

Table 1b. Summary of the number of relatives examined per family in the MacTel cohort 16

Table 2. Imaging techiniques used in the diagnosis of MacTel and typical findings 16

Table 3. Missense variants detected by Sanger sequencing in 8 MacTel probands 33

Table 4. Primer sequences for candidate gene screening 37

Table 5. Exclusion based on LOD scores less than -2 under autosomal dominance 64

Table 6. Variants detected by Sanger sequencing in two affected individuals 65

Table 7. Summary of whole exome variants reported 72

Table 8. Comparison of variants detected by exome sequencing, and genotyping 74

Table 9. Comparison of variants detected by exome and Sanger sequencing 74

Table 10. Unknown missense variants detected by exome sequencing on chromosome 1 75

Acknowledgements

There are many people who have influenced me, supported me, and shared this journey with me. I appreciate each of you. My children, Justin and Kelsey Soto, have endured the most, sacrificed the most, and have been my strongest supporters at every turn. There are no words to express what this means to me. I am immensely proud of the young adults they have become. I would like to particularly thank them both for humoring me on one particular Thanksgiving when I was asked what I was thankful for, and I responded "data." I am and will always be most thankful for the privilege of being their mother.

The credit for all that I am, and all that I do ultimately rests with my parents, Margaret Parmalee, and John Parmalee, who raised me to be curious, compassionate, and to never stop working towards the things I believe are important. Thank you, Mom and Dad.

From the bottom of my heart, my love and gratitude to my sister, Linda Parmalee, for all that she taught me.

My family is large enough for an extremely well-powered linkage study, and it would take many pages to list them all, and the many ways they have supported and nurtured me. I want to extend special thanks and much love to Jeanne, David, and Bruce, and their families, with whom I share no genetic material, but who have consistently proven that that is only one

definition of family. George, Franklin, Cristina—each of you have been with me at times on this road that I would not have been able imagine what I would have done without you. Many dear friends have encouraged me, laughed with me, and shared the best and worst of times, and everything in between. Robb, Danny, Amy, Casey, Robert, Greg, Veronika, Alissa, Jacqueline, Wendy—and many more—thank you only begins to cover it.

Many talented and dedicated scientists have taught me, encouraged me, and inspired me. I would like to particularly thank Michael O'Connor and Kavita Arora for providing my first and second lab positions, first as an undergraduate, and then as a technician. These were the lab experiences, and the lab groups that inspired me to want to go further.

Above all, I would like to thank my advisor, Rando Allikmets. Without his dedication and support, both to this project, and to me as a student, it would not have been possible for me to complete this work. He has challenged me, taught me, and pushed me to be a better scientist. I would not have been able to complete this degree if he had not decided to take a chance on me.

I owe an eternal debt of gratitude to my thesis advisory committee members, Dorothy Warburton and Joe Terwilliger for support, guidance, encouragement, and for sharing their insights and wisdom as this work has progressed. I am grateful to Susan Hodge and Ali Gharavi

for serving on my thesis committee, and for insightful and helpful discussion. I am equally indebted to Ginny Papaioannou and the Genetics Department training committee for believing in me and standing by me through the many stages of this journey. Fred Loweff, Stacy Warren, and Nina Steinberg each helped in more ways than I can count. Many thanks as well for assistance and friendship to my labmates, Jian Kong, Joanna Merriam, Jana Zernant, Carl Schubert and all the members of the Allikmets lab, past and present.

I have been privileged to work with and learn from many collaborators, here in the Columbia community, across the country, and around the world. Special thanks to Krzysztof Kiryluk, Lorraine Clark, Miguel Verbitsky, David Greenberg, Bill Stewart, and Joe Lee and the members of the Genome Group, and many others for all manner of assistance and insight. Special thanks to Carol Mason for mentorship, and for administering the Vision Science training grant, which supported me through much of this work.

Finally, I would to thank The MacTel Project, and the many collaborators whose joint efforts have informed this work, and are represented in this work. Deepest gratitude to the MacTel patients and families for their willingness to participate in this study. We share the hope that these efforts will lead to an understanding of this disease.

Dedication

To my children, Justin Soto and Kelsey Soto

And in memory of my sister

Linda Parmalee

1973-1989

Chapter One – Introduction

Macular telangiectasia type 2 and the MacTel Project

It is estimated that in 2002, more than 161 million people in the world suffered from visual impairment. At least 37 million people were blind, with the greatest proportion of cases of blindness occurring in people over age 50 (Resnikoff et al., 2004). Age-related macular degeneration (AMD) is the third leading cause of blindness in the world, and the primary cause of blindness in developed countries; the primary risk factor for blindness or low vision is considered to be aging. The World Health Organization estimates that more than 82% of all those who are blind are over age 50 (http://www.vision2020.org/main.cfm?type=FACTS). As global demographics shift, an increasingly greater segment of the population is at risk for losing functional vision. Research efforts to define the pathology of AMD and other macular degenerative disorders have intensified accordingly, and much progress has been made. However, much work remains to be done. The work described in this thesis investigates a comparatively rare macular degenerative disease, macular telangiectasia type 2, or MacTel. Patients usually present with complaints of visual distortion between the ages of 50 and 70. Frequently, the onset is slow and progressive, with some patients reporting that they first noticed changes in their vision many years before seeking treatment. While the burden of MacTel is far lower than that of more prevalent diseases, the impact on those who are affected is significant. Although MacTel is not considered a blinding disorder, the disease disrupts the delicate architecture of the macula and the fovea—the regions of the retina that provide the

high-resolution central vision needed for tasks such as reading and driving—causing considerable disruption to vision. As with many less prevalent diseases, MacTel had not been the focus of intense research efforts, given that it was thought to be very rare. One can imagine that patients who received a diagnosis of MacTel were confronted not only with the progressive loss of visual acuity that is the hallmark of the disease, but also with a lack of treatment options, and uncertainty about how the disease would progress and whether their children were likely to be at increased risk.

From the clinician's standpoint, retinal specialists who diagnose and treat MacTel patients have persistently expressed the view that MacTel has been underdiagnosed. The early signs of the disease are subtle changes in the distribution of pigment in the retina that are only detectable with specialized imaging, thus, many cases may be missed when this imaging is unavailable or is not routinely performed. In the late stages of the disease, the architecture of the retina can be profoundly disrupted, such that the key diagnostic signs can no longer be distinguished. These cases may be misdiagnosed as AMD, which also profoundly disrupts the organization of the retina.

The MacTel Project (established in 2005, with funding from the Lowy Medical Research Foundation) is an international consortium of clinicians and basic science researchers working toward understanding the pathogenesis and progression of MacTel, with the goal of eventually

translating basic science findings into treatment targets. The work contained in this thesis has been performed under the auspices of the MacTel Project.

Many insights into the molecular mechanisms of disease pathologies have been achieved by the discovery of genetic variants in rare diseases. The hope is that insights into the mechanism of pathology in MacTel could be extended to other diseases of the retina. The most prominent features of the MacTel phenotype are aberrant macular pigment distribution, and changes in the retinal vasculature. Macular pigments protect the retina from damaging short wavelength light, however the molecular pathways responsible for transporting and depositing pigments in the retina are still very poorly understood. Macular pigment is thought to play an important role in AMD pathogenicity (Loane et al., 2008a; Loane et al., 2008b), thus, learning more about the mechanisms of macular pigment transport through the study of the etiology of MacTel could yield important information to inform research efforts in AMD and other diseases. Similarly, dysregulation of retinal vessels is the catastrophic event in the progression of many retinal diseases, including AMD, diabetic retinopathy, retinopathy of prematurity, and a proportion of glaucoma cases (Rajappa et al., 2010).

In the following sections, I will describe the architecture of the retina to lay the groundwork for a detailed description of the MacTel phenotype. I will discuss the evidence that led to the hypothesis of MacTel as an inherited disease, and the important discovery of familial cases of MacTel, which shaped the rest of the project. Finally, I will introduce the research approaches I

have pursued during the course of this work, the reasoning behind the choice of methods, and both the positive and negative results achieved through these methods.

The architecture of the retina

In humans, the retina (Figure 1) develops before birth, as an outgrowth of the central nervous system. The neural retina contains six main types of cells: ganglion, amacrine, horizontal, bipolar, Mueller glia, and photoreceptors (Dowling, 1987). Each of these cell types has specialized subtypes. The retina is highly organized, with stereotyped patterning of cell types located in distinct regions (Figure 2). The most superficial cell layer (closest to the front of the eye) is the ganglion cell layer. Deep to that are the amacrine, bipolar, and horizontal cells. The Mueller glia cells extend from the outer nuclear layer to the vitreal face of the retina. The photoreceptors are located at the deepest layer. The inner limiting membrane separates the neural retina from the vitreous fluid that occupies most of the volume of the eye. The inner plexiform layer lies deep to the ganglion cell layer, and contains the synapses between ganglion, bipolar, amacrine, and horizontal cells. The inner nuclear layer consists of the cell bodies of amacrine, horizonal, and bipolar cells. The outer plexiform layer contains the projections of the photoreceptor cells that synapse with bipolar cells. The outer nuclear layer contains the nuclei of the photoreceptors, and the outer segment consists of the light responsive portion of the photoreceptors. The outer limiting membrane separates the outer segment of the photoreceptors from their nuclei. Additionally, the retinal pigment epithelium (RPE), which is not a part of the neural retina, lines the back of the retina and provides support to the photoreceptor outer segment.

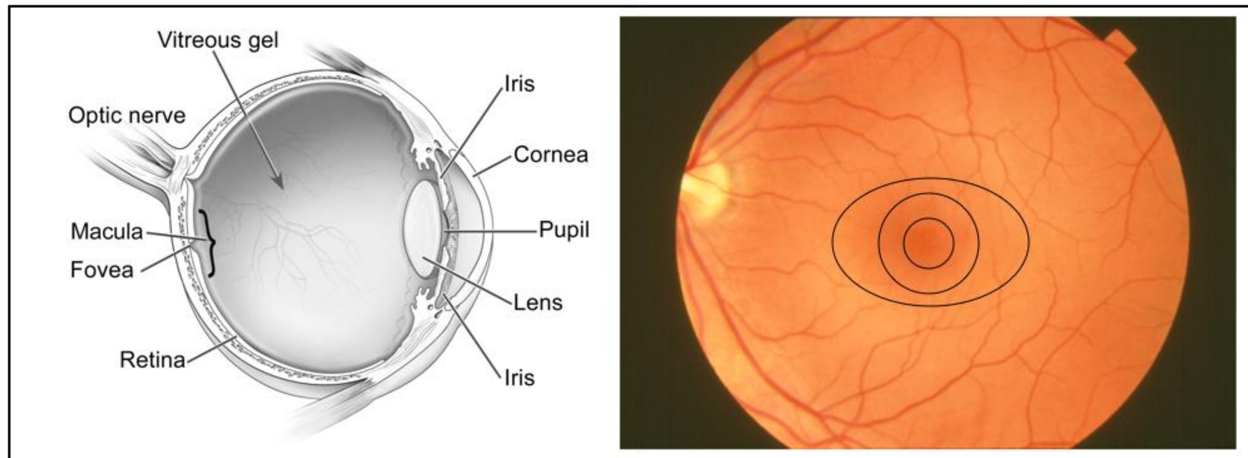


Figure 1. Anatomy of the eye. (Left) Overview of the eye. Light passes through the cornea, pupil, lens, and vitreous gel to reach the retina at the back of the eye, where photoreceptors are activated. (Right) Color fundus photograph of a normal retina. The optic disk, where the optic nerve reaches the eye, is seen as a light colored half circle on the leftmost edge of the image. Blood vessels extend from the optic disk to the periphery of the retina. The inner circle designates the fovea; the region between the the two concentric circles is the parafovea. The fovea and the parafovea lie within the macula, shown as the region within the oval (Photo credits: National Eye Institute).

The adult retina is supplied by two vascular systems: the choroid, which lies deep to the RPE, and the retinal vasculature, which is organized in an arcade of three layers, the first at the ganglion cell layer, and the second and third at the superficial and deep boundaries of the inner nuclear layer, respectively (Campochiaro, 2000).

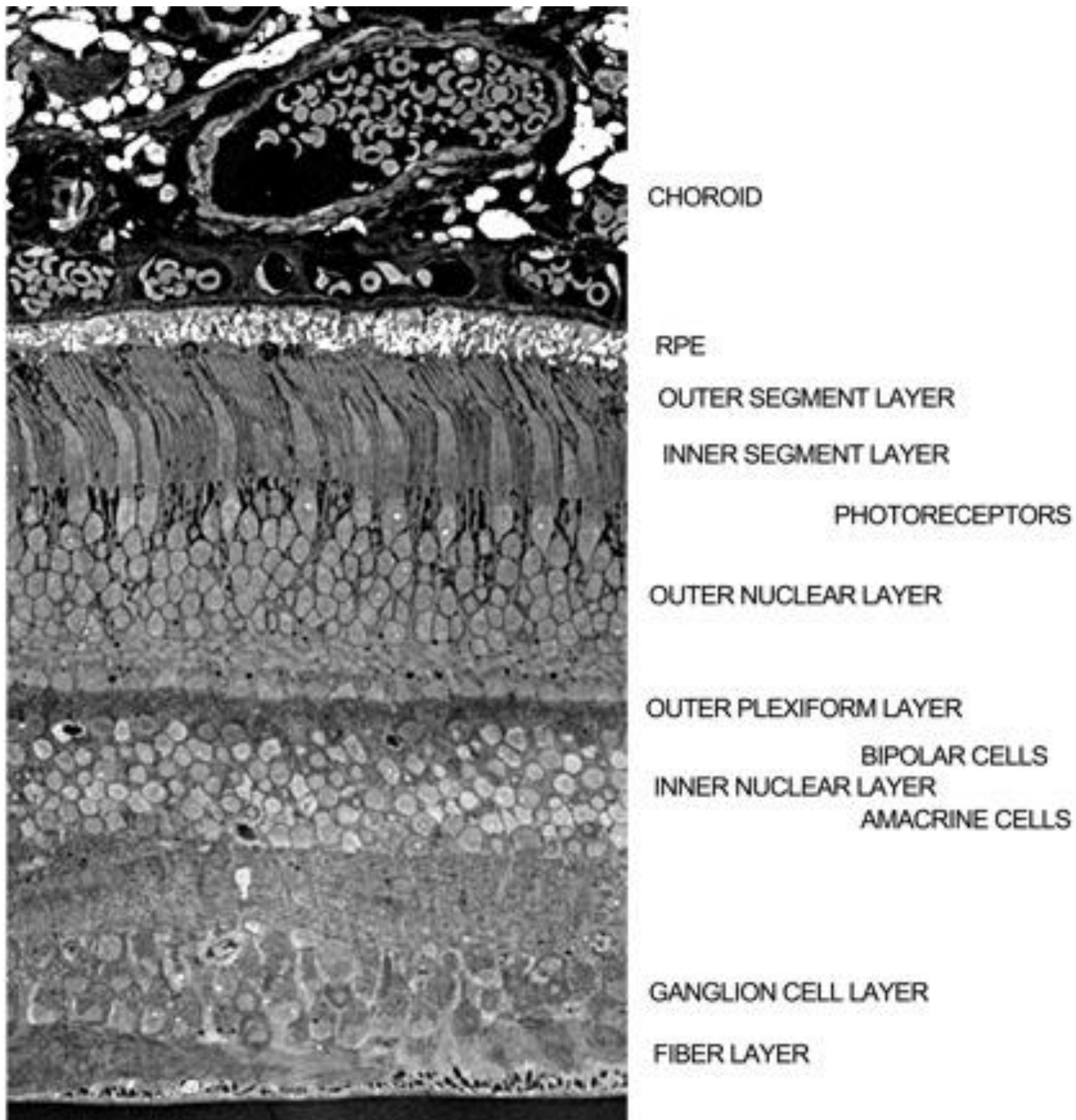


Figure 2. Vertical section of the human retina. The deepest layers (back of the eye) are at the top of the image, the superficial layers, at the bottom. (Image courtesy of Bryan William Jones, Ph.D. Marc Laboratory, Moran Eye Center. Used with permission.)

The macula is a specialized structure of the retina, found in only a few species, including primates (Finlay, 2008). At the center of the macula, is the fovea, surrounded by the parafovea. The fovea contains a dense concentration of cone photoreceptors, which provide a high degree

of central visual acuity. Disruption in this region compromises high-resolution vision, and can significantly impair the ability to perform activities such as reading and driving. The macular pigments lutein and zeathanthin are localized in the fovea and parafovea; zeathanthin is found primarily in the fovea, and lutein is found primarily in the parafovea (Loane et al., 2008b). Lutein and zeathanthin are dietary xanthophyll carotenoids. They are not synthesized physiologically, and must be obtained through the diet, converted to a biologically available form, and transported to the retina. The molecular mechanisms for their transport and deposition are poorly understood. One binding protein for zeathanthin, GSTP1, has been identified (Bhosale et al., 2004; Loane et al., 2008b), however the binding protein for lutein remains unknown.

The MacTel phenotype

Macular telangiectasia is also known as idiopathic juxtafoveal telangiectasia. The term idiopathic indicates that the cause of the disease is unknown. The first signs of disease are noted in the region adjacent to the fovea, termed the juxtafoveal region. The fovea is the part of the retina most critically responsible for functional, high-resolution vision. The term telangiectasia refers to blood vessels that are dilated, with irregular appearance.

Patients with MacTel generally experience a gradual decline in visual function, often with distortions in the visual field, known as metamorphopsia (Figure 3). Patients may experience a

scotoma, or blind spot, that can be very disruptive to activities that require attention at the center of the visual field. On examination, early signs of MacTel include changes in the transparency and reflectivity of the retina, especially with hyperreflective spots in the deep layers of the retina, depletion of the macular pigments, lutein and zeathanthin, that protect the retina from light damage, macular edema, macular holes, and dilated and tortuous blood vessels, often with capillary leakage. In advanced cases the blood vessels in the retina may become neovascular, with aberrant vessel growth invading the inner nuclear layer and disrupting the architecture of the retina. Right angle venules are frequently observed as a result of vascular dysregulation. These are vessels that do not exhibit a normal branching pattern, but rather sprout at a right angle to the vessel of origin. Vessels that extend toward the macula, which is avascular in the absence of disease, can disrupt the organization of the retina, resulting in vision loss. Atrophy of the macula occurs in the most progressed cases, with presumed loss of photoreceptors.

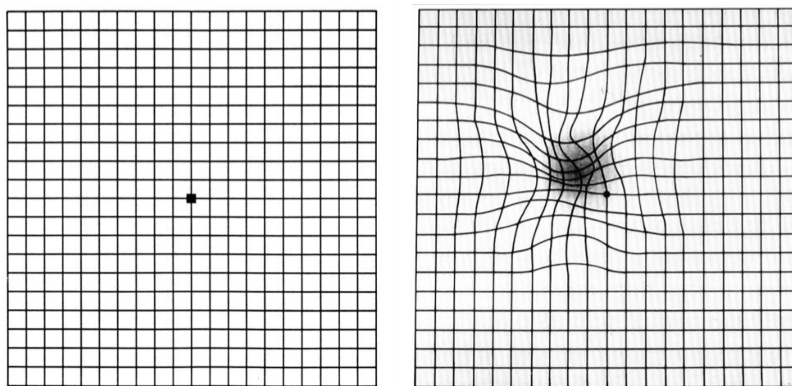


Figure 3. The Amsler grid (left) is a diagnostic tool used to screen for visual distortions, especially of the type experienced with macular disease. When viewed with one eye at a time, focusing on the center square, the lines should appear parallel, with no missing spaces. The right panel shows the way the Amsler grid might look to a patient with macular disease (Image credit: National Eye Institute).

The term telangiectasia is used to describe blood vessels that are dilated and tortuous. Normal vessels have an orderly appearance, tracing a smooth line, with few twists and turns. Telangiectatic vessels are erratic and wandering, with many twists, irregular vessel diameter, and a general loss of stereotyped patterning. Telangiectasias can occur systemically, as in ataxia telangiectasia, a neurodegenerative disorder, or locally, as in varicose veins. One of the most readily observable aspects of the MacTel phenotype is retinal telangiectasia. Severe telangiectasia is visible on color photographs of the retina, and by direct observation during an ophthalmic examination. Less pronounced disturbances can be observed with fluorescein angiography, in which a fluorescent dye is injected to image the retinal vessels. Gass and Blodi classified the retinal telangiectasias in 1993 (Gass and Blodi, 1993). Type 1, or aneurismal telangiectasia, is generally unilateral, and is most commonly seen in men. Type 2, MacTel, is generally bilateral, and affects both men and women. Type 3, occlusive telangiectasia, is extremely rare. Yannuzzi et. al. reclassified the disorders in 2006 (Yannuzzi et al., 2006), expanding on the clinical definitions with improved imaging techniques. This seminal publication emphasizes that while telangiectasia is the most prominent feature of each of these pathologies, it is not necessarily the first sign of the disease in the case of MacTel. The use of the term "telangiectasia" in naming the pathology may turn out to be a misnomer when the phenotype is more fully characterized; a vascular response to any perturbation is recognized among clinicians to be a stereotyped response to metabolic perturbations in the retina. Nonetheless, intraretinal neovascularization involves a specific tissue response, and thus, this aspect of the phenotype may yet prove to be informative in understanding the mechanism of the disease.

MacTel is characterized by changes in the transparency and reflectivity of the retina, due to changes in the distribution of the macular pigments, lutein and zeathanthin. Yannuzi noted that these changes are observed in MacTel patients without retinal aberrations, and proposed that macular pigment changes are likely primary to vascular dysregulation. Recent work by groups in the MacTel Project has expanded the ability to measure macular pigment and detected that a change in macular pigment distribution is one of the earliest signs observed in the disease (Helb et al., 2008). Work by Frank Holz, Peter Charbel-Issa, and Hendrik Scholl, collaborators in the MacTel project, has served to expand the technical capacity to capture imagery of the distribution of pigment in the macula (Charbel Issa et al., 2009a), helping to better describe the MacTel phonotype, and contributing to the overall understanding of the mechanism of macular pigment distribution. This work confirms that dysregulation of macular pigment deposition is detectable prior to, and in the absence of retinal vascular aberrations in MacTel patients.

While macular pigment distribution appears to be the more proximal facet of the MacTel phenotype, the specificity of the vascular aspect of the phenotype remains suggestive. Whether vascular effects have a primary mechanistic role in MacTel or not, there may be an underlying genetic predisposition that regulates the vascular response. If such a predisposition could be discovered, the findings might be extended to other diseases. Patients with late stages of MacTel exhibit vascular abnormalities in addition to telangiectasia. The highly ordered structure of the retina is relatively intolerant to disruption, both architecturally, and metabolically. Dysregulation of metabolic function resulting in hypoxia can lead to neovacularization (Arjamaa and Nikinmaa, 2006). The patterning of the embryonic retina occurs before birth in humans,

and is complete at the end of gestation (Fulton et al., 2009). Subsequent vessel growth, referred to as neovascularization, results in improperly patterned, and often leaky vessels. Physical disruption of the layers of the retina as a result of vessel growth, or hemorrhage from leaky vessels, can be devastating to vision. Most retinal disease involving neovascularization involves aberrant vessels arising from the choroid. Neovascularization arising from the intraretinal layer is less common, but has a significant impact in diabetic retinopathy, a subset of AMD cases, and is noted significantly in congenital blinding disorders (Fulton et al., 2009). The mechanism of the patterning and regulation of the intraretinal vessels remains somewhat obscure. It is significant that in MacTel, neovascularization arises from the intraretinal vessels, implying a different mechanism than that of choroidal neovascularization. While it appears that changes in macular pigment distribution precede vascular changes in MacTel, it is unknown whether dysregulation of the retinal vessels occurs only as a response to earlier events, whether a common system impinges on both macular pigment transport and vascular regulation, or whether genetic defects in both systems are required for the manifestation of both sets of symptoms.

Patients in early stages of the disease may exhibit signs of aberrant macular pigment distribution but not vascular defects. As the patients in the cohort are followed, it will likely become clear what proportion of cases advance to become more severe over time. Given that distinct molecular pathways are responsible for choroidal versus intraretinal vessel formation, it is likely that in MacTel, and in other neovascular diseases, an underlying genetic disposition

may be responsible for intraretinal vascularization, even if it is a response to disease conditions rather than a proximal phenotype.

The hypothesis of MacTel as an inherited disease

Indications that MacTel may be, at least in part, a genetic disease came from case reports of affected pairs of siblings, and concordant monozygotic twins (Hutton et al., 1978; Chew et al., 1986; Oh and Park, 1999; Menchini et al., 2000; Siddiqui and Fekrat, 2005; Hannan et al., 2007; Gillies et al., 2009). Given the perceived rarity of the disease, the discovery of families with multiple affected relatives suggested there could be a genetic component to the etiology of MacTel. However no systematic investigation of the incidence of the disease had been made. While anecdotal cases were reported of affected sibling pairs, it was unknown whether the relatives of MacTel probands had subclinical signs of the disease. Given that the early stages of the phenotype are subtle, and visual disturbances can manifest progressively over the course of many years, the hypothesis was made that more familial MacTel cases might be detected by systemically examining the relatives of probands.

Summary

Macular telangiectasia is a relatively rare retinal disease as compared to diseases such as AMD and glaucoma that have a very high societal burden. However, genetic dissection of this disease has the potential to elucidate aspects of macular pigment transport and retinal vascular regulation that are relevant to a number of retinal diseases that collectively pose an enormous

burden both on patients, and to society. In conjunction with the MacTel project, we have assembled a cohort of over 500 MacTel patients, and the first collection of MacTel families with multiple affected relatives. My analysis has consisted of candidate gene screening in the probands of multiplex families, linkage analysis in families, and whole exome sequencing in selected patients. This work has resulted in the identification of a region on chromosome 1 that shows statistically significant evidence of cosegregation with the MacTel phenotype in a subset of our families. The chapters that follow will discuss the details of each aspect of this work, the conclusions that can be drawn from amassed data from each inquiry, and the next steps in the genetic analysis of MacTel.

Chapter Two - Description of Sample and Introduction to Genetic Methods

Recruitment of MacTel patients and relatives

The MacTel study cohort was ascertained at 23 enrollment centers in seven countries: Australia, The United Kingdom, France, Germany, Switzerland, Israel, and the United States. Informed consent was obtained at each site in accordance with the Declaration of Helsinki. Enrollment protocols were centrally managed by the EMMES Corporation, in Rockville, Maryland. MacTel patients, and, whenever possible, relatives of the proband were given a full ophthalmic examination. Retinal images were taken and transmitted to the Reading Center at Moorfields Eye Hospital, in London. Imaging techniques consisted of color fundus photography, autofluorescent angiography, blue light reflectance imaging, and optical coherence tomography. Peripheral blood was obtained by venipuncture and sent to the Genetics Center at Columbia University. As of April 2011, 1,004 individuals had been enrolled in the genetics portion of the study. Of those, 480 were unrelated MacTel patients, 429 were relatives, and 95 were unrelated controls with no sign of the disease (Table 1a). Controls were recruited at the MacTel enrollment centers, usually as a spouse or other unrelated individual attending the clinic visit with the proband.

Relatives of 195 MacTel probands were enrolled and screened. In 114 families, 1 relative was screened in addition to the proband, in 37 families 2 relatives were screened, and in 44 families

three or more relatives were screened. The largest families were family 8, family 21, and family 156 with 12, 13, and 16 relatives screened respectively (Table 1b). The number of relatives screened was dependent on the size of the family, and the availability and willingness of relatives to participate.

Center	Location	Total	Probands	Relatives	Simplex	Controls
001	Paris, France	3	0	0	3	0
002	Nedlands, Australia	70	9	24	25	12
003	Creteil, France	4	0	0	4	0
004	Paris, France	56	12	22	22	0
005	Los Angeles, CA, USA	48	11	18	15	4
006	Melbourne, Australia	50	13	17	10	10
007	New York, NY, USA	35	10	18	3	4
008	London, UK	106	20	49	35	2
010	Cleveland, OH, USA	45	6	23	16	0
011	Sydney, Australia	164	27	86	27	24
012	San Diego, CA, USA	11	1	8	1	1
013	Munster, Germany	61	19	20	9	13
014	Tel Hashomer, Israel	37	9	13	13	2
015	New York, NY, USA	14	2	6	4	2
016	Fairfax, VA, USA	22	3	6	13	0
017	Bonn, Germany	38	6	15	15	2
019	Ann Arbor, MI, USA	73	15	28	20	10
020	Madison, WI, USA	31	6	11	10	4
021	Baltimore, MD, USA	5	1	1	2	1
022	Philadelphia, PA, USA	8	1	2	5	0
024	Bern, Switzerland	31	9	18	2	2
025	Salt Lake City, UT, USA	42	6	33	1	2
026	Miami, FL, USA	50	9	11	30	0

Table 1a. Summary of MacTel enrollment by center.

Number of relatives	Number of families
1	114
2	37
3	20
4	10
5	6
7	1
8	1
9	3
12	1
13	1
16	1
Total	195

Table 1b. Summary of the number of relatives examined per family in the MacTel cohort.

Four diagnostic categories were established by the Moorfields Reading Center: affected, possibly affected, probably not affected, and unaffected. Representative findings for each diagnostic category are given in Table 2.

Imaging Technique	Diagnostic category	Findings
Color fundus	Affected	Loss of transparency and crystal deposit temporal to the fovea
	Possibly affected	Temporal drusen
	Probably not affected	No abnormal features
	Unaffected	No abnormal features
Blue light reflectance	Affected	Clearly visible loss of transparency; telangiectatic temporal vessels; intraretinal crystals adjacent to region with loss of transparency
	Possibly affected	Slight loss of transparency temporal to the fovea
	Probably not affected	Questionable slight loss of transparency temporal to the fovea
	Unaffected	No abnormal features
Fundus autofluorescence	Affected	Patch of increased autofluorescence temporal to the fovea
	Possibly affected	Increased autofluorescence temporal to the fovea
	Probably not affected	Questionable increased autofluorescence on the temporal to the fovea. No other abnormal MacTel features
	Unaffected	No increased or decreased autofluorescence
Optical coherence tomography	Affected	Abnormal retina architecture with inner and outer empty space; break in the inner segment/outer segment line at the fovea; intraretinal pigment absent Inner retina collapsing temporally to the fovea due to both inner and outer empty spaces
	Possibly affected	Mostly normal inner segment/outer segment line, with a small break, absent inner and outer empty space. Absent intra-retina pigment. Questionable hyper-reflectivity on the temporal side.
	Probably not affected	Normal inner segment/outer segment line, no inner or outer empty space, normal retinal architecture. Absent intraretinal pigment.
	Unaffected	Normal inner segment/outer segment line, no inner or outer empty space, normal retinal architecture

Table 2. Imaging techniques used in the diagnosis of MacTel, and typical findings for each diagnostic category.

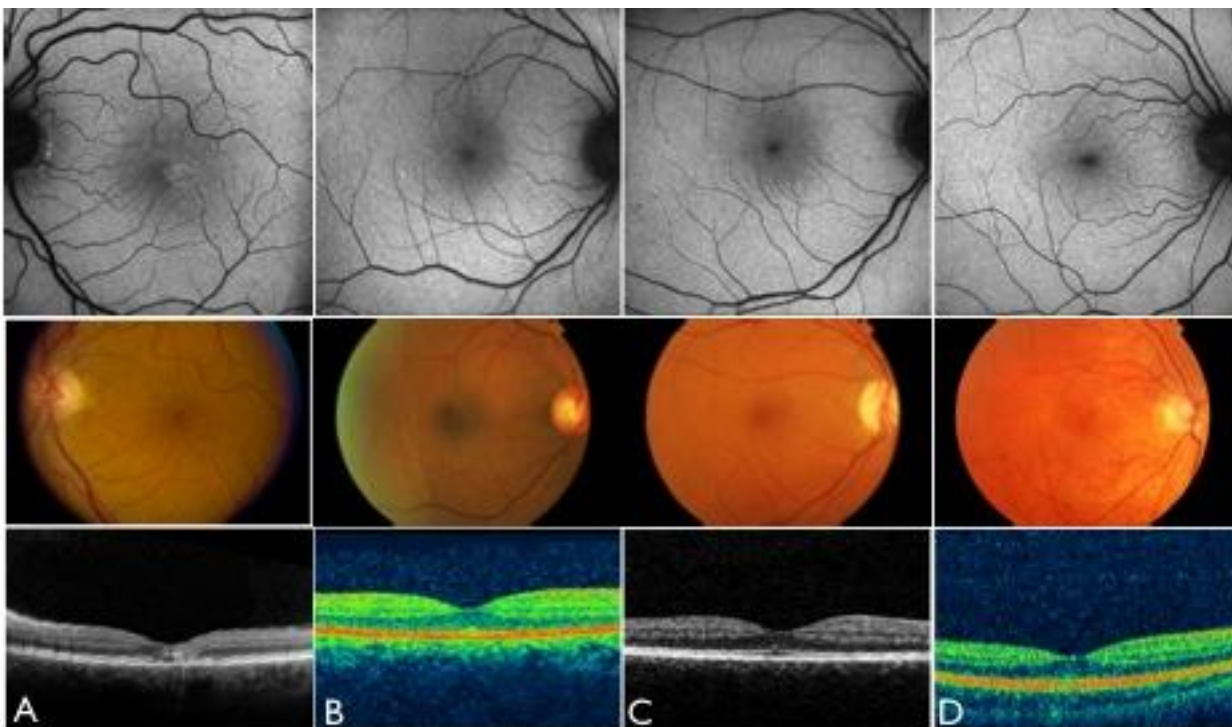


Figure 4. Retinal images from patients evaluated for MacTel. Top row: fundus autofluorescence; middle row: Color fundus; bottom row: optical coherence tomography. Panel A: Affected with MacTel; panel B: possibly affected; panel C: probably not affected; panel D: unaffected. Panel A shows the right eye, while panels B, C, and D show the left eye. Images were selected as those most representative of the diagnostic categories. Optical coherence tomography images (bottom row) are in color or black and white depending on how the image was saved by the technician.

At the outset of the MacTel Project, the inheritance pattern was thought to be complex, since families with more than one affected relative were rare, and did not appear to conform to expectations for a Mendelian inherited disease. Prior to this study, affected relatives had rarely been recorded; it was thought all but a few MacTel cases were simplex. This view changed as a result of diligent examination of the relatives of MacTel probands in the course of

investigations by researchers in the MacTel project, including the work that comprises this thesis.

A total of 51 families with more than one affected or possibly affected relative were identified in the course of this study. Of those families where all enrolled family members had been evaluated, 26% of the families had two or more relatives, including the proband, diagnosed as affected or possibly affected with MacTel. Notably, a large proportion of the simplex families consisted of families with only one relative screened in addition to the probands, suggesting that more familial cases of MacTel might be detected by screening additional relatives. A total of 429 relatives were enrolled, with confirmed diagnoses for 364 of those in April 2011. Of those relatives with a confirmed diagnosis, 32 were classified as affected, 38 as possibly affected, 61 as probably not affected, and 232 as unaffected. Altogether, 9% of relatives screened were diagnosed as affected. Including those relatives classified as possibly affected, 19% of relatives were deemed likely to have either frankly diagnosable MacTel, or very suggestive retinal findings. Notably, those relatives classified as probably not affected were frequently ascertained in pedigrees where other relatives were diagnosed as either strictly affected or possibly affected. This suggests that the relatives diagnosed as probably not affected may be exhibiting some degree of MacTel phenotype, however these cases were considered too ambiguous for analysis, and were not included in linkage studies.

Several factors complicate the assessment of inheritance patterns in families. MacTel is a late-onset disorder. At the age patients generally present to a clinic with vision loss severe enough to seek treatment, the parents of the patient are in many cases deceased or unavailable for examination; any offspring tend to be too young to exhibit signs of the disease. There are exceptions, and families with vertical transmission have been ascertained, but the majority of families with multiple affected relatives consist of sibling pairs or trios, simply because these are the relatives that are available to be enrolled.

Genetic Approaches

A decade after the announcement that the human genome had been sequenced (McPherson et al., 2001; Venter et al., 2001) the techniques available for genetic analysis are changing at such a rapid pace that methods that are reasonable at the end of this study were beyond the realm of consideration at its inception. Nonetheless, the most powerful techniques available to identify a susceptibility locus in an inherited disease rely on theoretical methods described near the beginning of the last century (Sturtevant, 1913; Fisher, 1922; Morton, 1955).

Fundamentally, the goal of genetic disease mapping is to discover a functional variant in the genome that causes or contributes to a disease. In the simplest case, for a fully penetrant, monogenic, Mendelian disorder, a variant in a single gene is sufficient to cause the disease. Many of the human monogenic traits were mapped in the 1980s; genome-wide maps of naturally occurring polymorphisms have been exploited by researchers to discover thousands

of pathogenic alleles (Antonarakis and Beckmann, 2006). These discoveries contributed to the understanding not only of the causes of genetic disease in humans, but also fueled basic research by identifying molecular pathways that could be further studied in model organisms.

Candidate gene screening

Given that most genes in the human genome are now known (current estimates place the number of protein-coding genes in the human genome at approximately 21,000 (Clamp et al., 2007; Lander, 2011)), and the function of many of these have been at least partially elucidated, one approach is to review the literature for genes with known function and expression for which there is plausible evidence that the gene could be connected to the observed phenotype. These genes can be directly sequenced to look for variants that could cause or contribute to the disease. This method has several serious limitations: while much is known about the function, expression pattern, and interactions of genes, the collected knowledge is incomplete. Given the number of genes in the genome, the odds of picking the right one to sequence are daunting. Next is the difficulty of recognizing a disease causing variant when it has been genotyped. Taken alone, detecting a base pair variant that differs from the reference sequence confers little information. Outside information is needed to estimate the frequency of the allele in the population, and to determine whether it affects gene function. Additionally, for an inherited disease presumed to be caused by the same inherited allele in all affected members of that family, the variant must be shown to have been inherited by all affected family members.

In this study, I used candidate gene screening by Sanger sequencing to evaluate genes with known retinal vascular or macular pigment transport phenotypes, genes implicated in diseases with phenotypic similarity to MacTel, and genes in a region of interest identified through family studies.

Linkage Analysis

The availability of families with multiple affected individuals greatly increases the power of a genetic study (assuming they are affected due to shared genetic factors), and is the main strength of this endeavor. Family members will share a predicted proportion of alleles identical by descent because of Mendel's laws. The expected shared proportion depends on the relationship of two individuals. For a parent-child pair, exactly one half of the genetic material is shared identical by descent. On average, a pair of siblings will share two alleles identical by descent across one quarter of the genome, one allele identical by descent across half the genome, and zero alleles identical by descent across one quarter of the genome. Relatives that share only one common ancestor will share one allele identical by descent across some fraction of the genome that becomes smaller the more distantly the two are related. The predicted allele sharing between family members can be used to look for regions of the genome that may harbor a gene that causes or contributes to a phenotype.

Linkage analysis is a method used to determine whether two loci segregate independently. When applied to the genetic study of a disease, linkage is used to try to identify a gene that is

involved in the pathology of the disease by identifying a locus that segregates with the disease more often than would be expected by chance.

With the availability of chip-based arrays to genotype single nucleotide polymorphisms, it is now trivial to determine the genotypes at millions of positions across the genome. Loci that are polymorphic can be used as markers in linkage analysis, meaning that it is possible to test many positions across the genome to determine whether a marker might be close to a gene involved in the disease.

By the law of independent assortment, during meiosis, there is an equal probability that either of two copies of a chromosome will be segregated to a gamete. By virtue of this fact, loci that are on different chromosomes have a 50:50 chance of segregating to the same gamete or to different gametes during meiosis. During meiosis, homologous chromosomes pair and crossing over occurs at least once in every pair of chromosomes. Two loci at opposite ends of the same chromosome will segregate independently as a result of the fact that a crossing over will almost certainly occur somewhere in between. When an odd number of cross over events occur between two loci, the resulting chromosome will be recombinant with respect to those two loci. Thus, when two loci are on different chromosomes, or far apart on the same chromosome, they will be inherited together 50% of the time, and they are said to be unlinked. When two loci are close together, a recombination event between them is less likely to occur. When this is the

case, alleles at the two loci will be inherited together more than 50% of the time, and the two loci are said to be linked.

The principles of genetic linkage were first defined in *Drosophila melanogaster* where controlled matings could be arranged and the phenotypes of the progeny observed (Morgan, 1910; Sturtevant, 1913). It was established that the recombination fraction, or the proportion of offspring in which a crossover occurred between two markers, could be quantified and analyzed mathematically to detect alleles that segregate together more often than would be expected by chance alone. It was later shown that this method could be applied to analysis of human families (Haldane, 1934; Penrose, 1953; Morton, 1955), where statistical power depends on the size and pedigree structure of the families available for analysis (Ott, 1999). A genetic map defines the relative distance between syntenic loci by observing the number of recombinations observed between them. Genetic distance is measured in centiMorgans (cM), where one cM is defined as the distance between loci such that on average 1 crossover is expected between the two loci in every 100 meiosis. The recombination rate varies across the genome; on average, one cM corresponds to 1 million base pairs. Genetic maps of the human genome were made using markers that are highly polymorphic (Botstein et al., 1980), allowing the localization of disease genes to a small section of the chromosome. This led to the discovery of the genetic basis of many diseases that are caused by variants of a single gene, such as Huntington's Chorea (Gusella et al., 1983) and cystic fibrosis (Rommens et al., 1988). Genetic maps were created using restriction fragment length polymorphisms (Donis-Keller et al., 1987), and later more closely spaced microsatellites (Dib et al., 1996), increasingly the resolution possible in a genome-wide linkage scan. More recently, high throughput chip technology

combined with data from the Human Genome Project and the HapMap Project has resulted in the availability of arrays to genotype thousands of single nucleotide polymorphisms (SNPs) very rapidly and affordably (Zhao et al., 1998; Li et al., 2008). SNPs are much less variable than microsatellites, as most SNPs are biallelic. The utility of a single polymorphism in linkage analysis is dependent on its frequency in the population, and on the accuracy of population data on allele frequencies (Albrechtsen et al., 2010). However, multiple SNPs can be combined in analysis to increase the information content, leading us to multipoint analysis of dense maps of informative markers, such that the resolution of linkage analysis is limited only by the number of informative meiosis available for analysis.

Association

At the outset of the MacTel project, it was unclear to what degree family studies would be possible. Until relatives of probands were examined systematically, the degree to which MacTel clusters in families was not evident. Several familial cases of MacTel had been reported, but most cases were thought to be simplex cases, with no additional affected relatives, as would be expected for a complex trait, caused by an unknown number of genetic and environmental factors. SNP chips have been used extensively in genome-wide association studies (GWAS) with mixed results (Weiss and Terwilliger, 2000; Altshuler et al., 2008; Hall et al., 2010; Rafiq et al., 2010). GWAS depends on the detection of association between a single marker allele and a disease trait in cases and controls. Association is tested by comparing the frequency of an allele in cases to its frequency in controls. Such associations would exist if there were linkage

disequilibrium between one of the markers and a variant predisposing to disease (Terwilliger et al., 1998; Terwilliger and Goring, 2000; Clark et al., 2010).

Association refers to a difference in allele frequencies between cases and controls. The presumption is that to identify a disease susceptibility locus, the identical allele will be over represented in a sample of cases as compared to controls. In the same way that the goal of linkage is to detect chromosomal segments that are inherited identical by descent in a subset of individuals in families (where the subset might be those people afflicted with a certain disease), association looks for segments of chromosomes inherited identical by descent in individuals who are very distantly related—so distantly that the last common ancestor is typically unknown, and the individuals in the sample would not consider themselves to be related. A statistically significant difference in the frequency of alleles at a locus between cases and controls may indicate that a functional variant exists in that region. A functional variant need not be the variant that is used to test association, but could be any variant that is syntenic with the marker and was also inherited identical by descent in a group of people with the disease or trait in question. Segments of chromosomes that have been inherited through many generations without recombination are referred to as linkage disequilibrium blocks, or LD blocks.

The availability of chip arrays to genotype hundreds of thousands to millions of single nucleotide polymorphisms makes it possible to test for association across the genome with the

assumption that nearly every position on the genome will be in linkage disequilibrium with some marker represented on the chip. Studies of this kind are referred to as genome-wide association studies, or GWAS. Chip arrays that have typically been used in GWAS have upwards of 600,000 markers, requiring correction for multiple testing to reduce the number of expected false positives to an acceptable value. If the traditional threshold for significance of 0.05 were used in GWAS with 1 million markers, meaning that 1 in 20 positive results would be expected to occur by chance alone, the expected number of false positives would be 50,000. To reduce this number, a more stringent threshold for significance is required, and large sample sizes are required to achieve significance.

In the MacTel study, GWAS was considered prior to the identification of multiplex families. As enrollment proceeded, multiplex pedigrees amenable to linkage analysis were identified early on, and the decision was made to pursue linkage studies rather than GWAS. I have used association where variants were detected in candidate gene screening. Where rare variants were detected by Sanger sequencing in candidate genes, these variants were screened in MacTel cases and in a library of controls recruited and screened for AMD studies. The principle of association when applied in this way is the same as in genome-wide association, however testing a single variant reduces the need for correction for multiple testing, which dramatically reduces the power of GWAS.

Exome Sequencing

Advances in next generation sequencing have recently progressed to the point that it is feasible for individual laboratories to carry out large scale sequencing rapidly, and at a reasonable cost (Ng et al., 2010b). The exome is defined as all regions of the genome that code for proteins (Ng et al., 2008). Whole exome sequencing became readily accessible in 2009 and 2010 as the cost of sequencing dropped, and sequence capture arrays were created to target the coding regions of the human genome (Hodges et al., 2007). In 2010, whole genome sequencing was not feasible for individual laboratories, though this has since changed. The rationale for targeting the exome is that many genetic diseases result from alleles in coding regions that alter the structure and function of proteins (Botstein and Risch, 2003; Ng et al., 2009). Coding variants are more tractable to identification, confirmation, and interpretation than non-coding disease causing variants. The main reasons to sequence the exome, as opposed the the genome, are cost, the lesser burden of dealing with large data sets (as compared to whole genome sequencing), and the notion that a change in amino acid sequence is easier to understand as a functional entity than a noncoding, or nongenic variant that may have a regulatory function (for example) that has not been characterized.

We employed an approach similar to Ng et al. (Ng et al., 2010a) in which detected variants were filtered to remove common polymorphisms, synonymous variants, and variants not shared between affected siblings to reduce the number of candidates. While exome sequencing has proved useful in discovering diseases alleles for several monogenic diseases(Ng et al., 2010a;

Yan et al., 2011; Zuchner et al., 2011), our analysis was hampered by technical constraints that will be discussed in detail. Our experience was shared by other groups utilizing what was at the time a very new technology. The lessons learned from this experience extend to other emerging technologies.

Analysis of risk haplotypes

Recombination mapping in sibling pairs identified chromosomal segments where at least one allele was shared identical by descent in the affected siblings. I sought to extend this analysis by determining phased haplotypes in families to identify the alleles on the risk haplotype in each family, and to compare these haplotypes across families to determine whether a founder mutation could be detected.

It was not expected that a single founder mutation would account for all cases of MacTel, but it was hoped that some proportion of cases would share a founder mutation that could be detected. Even allowing for a hypothesis of allelic or locus heterogeneity, haplotype analysis could theoretically reveal a risk haplotype enriched in the genomes of affected individuals. The goal of the analysis was to narrow the region of interest identified through linkage analysis.

A haplotype describes the alleles inherited together from one grandparent for a set of markers. Haplotype analysis across families is a way of looking for linkage disequilibrium, recognizing

that “unrelated” individuals share common distant ancestry (Terwilliger and Goring, 2009). The hypothesis underlying this experiment is that some proportion of families with inherited disease may share a common ancestor who carried a founder mutation in some gene causing the disease. The shared haplotypes that can be identified in sibling pairs are necessarily large (on average recombination occurs every 100Mb, so the length of shared segments between members of a sib pair should average about 50Mb). Inspecting two informative meioses (an affected sibling pair), it is expected that 25% of the genome will be inherited by two sibs in such a way that no alleles are shared identical by descent, 50% will have one allele IBD, and 25% of the genome would have both alleles inherited IBD. In more distant relatives, either within a family, or between families that share a distant common ancestor, a greater number of meioses have occurred, and therefore proportionally less of the genome will be shared IBD. Haplotype analysis attempts to exploit the information that can be inferred from many generations of unobserved meioses, where a greater proportion of the genome will have recombined away from the risk allele, resulting in a minimal candidate region that is more narrow than that which could be observed in any given sibling pair.

This method was chosen in this study in an attempt to narrow the minimal candidate region on chromosome 1q41-42. I will discuss how this method performed in families with parental genotypes present, and in families with missing parental genotypes, including the limitations of inferring phased haplotypes using population allele frequencies in a cohort with heterogeneous ethnicity, and the potential confounding factors related to genotyping error over a large number of markers.

INTRODUCTION TO CHAPTER THREE

The following chapter consists of published work reporting the results of candidate gene screening by Sanger sequencing in a subset of family probands with Macular Telangiectasia. This work was carried out early in the project, during the early phase of enrollment of patients and families. Several genes were compelling candidates based on mutant phenotypes in mouse models with plausible similarity to aspects of the MacTel phenotype, and known disease association with human retinal diseases. While no associations were found between variants in any of these genes and MacTel, this analysis allowed us to rule out the exonic regions of these genes as harboring a causative variant for MacTel. Additionally, a compound allele in the Wnt receptor Frizzled-4 that had beeen previously reported to be a causative variant in another disease was determined to be a benign polymorphism.

Molecular Vision 2010; 16:2718-2726 <http://www.molvis.org/molvis/v16/a290>
Received 11 June 2010 | Accepted 9 December 2010 | Published 14 December 2010

© 2010 Molecular Vision

Analysis of candidate genes for macular telangiectasia type 2

Nancy L. Parmalee,[1,2] Carl Schubert,[1] Joanna E. Merriam,[1] Kaija Allikmets,[1] Alan C. Bird,[3] Mark C. Gillies,[4] Tunde Peto,[3] Maria Figueroa,[5] Martin Friedlander,[6] Marcus Fruttiger,[7] John Greenwood,[7] Stephen E. Moss,[7] Lois E.H. Smith,[8] Carmel Toomes,[9] Chris F. Inglehearn,[9] Rando Allikmets[1,10]

[1]Department of Ophthalmology, Columbia University, New York, NY; [2]Department of Genetics and Development, Columbia University, New York, NY; [3]Moorfields Eye Hospital, London, UK; [4]Save Sight Institute, Department of Clinical Ophthalmology and Eye Health, The University of Sydney, Sydney, Australia; [5]The EMMES Corporation, Rockville, MD; [6]Department of Cell Biology, The Scripps Research Institute, La Jolla, CA; [7]Department of Cell Biology, University College London Institute of Ophthalmology, London, UK; [8]Department of Ophthalmology, Harvard Medical School, Children's Hospital Boston, Boston, MA; [9]Section of Ophthalmology and Neuroscience, Leeds Institute of Molecular Medicine, St James's University Hospital, Leeds, UK; [10]Department of Pathology and Cell Biology, Columbia University, New York, NY

Purpose: To find the gene(s) responsible for macular telangiectasia type 2 (MacTel) by a candidate-gene screening approach.
Methods: Candidate genes were selected based on the following criteria: those known to cause or be associated with diseases with phenotypes similar to MacTel, genes with known function in the retinal vasculature or macular pigment transport, genes that emerged from expression microarray data from mouse models designed to mimic MacTel phenotype characteristics, and genes expressed in the retina that are also related to diabetes or hypertension, which have increased prevalence in MacTel patients. Probands from eight families with at least two affected individuals were screened by direct sequencing of 27 candidate genes. Identified nonsynonymous variants were analyzed to determine whether they co-segregate with the disease in families. Allele frequencies were determined by TaqMan analysis of the large MacTel and control cohorts.
Results: We identified 23 nonsynonymous variants in 27 candidate genes in at least one proband. Of these, eight were known single nucleotide polymorphisms (SNPs) with allele frequencies of >0.05; these variants were excluded from further analyses. Three previously unidentified missense variants, three missense variants with reported disease association, and five rare variants were analyzed for segregation and/or allele frequencies. No variant fulfilled the criteria of being causal for MacTel. A missense mutation, p.Pro33Ser in frizzled homolog (*Drosophila*) 4 (*FZD4*), previously suggested as a disease-causing variant in familial exudative vitreoretinopathy, was determined to be a rare benign polymorphism.
Conclusions: We have ruled out the exons and flanking intronic regions in 27 candidate genes as harboring causal mutations for MacTel.

Macular telangiectasia type 2 (MacTel) is a rare, adult onset retinal disease that results in tortuous and dilated retinal vessels, macular pigment changes, macular edema, and in some cases macular holes. It is also referred to in the literature as idiopathic juxtafoveal, or juxtafoveolar, telangiectasia. A classification system was introduced by Gass and Blodi [1] and updated in 2006 by Yannuzzi [2], distinguishing the features of three types of macular telangiectasias. MacTel, or type 2, is a bilateral disease that affects both genders, as opposed to type 1, which is often unilateral, with aneurism and exudates, and generally presents only in men. Type 3, characterized by occlusive telangiectasia, is very rare. The three forms of idiopathic macular telangiectasia are described together to distinguish and classify phenotypically similar pathologies; however, the findings and progression of the three are distinct, and it is believed that each arises from a distinct etiology.

In 2005, The MacTel Project—an international consortium of clinicians and basic science researchers—was established to study the cause, natural history, progression, and epidemiology of the disease, and to explore potential treatments. Publications from collaborators in the MacTel Project have further described the clinical findings, diagnostic methods, and epidemiology related to the disease [3-11]. In advanced cases, neovascularization may be present, arising from the intraretinal vascular plexus. While most reported cases are sporadic, affected sib pairs and affected pairs of monozygotic twins were described in the literature [12,13], leading to the hypothesis that MacTel had a genetic cause in a significant proportion of cases. Ophthalmic examination of relatives of MacTel patients revealed that many families had family members who experienced no vision loss, but did exhibit subtle signs of the disease [14]. Since vertical transmission is observed in several families, an autosomal

Correspondence to: Rando Allikmets, PhD, Eye Institute Research Addition, Columbia University, 160 Fort Washington Avenue, 2nd Floor, Room 202, New York, NY,10032; Phone: (212) 305-8989; FAX: (212) 305-7014; email: rla22@columbia.edu

dominant model of Mendelian inheritance with reduced penetrance is assumed.

In conjunction with the MacTel Project, we have assembled a cohort of MacTel patients and their relatives for a genetic study. One mode of investigation in our studies has been to identify the disease-causing gene(s) by the direct sequencing of candidate genes in family probands, followed by segregation analysis of families. The current hypothesis is that a dominant causal mutation would be a rare heterozygous variant in all affected individuals. Allele frequencies were determined by screening cases and controls for unknown variants, and for known variants where population frequency data was unavailable.

We selected candidate genes that were known to be causative in diseases with phenotypic similarity to MacTel (e.g., familial exudative vitreoretinopathy [FEVR] and Norrie disease [15-19]), genes with a known molecular genetic role in retinal vascularization or macular pigment transport, and genes of functional interest that lay in regions of interest based on familial linkage studies. Twenty-seven genes were identified and screened based on these criteria.

METHODS

Study population: Participants were enrolled at 23 centers in seven countries (Australia, Germany, France, the UK, Israel, Switzerland, and the United States). Each center received approval from their governing human subjects review board. Informed consent was obtained in accordance with the Declaration of Helsinki. Records of informed consent and human subject approvals for all participating centers were centrally managed by the EMMES Corporation. Study subjects were enrolled based on a diagnosis of MacTel by the principal investigator at each center. Further criteria for enrollment were that patients be at least 18 years of age, be of European ancestry, and be free of diabetic retinopathy or other retinal disease [20]. Relatives of patients diagnosed with MacTel were actively recruited. Age- and ethnicity-matched controls were concurrently recruited—often a spouse or other unrelated individual present at the clinic appointment with the proband. At the time screening was performed, the MacTel cohort consisted of DNA from 200 unrelated probands diagnosed with MacTel. Since March 2010, study enrollment has been ongoing; a total of 735 samples have been sent to Columbia University's Center for Human Genetics. Of those samples, 360 are unrelated cases, and the rest are mostly unaffected relatives. The youngest MacTel proband in the study was 25 years old at the time of enrollment, and the oldest was 85. The majority of probands were between 50 and 70 years of age when they were enrolled.

The families of the probands sequenced consisted of five affected sibling pairs (ASPs), two affected sibling trios, one affected parent and child duo, and additional relatives. Specifically, family 8 (Figure 1) consisted of 11 individuals, including an ASP, two additional unaffected siblings, one affected and one unaffected parent, two uncles, and three cousins. Family 9 consisted of an ASP and three unaffected and one possibly affected offspring of the siblings. Families 22, 30, and 81 were ASPs. Family 29 was an affected sibling trio plus six unaffected adult offspring of the siblings. Family 42 was a three-generation family consisting of an affected trio, an affected uncle, three unaffected individuals in the second generation, and three unaffected adult offspring in the third generation. Family 101 was a discordant sibling pair with one affected and one unaffected parent.

The Columbia University control cohort consisted of DNA samples from individuals recruited as controls for studies of age-related macular degeneration (AMD). Participants were matched by age and ethnicity to the AMD cohort, and were found free of macular disease at the time of recruitment, as previously described [21,22]. Briefly, controls underwent ophthalmic examination to screen for retinal disease. Stereo fundus photographs were evaluated using standard classification systems. Individuals accepted as controls had no family history of AMD, and were determined to be free of retinal disease. Three hundred and sixty-eight controls from the Columbia cohort were screened for this study. In addition, a cohort of 639 AMD patients was screened for selected variants. Data obtained from screening the AMD cohort was beneficial in that the patients in this cohort were of advanced age and had undergone thorough retinal examination, thereby ruling out MacTel and increasing the number of individuals classified as controls.

Diagnosis of MacTel: Participants were given a standardized ophthalmic examination, including best corrected visual acuity, fundus examination with photography, fluorescein angiography, optical coherence tomography, and blue light reflectance. Images were taken and sent to Moorfields Eye Hospital's Reading Centre, in London, England, for evaluation. The criteria for diagnosis are described by Clemons et al. [20]. Diagnostic features of MacTel are based on the Gass and Blodi criteria [1], and include loss of transparency in the perifoveal region, dilated and telangiectatic blood vessels, especially in the temporal retina, and crystalline deposits. In cases where the adjudication made at the reading center was not in accordance with the diagnosis made at the recruiting center, the reading center diagnosis was used to code the sample for genetic studies. Each sample was assigned to one of four diagnostic categories: affected, possibly affected, probably not affected, or unaffected. Patients were reevaluated at regular intervals over the course of the study.

Sequencing and genotyping: DNA was isolated from whole blood by column purification (DNA Blood Maxi, 51194; Qiagen, Valencia, CA). Eight probands of families with more than one affected individual were screened by direct Sanger sequencing for mutations in 27 candidate genes. For each

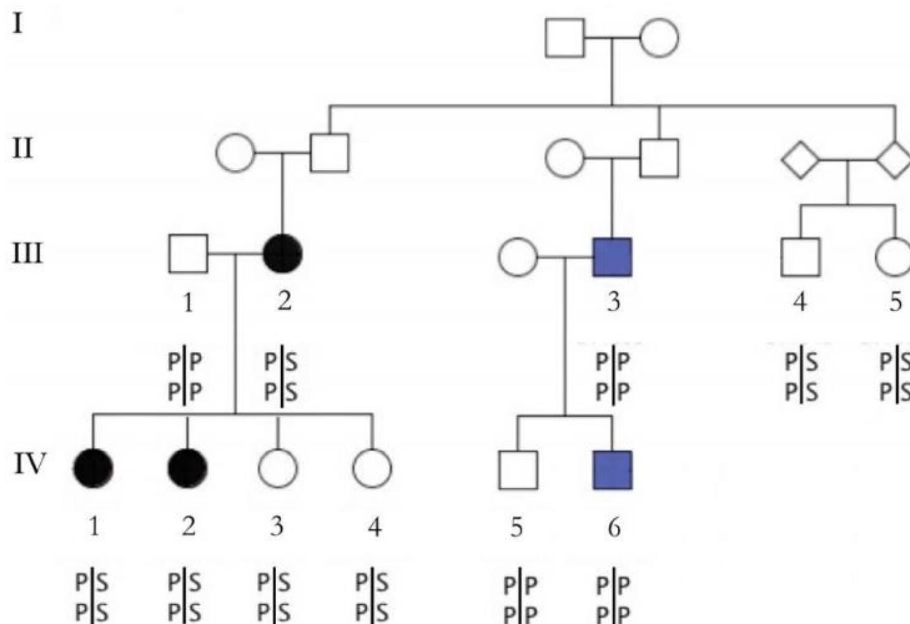


Figure 1. Segregation of the p.Pro33Ser (top)/p.Pro168Ser (bottom) compound variant in a family with inherited macular telangiectasia type 2. Black filled circles represent affected family members; blue filled circles represent possibly affected family members. The numbered individuals are those for whom DNA was available for analysis.

gene, primers were designed to amplify each exon and flanking intronic sequences. Primer sequences are listed in Appendix 1. PCR reactions were performed with 2 ng of genomic DNA in a total volume of 25 µg, using 25 pmol each of forward and reverse primer, 200 µM dATP, dCTP, dGTP, dTTP, 2.5 mM $MgCl_2$, 1.5 U Taq Polymerase (Hot Fire DNA Polymerase, Solis Biodyne, Tartu, Estonia, or AmpliTaq Gold, Applied Biosystems, Carlsbad, CA), and 10× buffer supplied by the manufacturer. Thermocycling was performed using either the Stepdown protocol, or the Touchdown (68–55 °C) protocol. Stepdown: an initial 12 min denaturation step at 95 °C was followed by 12 cycles of 95 °C for 12 s, 65 °C for 20 s (with a 0.5 °C reduction in temperature for each cycle), and 72 °C for 55 s. This was followed by 30 cycles of 95 °C for 12 s, 50 °C for 20 s, and 72 °C for 55 s, with a final 7 min extension at 72 °C. Touchdown (68–55 °C): an initial 12 min denaturation step at 95 °C was followed by 26 cycles of 95 °C for 15 s, 68 °C for 20 s (with a 0.5 °C reduction in temperature for each cycle), and 72 °C for 45 s. This was followed by 15 cycles of 95 °C for 15 s, 55 °C for 20 s, and 72 °C for 45 s, with a final 7 min extension at 72 °C. Sequencing was performed by Genewiz (South Plainfield, NJ). One hundred and twelve familial samples were genotyped on the Illumina (Illumina, Inc., San Diego, CA) 1M chip for ongoing linkage studies (data not shown). These data were used to evaluate the segregation of variants detected by sequencing when the variant was a known single nucleotide polymorphism (SNP) that was genotyped on the chip.

Sequences were analyzed for known or unknown variants that differed from the reference sequence. Nonsynonymous coding variants were evaluated using the population frequencies published in the Single Nucleotide Polymorphism database (dbSNP). Unknown, infrequent, or disease-associated variants were evaluated to determine whether they co-segregated with the disease within the families in which they were discovered.

For known variants that were present in the Illumina 1M chipset, genotypes of relatives in the family were inspected to determine whether the variant co-segregated with the disease. For unknown variants, and for rare variants not included on the 1M chip, tests for co-segregation were performed by sequencing the family members of the proband in whom the variant was detected. For variants detected in a family with vertical transmission, where both parents were available, segregation was assessed based on whether the allele was inherited from the affected or the unaffected parent. For variants detected in affected sib pairs whose parents were not available, if the variant was present in both siblings, follow-up analysis was performed by TaqMan assay (Applied Biosystems, Foster City, CA) in the cohort of MacTel probands and in controls, to determine allele frequencies in these cohorts. Variants that were determined to be common polymorphisms based on population frequencies from dbSNP were not pursued further. In some cases, population frequencies were not available in dbSNP, yet the variant was highly polymorphic in the genotyped cohort and was present in both affected and unaffected individuals. Such variants were classified as frequent variants and not pursued further. Variants that merited further analysis either because co-segregation of the variant with the disease could not be ruled out, or because the allele detected was rare or had been associated with disease, were screened by TaqMan assay in

the entire MacTel cohort, and in a cohort of controls. For selected variants, we also screened an available cohort of AMD patients to determine allele frequencies more precisely. Analyzing the large AMD cohort, in addition to the MacTel and control cohorts, provided additional allele frequency data for previously unknown variants and known variants where population frequency data was unavailable. Allele and genotype frequencies were compared between cases and controls with standard statistical tests, such as a 2x2 contingency table and Fisher's exact text.

RESULTS

A summary of the screened genes grouped by selection criteria is shown in Table 1. Descriptions of the genes screened and the rationale for selecting candidate genes were as follows.

Genes involved in angiogenesis: One of the defining phenotypes of MacTel is the presence of telangiectatic blood vessels in the retina and, in advanced disease, intraretinal neovascularization. This form of neovascularization is prevalent in diabetic retinopathy and in retinopathy of prematurity (ROP), and represents about 10%–15% of neovascular changes in AMD [23-27]; the remaining 85%–90% of neovascularization in AMD involves aberrant vessels arising from the choroidal vasculature (CNV). The vasculature of the retina is highly specialized; thus, the specific location of the aberration is likely a result of highly tissue-specific molecular genetic interactions. We selected 11 genes related to angiogenesis (Table 1), including the Wnt receptor frizzled-4, its ligand, norrin (*NDP*), and the co-receptor, low-density lipoprotein receptor-related protein 5 (*LRP5*). Mutations in these genes have been implicated in FEVR, Norrie disease, and ROP [28-30]. Mouse knockout models of frizzled homolog 4 (*Drosophila*); (*FZD4*) [15] and NDP [31] show a lack of intraretinal vessel formation. Other genes involved in angiogenesis or vessel regulation that were screened are angiogenic factor with G patch and FHA domains 1 (*AGGF1*) [32], angiopoietin 1 (*ANG1*) [33], dickkopf homolog 1 (*Xenopus laevis*); (*DKK1*) [34], hypoxia inducible factor 1, alpha subunit (basic helix–loop–helix transcription factor); (*HIF1A*) [35], serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1 (*PEDF*) [36], thrombospondin 1 (*THBS1*) [37], tyrosine kinase, endothelial (*TIE2*) [38], and von Hippel-Lindau tumor suppressor (*VHL*) [39].

Genes involved in macular pigment transport: MacTel is also characterized by macular pigmentary changes, with advanced cases often lacking macular pigment altogether. Little is known about the molecular genetics of macular pigment transport in the retina. In the healthy retina, the two macular pigments, lutein and zeaxanthin, filter damaging blue light [40,41]. The proteins responsible for lutein transport are unknown. The macular pigment genes screened were glutathione S-transferase pi 1 (*GSTP1*), a binding protein for zeaxanthin [42], and scavenger receptor class B, member 1 (*SCARB1*), which has been proposed as a lipid transporter in the retina [43].

Genes identified from expression arrays: Several genes were screened that were identified as differentially expressed in mouse models intended to mimic some aspects of the MacTel phenotype (data not shown). From the top of this list, five genes were screened that were identified as also having possible functional relevance by playing a role in MacTel: apelin receptor (*AGTRL1*), apelin (*APLN*) [44], complement factor B (*CFB*) [45], leucine-rich alpha-2-glycoprotein 1 (*LRG1*), and plasmalemma vesicle associated protein (*PLVAP*) [46].

Genes identified from suggestive linkage regions: Linkage studies were performed using families in which at least one family member in addition to the proband was diagnosed as affected by MacTel. While these results will be reported separately, several genes of possible functional interest were identified in regions of suggestive linkage on chromosomes 1, 7, 10, and 12 during the course of analysis. Cerebral cavernous malformation 2 (*CCM2*) [47], insulin-like growth factor binding protein 3 (*IGFBP3*) [48], sarcospan (Kras oncogene-associated gene); (*SSPN*) [49], and transforming growth factor, beta 2 (*TGFB2*) [50] were selected and screened as candidates under these criteria. Each of these genes has been proposed to be involved in angiogenesis or regulation of blood vessels; *IGFBP3* and *TGFB2* have also been proposed as genes related to diabetes. Linkage studies are ongoing as additional families are recruited.

Genes involved in diseases with related phenotypes: An increased prevalence of hypertension and diabetes are found in MacTel patients [51]. Genes involved in these diseases, which are also expressed in the retina, especially in the vasculature, were considered as candidates. Succinate receptor 1 (*SUCNR1*) [52], angiotensin II receptor, type 1 (*AGTR1*) [53], aldehyde dehydrogenase 3 family, member A2 (*ALDH3A2*) [54,55], very low density lipoprotein receptor (*VLDLR*) [56], and oxoglutarate (alpha-ketoglutarate) receptor 1 (*OXGR1*) were screened on this basis, in conjunction with linkage or expression array data, or personal communication with collaborators.

Table 1 summarizes the variants detected by the complete sequencing of all exons and adjacent intronic sequences in 27 candidate genes in eight MacTel probands. In total, we discovered three unknown and 20 known missense changes, and 22 synonymous and 61 intronic variants. Frequent variants with reported minor allele frequencies (MAFs) over 0.10 were not analyzed further unless the variant was reported to be disease-associated (*PEDF* p.Thr72Met and *GSTP1* p.Ile105Val were screened in cases and controls, though they are frequent variants). Variants with published population frequencies between 5%–10% were assessed for further

Molecular Vision 2010; 16:2718-2726 <http://www.molvis.org/molvis/v16/a290> © 2010 Molecular Vision

Table 1. Genes sequenced and missense variants detected by Sanger sequencing in 8 macular telangiectasia probands.

Gene/ category	Variant	rs number	MAF MacTel cases (400 chromosomes) /MAF AMD cases (1278 chromosomes)	MAF Controls (736 chromosomes)	MAF dbSNP	Notes
Vascular/angiogenic						
AGGF1	p.Pro698Thr	rs34400049	NS	NS	0.28	FV, DNS
ANG1	p.Thr257Arg	-	0.005 (1/400)	0 (0/736)	-	DNS
DKK1	None					
FZD4	p.Pro33Ser	rs61735304	0.02 (6/400)/.03 (17/1278)	0.01 (13/736)	ND	DNS, MT, AMD
	p.Pro168Ser	-	NS	NS	-	Allelic with P33S
HIF1A	None					
LRP5	p.Val667Met	rs4988321	0.07 (28/400)	0.06 (41/736)	0.03	MT, DNS
	p.Gln1192Arg	-	0.005 (1/400)	0 (0/736)	-	MT, DNS
	p.Ala1130Val	rs3736228	NS	NS	0.12	FV, DNS
NDP	None					
PEDF	p.Thr72Met	rs1136287	0.37 (148/400)/.34 433/1278)	0.33 (245/736)	0.36	FV, MT, AMD
THBS1	p.Thr523Ala	rs2292305	NS	NS	0.1	FV, DNS
TIE2	p.Val486Ile	rs1334811	NS	NS	0.03	DNS
	p.Val600Leu	rs35030851	NS	NS	0.04	DNS
VHL	None					
Expression microarray						
AGTRL1	None					
APLN	None					
CFB	p.Leu9His	rs4151667	0.04 (16/400)/.02 (23/1278)	0.04 (29/736)	0.07	DNS, MT, AMD
	p.Arg32Gln	rs641153	0.10 (40/400)/.04 (51/1278)	0.12 (88/736)	0.1	DNS, MT, AMD
	p.Gly252Ser	rs4151651	NS	NS	0.04	DNS
	p.Lys533Arg	-	0.04/.02 (23/1278)	0.04 (29/736)	-	DNS, AMD, LD
LRG1	p.Pro133Ser	rs966384	NS	NS	0.35	DNS, FV
PLVAP						
Macular pigment						
GSTP1	p.Ile105Val	rs1695	0.33 (132/400)/.30 (383/1278)	0.3	0.39	MT, AMD, FV
	p.Ala114Val	rs1138272	0.08 (32/400)/.07 (87/1278)	0.06 (46/736)	0.12	MT, AMD, FV
SCARB1	p.Val135Ile	rs5891	0.03 (12/400)	0.02 (15/736)	0.01	MT DNS
	p.Pro376Leu	rs74830677	0.01 (3/400)	0.01 (5/736)	0.05	MT, DNS
Genes under suggestive linkage peaks						
CCM2	p.Val120Ile	rs11552377	NS	NS	0.17	FV, DNS
IGFBP3	None					
SSPN	None					
TGFB2	None					
Disease related genes						
AGTR1	p.Ala244Ser	rs12721225	NS	NS	0.03	DNS
ALDH3A2	None					
OXGR1	p.Thr205Ala	-	0.005 (1/400)	0 (0/736)	-	DNS
SUCNR1	None					
VLDLR	None					

NS represents not screened, ND represents not determined, FV represents frequent variant, DNS represents Does not segregate with disease, MT represents Screened in MacTel cases and controls, AMD represents Screened in AMD cohort. LD represents the *CFB* p.Lys533Arg variant is in complete linkage disequilibrium with p.Leu9His. No comparison between MacTel cases and controls reached statistical significance.

screening, based on whether the variant had been reported to be associated with any diseases with phenotypes similar to that of MacTel. Twelve missense variants were screened by TaqMan assay (Table 1) in MacTel cases and unaffected controls; six of these variants were also screened in a large AMD cohort. Of the variants detected, three were unknown (*ANG1* p.Thr257Arg, *LRP5* p.Gln1192Arg, and *SCARB1* p.Ile135Val), three had been reported as possibly disease-associated (*FZD4* p.Pro33Ser, *GSTP1* p.Val105Ile, and *PEDF* p.Thr73Met). The remainder had low reported MAFs. None of the variants found by sequencing segregated with the disease. Of the variants screened by TaqMan assay, only *GSTP1* p.Val105Ile showed a trend toward a statistically significant frequency difference between cases and controls (p=0.09), suggesting it could be a possible modifier, but not a causal gene for MacTel.

The *FZD4* variants p.Pro33Ser and p.Pro168Ser were detected in the proband III2 (family 8, Figure 1). We sequenced all members of family 8 and found both the p.Pro33Ser and p.Pro168Ser variants present in two affected daughters, one unaffected daughter, and one unaffected cousin of the proband, indicating that the complex allele containing both mutations did not segregate with the disease (Figure 1). The p.Pro33Ser variant was analyzed by TaqMan assay in 200 MacTel cases, 368 unrelated controls, and 639 AMD cases to determine allele frequencies. This variant was found in one other unrelated MacTel proband (A5). The p.Pro168Ser mutation was also present in each individual carrying p.Pro33Ser. Thirteen controls were heterozygous for p.Pro33Ser (MAF=0.018). In 639 unrelated AMD samples, 16 heterozygotes and one homozygote for p.Pro33Ser were detected (MAF=0.013). In conclusion, there was no statistically significant difference in allele frequencies between cases and controls.

DISCUSSION

We have shown that the *FZD4* p.Pro33Ser /p.Pro168Ser complex allele, which has been reported as causative in FEVR and ROP [57,58], is present in 2% of unaffected controls, and therefore is not a disease-causing variant in MacTel, FEVR, or ROP. The same allele was detected in one MacTel patient, prompting the hypothesis that FEVR and MacTel may be allelic diseases. Given that FEVR has phenotypic similarities to MacTel, in that the intraretinal vascular plexus is perturbed in both diseases, dysregulation of *FZD4* was a plausible hypothesis in the etiology of MacTel. Both FEVR and MacTel also exhibit variable expressivity [59]. In both diseases, affected family members are often unaware that they are affected until a diagnosis is made after thorough examination. Segregation analysis in one MacTel family and case-control association analysis using a large cohort of controls revealed that p.Pro33Ser /p.Pro168Ser, which had been reported as a disease causing mutation, rather, is a benign polymorphism present at a low frequency in the general population. Accordingly, we conclude that it is not causative in either MacTel or FEVR, because our control cohort had no documented retinal disease. This result highlights the importance of segregation analysis in families, and of screening a sufficient number of controls to distinguish between causal mutations and rare benign polymorphisms.

Candidate gene screening is a widely used method for detecting disease-associated variants and genes. While often criticized as a "needle in a haystack" approach, it has been successful in determining some disease-associated genes, most notably the major AMD-associated genes, *CFH* and *CFB* [21,22]. In this study, 27 possible candidate genes for MacTel were selected based on a combined set of criteria. The exons and flanking intronic regions of all genes were screened by direct sequencing, with follow-up segregation analysis in families and TaqMan genotyping in large cohorts for rare, unknown, or previously disease-associated missense variants. No variants segregated with the disease, and none showed significant association with MacTel, allowing exclusion of the coding regions of these genes as harboring a causal mutation for MacTel. The causal gene(s) for MacTel are currently being searched for by a combination of linkage analyses and whole-genome sequencing.

ACKNOWLEDGMENTS

The authors are very thankful to collaborators in the entire project, including the coordinating center, the reading center, clinical enrollment sites, and basic science laboratories. For more information on the MacTel Project, and for a list of collaborators and enrollment centers, please see the MacTel Project website. The authors also wish to express our gratitude to the Lowy Medical Research Foundation for their generous support in making this work possible. We offer our deepest thanks to the macular telangiectasia patients and their families for choosing to participate in this work.

REFERENCES

1. Gass JD, Blodi BA. Idiopathic juxtafoveolar retinal telangiectasis. Update of classification and follow-up study. Ophthalmology 1993; 100:1536-46. [PMID: 8414413]
2. Yannuzzi LA, Bardal AM, Freund KB, Chen KJ, Eandi CM, Blodi B. Idiopathic macular telangiectasia. Arch Ophthalmol 2006; 124:450-60. [PMID: 16606869]
3. Charbel Issa P, Finger RP, Helb HM, Holz FG, Scholl HP. A new diagnostic approach in patients with type 2 macular telangiectasia: confocal reflectance imaging. Acta Ophthalmol (Copenh) 2008; 86:464-5.
4. Charbel Issa P, Berendschot TT, Staurenghi G, Holz FG, Scholl HP. Confocal blue reflectance imaging in type 2 idiopathic macular telangiectasia. Invest Ophthalmol Vis Sci 2008; 49:1172-7. [PMID: 18326746]
5. Charbel Issa P, van der Veen RL, Stijfs A, Holz FG, Scholl HP, Berendschot TT. Quantification of reduced macular pigment optical density in the central retina in macular telangiectasia type 2. Exp Eye Res 2009; 89:25-31. [PMID: 19233170]
6. Aung KZ, Wickremasinghe SS, Makeyeva G, Robman L, Guymer RH. The prevalence estimates of macular

telangiectasia type 2: the Melbourne Collaborative Cohort Study. Retina 2010; 30:473-8. [PMID: 19952995]

7. Chew E, Gillies M, Bird A. Macular telangiectasia: a simplified classification. Arch Ophthalmol 2006; 124:573-4. [PMID: 16606887]
8. Charbel Issa P, Holz FG, Scholl HP. Metamorphopsia in patients with macular telangiectasia type 2. Doc Ophthalmol 2009; 119:133-40. [PMID: 19711108]
9. Finger RP, Charbel Issa P, Fimmers R, Holz FG, Rubin GS, Scholl HP. Reading performance is reduced by parafoveal scotomas in patients with macular telangiectasia type 2. Invest Ophthalmol Vis Sci 2009; 50:1366-70. [PMID: 18997085]
10. Charbel Issa P, Scholl HP, Gaudric A, Massin P, Kreiger AE, Schwartz S, Holz FG. Macular full-thickness and lamellar holes in association with type 2 idiopathic macular telangiectasia. Eye (Lond) 2009; 23:435-41. [PMID: 18259211]
11. Charbel Issa P, Helb HM, Rohrschneider K, Holz FG, Scholl HP. Microperimetric assessment of patients with type 2 idiopathic macular telangiectasia. Invest Ophthalmol Vis Sci 2007; 48:3788-95. [PMID: 17652753]
12. Hannan SR, Madhusudhana KC, Rennie C, Lotery AJ. Idiopathic juxtafoveolar retinal telangiectasis in monozygotic twins. Br J Ophthalmol 2007; 91:1729-30. [PMID: 18024833]
13. Menchini U, Virgili G, Bandello F, Malara C, Rapizzi E, Lanzetta P. Bilateral juxtafoveolar telangiectasis in monozygotic twins. Am J Ophthalmol 2000; 129:401-3. [PMID: 10704569]
14. Gillies MC, Zhu M, Chew E, Barthelmes D, Hughes E, Ali H, Holz FG, Scholl HP, Charbel Issa P. Familial asymptomatic macular telangiectasia type 2. Ophthalmology 2009; 116:2422-9. [PMID: 19815294]
15. Xu Q, Wang Y, Dabdoub A, Smallwood PM, Williams J, Woods C, Kelley MW, Jiang L, Tasman W, Zhang K, Nathans J. Vascular development in the retina and inner ear: control by Norrin and Frizzled-4, a high-affinity ligand-receptor pair. Cell 2004; 116:883-95. [PMID: 15035989]
16. Kondo H, Hayashi H, Oshima K, Tahira T, Hayashi K. Frizzled 4 gene (FZD4) mutations in patients with familial exudative vitreoretinopathy with variable expressivity. Br J Ophthalmol 2003; 87:1291-5. [PMID: 14507768]
17. Warden SM, Andreoli CM, Mukai S. The Wnt signaling pathway in familial exudative vitreoretinopathy and Norrie disease. Semin Ophthalmol 2007; 22:211-7. [PMID: 18097984]
18. Tasman W, Augsburger JJ, Shields JA, Caputo A, Annesley WH Jr. Familial exudative vitreoretinopathy. Trans Am Ophthalmol Soc 1981; 79:211-26. [PMID: 7342402]
19. Chen ZY, Battinelli EM, Fielder A, Bundey S, Sims K, Breakefield XO, Craig IW. A mutation in the Norrie disease gene (NDP) associated with X-linked familial exudative vitreoretinopathy. Nat Genet 1993; 5:180-3. [PMID: 8252044]
20. Clemons TE, Gillies MC, Chew EY, Bird AC, Peto T, Figueroa MJ, Harrington MW, MacTel Research Group. Baseline characteristics of participants in the natural history study of macular telangiectasia (MacTel) MacTel Project Report No. 2. Ophthalmic Epidemiol 2010; 17:66-73. [PMID: 20100102]
21. Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, Stockman HA, Borchardt JD, Gehrs KM, Smith RJ, Silvestri G, Russell SR, Klaver CC, Barbazetto I, Chang S, Yannuzzi LA, Barile GR, Merriam JC, Smith RT, Olsh AK, Bergeron J, Zernant J, Merriam JE, Gold B, Dean M, Allikmets R. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proc Natl Acad Sci USA 2005; 102:7227-32. [PMID: 15870199]
22. Gold B, Merriam JE, Zernant J, Hancox LS, Taiber AJ, Gehrs K, Cramer K, Neel J, Bergeron J, Barile GR, Smith RT, Hageman GS, Dean M, Allikmets R. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. Nat Genet 2006; 38:458-62. [PMID: 16518403]
23. Yannuzzi LA, Negrao S, Iida T, Carvalho C, Rodriguez-Coleman H, Slakter J, Freund KB, Sorenson J, Orlock D, Borodoker N. Retinal angiomatous proliferation in age-related macular degeneration. Retina 2001; 21:416-34. [PMID: 11642370]
24. Rajappa M, Saxena P, Kaur J. Ocular angiogenesis: mechanisms and recent advances in therapy. Adv Clin Chem 2010; 50:103-21. [PMID: 20521443]
25. Heidary G, Vanderveen D, Smith LE. Retinopathy of prematurity: current concepts in molecular pathogenesis. Semin Ophthalmol 2009; 24:77-81. [PMID: 19373690]
26. Bradley J, Ju M, Robinson GS. Combination therapy for the treatment of ocular neovascularization. Angiogenesis 2007; 10:141-8. [PMID: 17372853]
27. Politoa A, Napolitano MC, Bandello F, Chiodini RG. The role of optical coherence tomography (OCT) in the diagnosis and management of retinal angiomatous proliferation (RAP) in patients with age-related macular degeneration. Ann Acad Med Singapore 2006; 35:420-4. [PMID: 16865194]
28. Hendrickx M, Leyns L. Non-conventional Frizzled ligands and Wnt receptors. Dev Growth Differ 2008; 50:229-43. [PMID: 18366384]
29. Nikopoulos K, Venselaar H, Collin RW, Riveiro-Alvarez R, Boonstra FN, Hooymans JM, Mukhopadhyay A, Shears D, van Bers M, de Wijs IJ, van Essen AJ, Sijmons RH, Tilanus MA, van Nouhuys CE, Ayuso C, Hoefsloot LH, Cremers FP. Overview of the mutation spectrum in familial exudative vitreoretinopathy and Norrie disease with identification of 21 novel variants in FZD4, LRP5, and NDP. Hum Mutat 2010; 31:656-66. [PMID: 20340138]
30. Ells A, Guernsey DL, Wallace K, Zheng B, Vincer M, Allen A, Ingram A, DaSilva O, Siebert L, Sheidow T, Beis J, Robitaille JM. Severe retinopathy of prematurity associated with FZD4 mutations. Ophthalmic Genet 2010; 31:37-43. [PMID: 20141357]
31. Luhmann UF, Lin J, Acar N, Lammel S, Feil S, Grimm C, Seeliger MW, Hammes HP, Berger W. Role of the Norrie disease pseudoglioma gene in sprouting angiogenesis during development of the retinal vasculature. Invest Ophthalmol Vis Sci 2005; 46:3372-82. [PMID: 16123442]
32. Brod RD, Shields JA, Shields CL, Oberkircher OR, Sabol LJ. Unusual retinal and renal vascular lesions in the Klippel-Trenaunay-Weber syndrome. Retina 1992; 12:355-8. [PMID: 1336616]

Molecular Vision 2010; 16:2718-2726 <http://www.molvis.org/molvis/v16/a290> © 2010 Molecular Vision

33. Nambu H, Nambu R, Oshima Y, Hackett SF, Okoye G, Wiegand S, Yancopoulos G, Zack DJ, Campochiaro PA. Angiopoietin 1 inhibits ocular neovascularization and breakdown of the blood-retinal barrier. Gene Ther 2004; 11:865-73. [PMID: 15042118]
34. Fedi P, Bafico A, Nieto Soria A, Burgess WH, Miki T, Bottaro DP, Kraus MH, Aaronson SA. Isolation and biochemical characterization of the human Dkk-1 homologue, a novel inhibitor of mammalian Wnt signaling. J Biol Chem 1999; 274:19465-72. [PMID: 10383463]
35. Hughes JM, Groot AJ, van der Groep P, Sersansie R, Vooijs M, van Diest PJ, Van Noorden CJ, Schlingemann RO, Klaassen I. Active HIF-1 in the normal human retina. J Histochem Cytochem 2010; 58:247-54. [PMID: 19901273]
36. Kozulin P, Natoli R, O'Brien KM, Madigan MC, Provis JM. Differential expression of anti-angiogenic factors and guidance genes in the developing macula. Mol Vis 2009; 15:45-59. [PMID: 19145251]
37. Sun J, Hopkins BD, Tsujikawa K, Perruzzi C, Adini I, Swerlick R, Bornstein P, Lawler J, Benjamin LE. Thrombospondin-1 modulates VEGF-A-mediated Akt signaling and capillary survival in the developing retina. Am J Physiol Heart Circ Physiol 2009; 296:H1344-51. [PMID: 19304944]
38. Hackett SF, Ozaki H, Strauss RW, Wahlin K, Suri C, Maisonpierre P, Yancopoulos G, Campochiaro PA. Angiopoietin 2 expression in the retina: upregulation during physiologic and pathologic neovascularization. J Cell Physiol 2000; 184:275-84. [PMID: 10911358]
39. Kapitsinou PP, Haase VH. The VHL tumor suppressor and HIF: insights from genetic studies in mice. Cell Death Differ 2008; 15:650-9. [PMID: 18219317]
40. Roberts RL, Green J, Lewis B. Lutein and zeaxanthin in eye and skin health. Clin Dermatol 2009; 27:195-201. [PMID: 19168000]
41. Landrum JT, Bone RA. Lutein, zeaxanthin, and the macular pigment. Arch Biochem Biophys 2001; 385:28-40. [PMID: 11361022]
42. Loane E, Nolan JM, O'Donovan O, Bhosale P, Bernstein PS, Beatty S. Transport and retinal capture of lutein and zeaxanthin with reference to age-related macular degeneration. Surv Ophthalmol 2008; 53:68-81. [PMID: 18191658]
43. Duncan KG, Hosseini K, Bailey KR, Yang H, Lowe RJ, Matthes MT, Kane JP, LaVail MM, Schwartz DM, Duncan JL. Expression of reverse cholesterol transport proteins ATP-binding cassette A1 (ABCA1) and scavenger receptor BI (SR-BI) in the retina and retinal pigment epithelium. Br J Ophthalmol 2009; 93:1116-20. [PMID: 19304587]
44. Kunduzova O, Alet N, Delesque-Touchard N, Millet L, Castan-Laurell I, Muller C, Dray C, Schaeffer P, Herault JP, Savi P, Bono F, Valet P. Apelin/APJ signaling system: a potential link between adipose tissue and endothelial angiogenic processes. FASEB J 2008; 22:4146-53. [PMID: 18708591]
45. Anderson DH, Radeke MJ, Gallo NB, Chapin EA, Johnson PT, Curletti CR, Hancox LS, Hu J, Ebright JN, Malek G, Hauser MA, Rickman CB, Bok D, Hageman GS, Johnson LV. The pivotal role of the complement system in aging and age-related macular degeneration: hypothesis re-visited. Prog Retin Eye Res 2010; 29:95-112. [PMID: 19961953]
46. Carson-Walter EB, Hampton J, Shue E, Geynisman DM, Pillai PK, Sathanoori R, Madden SL, Hamilton RL, Walter KA. Plasmalemmal vesicle associated protein-1 is a novel marker implicated in brain tumor angiogenesis. Clin Cancer Res 2005; 11:7643-50. [PMID: 16278383]
47. Faurobert E, Albiges-Rizo C. Recent insights into cerebral cavernous malformations: a complex jigsaw puzzle under construction. FEBS J 2010; 277:1084-96. [PMID: 20096036]
48. Meyer-Schwickerath R, Pfeiffer A, Blum WF, Freyberger H, Klein M, Losche C, Rollmann R, Schatz H. Vitreous levels of the insulin-like growth factors I and II, and the insulin-like growth factor binding proteins 2 and 3, increase in neovascular eye disease. Studies in nondiabetic and diabetic subjects. J Clin Invest 1993; 92:2620-5. [PMID: 7504689]
49. Fort P, Estrada FJ, Bordais A, Mornet D, Sahel JA, Picaud S, Vargas HR, Coral-Vazquez RM, Rendon A. The sarcoglycan-sarcospan complex localization in mouse retina is independent from dystrophins. Neurosci Res 2005; 53:25-33. [PMID: 15993965]
50. Eichler W, Yafai Y, Wiedemann P, Reichenbach A. Angiogenesis-related factors derived from retinal glial (Muller) cells in hypoxia. Neuroreport 2004; 15:1633-7. [PMID: 15232297]
51. Tikellis G, Gillies MC, Guymer RH, McAllister IL, Shaw JE, Wong TY. Retinal vascular caliber and macular telangiectasia type 2. Ophthalmology 2009; 116:319-23. [PMID: 19187824]
52. Macaulay IC, Tijssen MR, Thijssen-Timmer DC, Gusnanto A, Steward M, Burns P, Langford CF, Ellis PD, Dudbridge F, Zwaginga JJ, Watkins NA, van der Schoot CE, Ouwehand WH. Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins. Blood 2007; 109:3260-9. [PMID: 17192395]
53. Mottl AK. Shoham, David A., North, Kari E. Angiotensin II type 1 receptor polymorphisms and susceptibility to hypertension: A HuGE review. Genet Med 2008; 10:15.
54. van der Veen RL, Fuijkschot J, Willemsen MA, Cruysberg JR, Berendschot TT, Theelen T. Patients with Sjogren-Larsson syndrome lack macular pigment. Ophthalmology 2010; 117:966-71. [PMID: 20163870]
55. Rizzo WB, Carney G. Sjogren-Larsson syndrome: diversity of mutations and polymorphisms in the fatty aldehyde dehydrogenase gene (ALDH3A2). Hum Mutat 2005; 26:1-10. [PMID: 15931689]
56. Loewen N, Chen J, Dudley VJ, Sarthy VP, Mathura JR Jr. Genomic response of hypoxic Muller cells involves the very low density lipoprotein receptor as part of an angiogenic network. Exp Eye Res 2009; 88:928-37. [PMID: 19233325]
57. MacDonald ML, Goldberg YP, Macfarlane J, Samuels ME, Trese MT, Shastry BS. Genetic variants of frizzled-4 gene in familial exudative vitreoretinopathy and advanced retinopathy of prematurity. Clin Genet 2005; 67:363-6. [PMID: 15733276]
58. Drenser KA, Dailey W, Vinekar A, Dalal K, Capone A Jr, Trese MT. Clinical presentation and genetic correlation of patients with mutations affecting the FZD4 gene. Arch Ophthalmol 2009; 127:1649-54. [PMID: 20008721]
59. Toomes C, Bottomley HM, Scott S, Mackey DA, Craig JE, Appukuttan B, Stout JT, Flaxel CJ, Zhang K, Black GC, Fryer

Molecular Vision 2010; 16:2718-2726 <http://www.molvis.org/molvis/v16/a290> © 2010 Molecular Vision

A, Downey LM, Inglehearn CF. Spectrum and frequency of FZD4 mutations in familial exudative vitreoretinopathy. Invest Ophthalmol Vis Sci 2004; 45:2083-90. [PMID: 15223780]

Appendix 1. Primer sequences of candidate genes sequenced (5' -3').

Gene	Exon	Forward primer sequence	Reverse primer sequence
AGGF1	1	ccgttcggctacaagtgagt	tggaattgtaactgcgctga
	2	agggaaacttgctgctcttg	ttcctccagaagaacaaattga
	3	tgtgtttaaatgccagtgttttg	tatgctgtgacggaccaaag
	4	caggtcttgagttcagccaac	gatcactccactgcactcca
	5	ttgaatgtatttggcattttcg	tgggtatgtttgaaatgcaga
	6	cgaattcgttgaaacttgatg	catttttaatggttttgctctttt
	7	ccttacccattgctgcgtat	cagagcacaactctgtctcaaaa
	8	cactgtgtgaaagcacaggaa	aaacagatctgcactgccttt
	9	aaggctcaagttaacatcatcatatc	cagcctttctggaaattcctt
	10	ggtgggtttgacaagttgct	tctaacctgccctggttcac
	11	tgctgaaaggacatcatcaca	caaatcacgaaatcatactacacaa
	12	cagcttcccaaaatgctagg	agccctgaacatcagaccac
	13	aatgccaagcggttttctaa	tgccacaaaggcagtaatca
	14	aagttggacacagcagatttca	tgtgctgctgattcttttgg
AGTRL1	1A	ttgcagagtgggtgacagag	tgggcgatgaagaagtaaca
	1B	gggacttggagaacaccact	ttcagaaagcaagggcaaag
AGTR1	1A	cgtttatgactgagaaatgaatgttt	tgcaggtgactttggctaca
	1B	tgtgtttctactcacgtgtctca	ttctccttcaattctgaaaagtagc
ALDH3A2	1	ctgcacctgcatgcttcc	aaacaaaacccagccgaag
	2	tctgacattcagggccaagt	agccactatccttaatttccaa
	3	agcctgttcttcccactgaa	agagctgggtgactccacac
	4	tgttgaagagattgctgatgtt	gccataaggtatgggttca
	5	tgcaagacattcaaacaacaca	caaccttgctcccagcata
	6	gcatggctggattttgtactt	ggtacagtataccagaaagcagga
	7	tatgacttgggtgggagagg	cagcaaagaaggaacatgactc
	8	aactgagcacacagccctct	acatttgccctatgctggag
	9	tttcccggtcgttgttagac	atatgcatctggcagccatc
	10	ctgggcaatagagcgagact	caacttggcccaccttacat
	11	aaccgtgcttcctaggcttc	ggaatgttgggacgtagttga
ANG1	1	ggggaaagagtcaaacaaa	caaatgctcccctaaggaac
	2	ccatttctccaactcccaga	tcacagtgctgaaatgtgctc
	3	atgtttgccaggatggtctc	ttggcagagaggtgaaggat
	4	cacccaaggaaaccaagaaa	cagcaatcaatgcactgtagaa
	5	tggaggaaatgaagacatgc	tcctttatggacctgggaaa
	6	gcagacctgttcgccttatt	acgcacacgtggaaatcat
	7	tgaaatgtcaatcactggagaaa	caattattaagtggctaggtaaaagg
	8	caagctggtcttctgggttc	aattcgtgggcagaacaaaa
	9	gaacaagcaagtttcccatca	tttctcacctggcagcttct
APLN	1	ccaaggagcagcatgaatct	gcgctctgtcaatctgctg
	2	cgtttccaacaaccttcctc	gtccttctccctccgcttta
CCM2	1A	tgtttggattttgccaacag	tagctcacaatgggcaagc
	1B	ccgggtcgagcatgtagc	gaaagcccacctgaccatt

	2	gggggccatggtagtagttt	ggtgtttgtgtggctgtgtc
	3	tagtaggagatggggcagca	actggctcattcctcacagc
	4	ttggtgtccctggaataagc	caacacgaagctgcaaaaac
	5	tgagccaggtaaaggtcctc	gcaatggcagaggaagtagc
	6	gcagagggacacattctgg	acaaaaggccaccatgtca
	7	gcacacagcacattccagag	gcctcagccattcactaagc
	8	acacacggcatggactaggt	ccaggaaaaggcagtgatgt
	9	ttagtggctgtggcaaggt	ctcatgccatctctggttca
	10	ccaggatgctagatgcagtg	ccctgctccttcacagtaga
CFB	1	cgggaaagtgatgtgggttag	aaaggccaaggagggatg
	2	accttcccttcctgctgtct	caccctgcctagtctcatcc
	3	cgagaccaggagggatacac	agtgtccggagccgagtg
	4	tgctctctaccttgctcacg	gaggtggggagagaagacagt
	5	atgatcccccgtctctttg	tcaggctccagcattaacag
	6	actggggggaaaatggagaag	cagggagctagtcctggaag
	7	gcaaagtcgctgaaatctcc	gcaaagtcgctgaaatctcc
	8	tgtgggatttcagttgcaga	ggacttaagggccacatgct
	9	ccatgtgtatccctgccttt	aggaaggatgagggtccaag
	10	tctgggtgagtaacctgccta	agaaggctttccaggcaact
	11	gcctttgctccccatagact	tcccatttcaccttgacctc
	12	atgggagtaggggaaggaaa	ggaggtgttgcctggagag
	13/14	aatgacacggggccagag	cccacttgtctcttgggatct
	15	agaagggcttaggggacatc	gaaggattggggcctagagt
	16/17	ctaggccccaatccttccta	gcattgagctttcctgcttt
	18	ccatgcttccaggattagga	cagtgcctgttccccact
DKK1	1	ctccctgcagtcaggactct	taccaaatgctccctgaacc
	2	gggcagtaacaggttttgga	atagacgctcaaaggctgga
	3	ggcataacagactgccactg	cacagtctgatgaccggaga
	4	tccagaagaaccaccttgtct	ccttaggattgagttgatgaagg
FZD4	1	gtgcaaactgggggtgtct	tagtaggatcgggcttcctc
	2A	actcagctttgtgggagcat	acacaaagtagaggtgacg
	2B	aggagcctgaactgtgtgct	ataaaggtgtaacgtcggt
	2C	tttggcagcaggactcaaat	ggacaataaaaatactttagaggtt
	2D	agttcctgcaacgtgtgtga	gtcgtaaaacctcctttaaga
GSTP1	1	cgggaccctccagaagag	cctgatgctgcgggttgg
	2	gccttggcatcctcccc	gtaaaagccacacgacggag
	3	gaaatcttcggaggaacctg	tctccgtcctggaacttgg
	4	cttagggggctgtgactagg	ggtctcagtttcctccactccc
	5	tccccagtgactgtgtgttg	agcccctttctttgttcagc
	6	gctgggagggatgagagtag	aaacaaatggctcacacctg
	7	acacaggtgtgaggcatttg	ctggagaaaggaaggcaaac
HIF1A	1	tggacttgcctttccttctc	ggcaagcagggctcagac
	2	ccttctgtgataagcagaaatgtaa	gaggtatacaaaatcaaaacattgc
	3/4	gcgagaaaactttgtaaaaacatc	tgggtaagtacaatagcaaagtttaat

OXGR1	1A	tcatattgccaaactgaactct	ggctgatctgttggtcctgt
	1B	ccaatgagctgcttttccat	ttccattgagggagacatcc
PEDF	1	ggtgggggaaagtgactagc	ctctgttttgctgctggaga
	2	ccctgaactcaaacccaaga	ccctgagctcctgcttacac
	3	cctacttgggctctcagcag	agcgagactccgtctcaaaa
	4	ctcaaagacgggatgcttgt	gcaccactgcactccagtta
	5	ttttcaacaacgtgctctgg	ttcccactaccctgttttgc
	6	agctcctggctgtgtctgtc	tgcgttctgcttagcacagt
	7	atcccttggttggggtgt	ttcgtgtcctgtggaatctg
PLVAP	1	ggatcgagcaaatgggtct	agcattggtcctgtgatcgt
	2	gaccatgcccttgtgaactc	accaggagacccaggaagat
	3A	gtggactccctgcccagt	caggggatggtagaggttgt
	3B	aggagcaactgcaaaaggtg	accctcttggcatctggac
	4/5	agaagcctgtctgggaagc	caccctcagtcctggtccta
	6	ggtccccaggatgttgtgt	caccacggtgatatgtgacg
SCARB1	1	cataaaaccactggccacct	cctttagggcgggtcagg
	2	gccctcatcactctcctcac	ccgactatgacttgcctgct
	3	aggcgagtagaagggaaacg	cacctcagggactgctctct
	4	atttcccttcatgggctctt	gaggggaagacaggacacag
	5	cagcccagaatgttcagacc	agtgttcatcctcccagcac
	6	cagcacctgagagggcttat	tttggtggtgtgaccaaaga
	7	gggagaagtggctggatct	cagacagcactgggcagata
	8	tggggctttttacagaatgg	gtgctccaaccaggaatcac
	9	agcctgtggcttgttttagg	cactggagtctgggaccact
	10	cctggggacagttttgtcat	tccatgtcacagtccggtta
	11	aagcttgaggtggggaaaag	cagagggtccgatctgagg
	12	tcacttctgacagcctggtg	ccctacaagtcccttcagca
	13	atcgttgagggttgttggac	taagccgccagcattctact
SSPN	1	agtcggctccaaatgttttc	gagcagagcgcgagttaga
	2	ccaccagttgaaagaaaaaca	gtctggatcaatgccctctt
	3	gttttgatgaattcgctttgc	tttaacctcagctactccatgtg
SUCNR1	1	ggcaatatggaaacaagtaaaagc	atggagggtggcagtgac
	2A	ttcaccattcaagtgtttttca	aggtggtgccattgtcagtt
	2B	tgccaacctctataccagca	ttgtcgtcccattcttttca
TGFB2	1	ctttgagaattgttgatttcttttt	agctgctgcggttttgag
	2	cccaggtagcagggacctta	tctaggcacctttgtgatcg
	3	ttaaactggccgttggaaac	agctgtgatcatgccactgt
	4	gaccatgcaattgagatgaca	caaggaggtcatttgtatgtgtg
	5	ttttcctcccaagatgttcag	tccaatcctttctcccttca
	5	taactgttgccagctgatgc	agaaagggcaaaggcagttt
	6	tgagggtggtgaatcagctt	tgcgaatacttttgcaccaa
	8	ttgcctactcagtgctgtgac	cactgatgaaccaaggctct
THBS1	1	ccagcatctctttcctccac	caagagggagaggacagcac
	2	agccccctactgctggtc	ggatttcctgtagccctcct

	3	tttggcgattaatgttgctg	ccccttggagtcacttctgt
	4/5	acccctctacctgcatcctt	tcaaacaaaagccgtcgatt
	6	gacagccctgaccatgtttt	tcaaacaaaagccgtcgatt
	7	ccctgaacagagccctctaa	tgcagcaagaaactgaatgg
	8	tgatgggagcctctatttgg	ttggtgacgacaacaaggac
	9	tcttgaacgggcttaggaga	gcaggactttgcttctgtttg
	10/11	gcagtatggcagcttagacca	ggacagctctcctgtcctct
	12	ttccaaatggagcctctgtc	ataaggtgatgggccctctt
	13	aggtacactccctcgtgcat	tcttggcaacatctacatcctt
	14/15	ggcttgagctgttttcaagg	caaaggccatgaaagggata
	16	ctgtggtgggggcataagt	caaatgctttgagccttggt
	17	gacaggatgaagggaccaaa	tttgtttttccccatgtatgc
	18	gggtgctgaggatgtctagg	ctggtgatgctgggaacttt
	19/20	gcattgtttctgatggaatgaa	cccataaccgtcatccttga
	21	tgctaatgtgcttgaacaacct	attcattggaggctgaatgc
TIE2	1	ctccctgtgctcagacagaa	ttctaacacaggggccaatc
	2	aggctttgtgagcaccagtt	gctgccaagacaaaaggtgt
	3	ggaagtgaccatccctttca	gtaaccaaatatgtcacctaagagg
	4	tcattcttctcccacatcca	attcatttgctttgcagcat
	5	ttcaccattgtccactgaatg	tctgacctaacagaggccattt
	6	ctctgtggggcagtttcatt	tgacagcaaactggatctcct
	7	ccccagctttttaagttcctg	gccggcacaagaattacact
	8	cccggtgcatgacttactaaa	ggcattttatcaaagccaaca
	9	atggacctttgcgtttatgc	atcctggaaattaccccaaa
	10	gcaaaagcaaagcagagagc	aattctcctggtgccacatc
	11	cacatcgcaataacaacaacc	ctgtcactttcccacccatt
	12	tggggacactaatccaaacta	cggtagccccttattgct
	13	ccaattgattggggtaccat	gctgcctatagggctgcac
	14	acggtgtgggtctgtttctc	gaggcttgcctaagggaaat
	15	ggatgccaaccagaagacat	gaggaaagaatgagcaggaaa
	16	tggtgactgagggtagctga	aaaacaaggcaaaccacagc
	17	cctgggtggtgttgctagat	ctttagagggaactccacagg
	18	cctgttccccaaagttttca	tttcccagggcacacagtat
	19	gctgggacatacacaaagca	tgattctgcactcctctgga
	20	gtgtgcaagggcctatccta	catacaaacggcatcctgtg
	21	tcaccctctcttgccatacc	aaggggaggggttcctattt
	22	tcctaggggctgtactttgg	tggagggacatcctttgttt
	23	tttctcagcaggtgtcaagg	tttctcagcaggtgtcaagg
VHL	1	cgcgaagactacggaggt	cttcagaccgtgctatcgtc
	2	ggacggtcttgatctcctga	tggataacgtgcctgacatc
	3	gttggcaaagcctcttgttc	aaggaaggaaccagtcctgt
VLDLR	1	caactccttcccctccttct	gtgggcaaacggagacctac
	2	tccccatccatgggtattag	tgataaccccacgtcaaaca
	3	agtgcccattgactcagctt	aatgatttgactgcccttgg

	4	cacaggtattagcgcttcca	tcagtgcgtcttcaggctta
	5A	tgaacggaccaatcttgatg	cagagccgcactggattt
	5B	agtgatgagctggactgtgc	ggaggaaacaccatggaatg
	6/7	tcaactgggacaatttgctg	tggattgtgtcaaacttccag
	8	gacaaatcgtgggtttgacc	tctgggagggaaatcagaat
	9	gcaggtgatgggaaaggata	gaggttagcagcattgttctca
	10	tgggaggaggtggtttagaa	cccacagatcacttcccaag
	11	caaaaagtccattctccaagc	gaaagctcctgacctacacaga
	12	gccttgagttttctgctcaa	ggcggaagcagtagagtcat
	13	ggggaaagaaccgtgaaagt	ctcctcccgctatgtctgtt
	14	tgtcccagttcagcattcag	ttctgtgtcatgctgcttcc
	15	ggcaaggactcaggtcttca	cccggcatacaatagcagat
	16	acagctagccatgctggaac	ccaattggggcagaaatagt
	17	tttggctccttacctgatgg	catgaaaggggtagctccag
	18	tggattcctgaacgttattacc	tcacccaggtctcctttctg
	19	caactcaaaagcaaggtcca	agtgcatatttggccagagg

Summary of Relevance

This work represented the initial, hypothesis driven, effort to identify a causative allele for MacTel. While candidate gene screening has been successful in a limited number of cases, the odds of finding a risk allele by this method are extremely slim. Constructing a hypothesis for a role of any given gene in the pathology of a disease depends on prior knowledge of gene function, and the intracacies of networks of interacting gene products and signaling cascades. Our known universe in regard to molecular genetics is limited, whereas the true universe is vast. Even with this limitation, exhaustively testing the hypotheses that it is possible to construct based on reports of gene function in the literature is an unreasonable undertaking with low-throughput methods. That said, there were compelling candidate genes that we were able to rule out as harboring causative variants in coding regions by this method. At the time this work was carried out, the study cohort was not yet large enough to have any power to detect causative alleles through linkage, much less genome-wide association. This undertaking enabled me to construct criteria to evaluate variants by examining population allele frequencies from public databases, cosegregation of an allele with the disease in families, and allele frequencies in cases and controls in our cohorts. I later applied these criteria in evaluating variants in the linkage interval that was subsequently identified, as well as in defining an approach to evaluate variants indentified in whole exome sequencing.

INTRODUCTION TO CHAPTER FOUR

The work described in this chapter has been submitted for publication and accepted pending minor revisions. This chapter describes the first genome-wide linkage scan performed in MacTel families with multiple affected relatives. This work resulted in the identification of a region of statistically significant linkage on chromosome 1. Allele sharing identical by descent was determined in affected sibling pairs in an attempt to narrow the region of interest, and in preparation for whole exome sequencing, to define regions that could be excluded by virtue of no alleles shared identical by descent. The exons of genes in the region flanking the maximum LOD score were sequenced during a long delay pending the results from whole exome sequencing and while additional families were being recruited for subsequent linkage analysis.

Chapter Four – Macular telangiectasia type 2: Linkage to 1q41-42 and analysis of the locus

Abstract

Macular Telangiectasia type 2 (MacTel) is a relatively rare macular disease of adult onset presenting with distortions in the visual field and leading to progressive reduction in central visual acuity and loss of visual acuity. Several pedigrees were identified with multiple affected family members. Inheritance was consistent with autosomal dominant segregation with reduced penetrance. Genome-wide linkage analysis, analysis of recombination breakpoints to identify alleles shared identical by descent, and Sanger sequencing of positional candidate genes was performed to identify the genetic cause of MacTel. Seventeen families with a total of 71 individuals (45 affected or possibly affected) were enrolled at clinical centers in 7 countries under the auspices of the MacTel Project. Genome-wide linkage analysis identified a single peak with multi-point LOD score of 3.45 on chromosome 1 at 1q41-42 under a dominant model. Recombination mapping identifying chromosomal segments shared by all affected individuals in a subset of families defined a 15.6 Mb candidate region encompassing the 1q41-42 linkage peak . Sequencing of 14 genes flanking the maximum LOD score revealed no obvious disease-associated variant in these pedigrees.

INTRODUCTION

Macular telangiectasia is a group of diseases characterized by Gass and Blodi in 1993 (Gass and Blodi, 1993) and recharacterized by Yannuzzi in 2006 (Iturralde et al., 2006). Macular telangiectasia type 2 (MacTel) generally presents bilaterally between the 5th and 7th decades of life with reduction in central vision and distortion in the visual field. The cause of the disease is unknown and there is no treatment.

Clinical characteristics of MacTel include loss of retinal transparency, autofluorescence changes in the macula, macular edema, presence of intraretinal crystals, and disruption of macular pigment transport. Symptoms of advanced disease include the presence of a macular hole, dilated and tortuous vessels in the perifoveal region, leakage from retinal vessels and neovascularization arising from the intraretinal vessels (Chew et al., 2006; Charbel Issa et al., 2007; Hannan et al., 2007; Charbel Issa et al., 2008a; Charbel Issa et al., 2008b; Charbel Issa et al., 2009b; Charbel Issa et al., 2009c; Finger et al., 2009; Aung et al., 2010). Patients experience distortions in central vision, including parafoveal scotoma, and metamorphopsia. Both genders are affected equally

While MacTel had been presumed to be a very rare disease, recent epidemiological studies suggest that it is under-diagnosed and, therefore, more common than previously thought The Beaver Dam Eye Study recently reported a prevalence of 0.1% in a retrospective study of 4,790 individuals, aged 43-86 years of age (Klein et al., 2010). The Melbourne Collaborative Cohort

estimated a probable prevalence of 0.0045% based on evaluation of 3,784 images where macular disease was noted, out of a study population of 22,415 participants (Aung et al., 2010). Both studies used available population data where retinal images had been obtained to assess other macular diseases in populations. In both studies, however, images had not been taken with the intent to diagnose MacTel; therefore both studies concluded that subtle features of MacTel were likely missed without specialized imaging, such as fluorescein angiography and blue light reflectance imaging.

MacTel was proposed to have a genetic component based on case reports of affected sibling pairs and concordant monozygotic twins (Hutton et al., 1978; Chew et al., 1986; Oh and Park, 1999; Menchini et al., 2000; Siddiqui and Fekrat, 2005; Hannan et al., 2007; Gillies et al., 2009). To test the hypothesis that MacTel is an inherited disease, family members of probands were actively recruited and given full ophthalmic examinations. Gillies et al. (Gillies et al., 2009) have previously reported four multiplex families included in this study. Additional multiplex families were subsequently identified, strengthening the hypothesis that variants in one or more genes underlie in the etiology of MacTel.

The MacTel Project was established as a consortium of researchers and clinicians in order to study the natural history, identify the cause(s) of the disease, and propose targets for treatment. Patients were screened and enrolled at 23 clinical centers in seven countries (Australia, Germany, France, the U.K., Israel, Switzerland, and the United States). Family

members were actively recruited and given complete ophthalmic examinations. Seventeen multiplex families were identified that were informative for linkage analysis, together with additional parent-child duos Altogether, these data provided a basis for genome-wide linkage and recombination mapping analyses followed by sequencing of selected candidate genes from identified locus to determine the genetic cause of MacTel.

Results

Study population

Seventeen families with a total of 71 individuals (45 affected or possibly affected) were analyzed for linkage. The inheritance pattern in families with more than one affected individual was consistent with autosomal dominant transmission. MacTel exhibits reduced penetrance based on the observation that in some multiplex families neither parent is clearly affected with the disease. Variable disease expressivity is evident in many pedigrees in this cohort; while probands presented to the clinic experiencing vision loss, some relatives were given a diagnosis of MacTel only after a complete ophthalmic exam as a part of this study. Retinal images from all recruitment centers (fundus photos, AF, OCT, etc.) were evaluated by a central reading center. The adjudicator was not informed of the diagnosis assigned by the recruitment center or whether the subject was a enrolled as a proband, relative, or control. All subjects were categorized as definitely affected, possibly affected, probably not affected, or definitely not affected. This clearly illustrates the variable expressivity of MacTel, complicating genetic analysis. No gender bias was observed in patients, and male to female and female to male transmissions were both observed in the pedigrees. Most families are too small to make a reliable estimation of the ratio of affected offspring, however, four large families had ratios of affected offspring consistent with autosomal dominant inheritance. Family 1 (Figure 6A) had four offspring, with two offspring affected, and one affected parent. Family 2 (Figure 6B) had seven offspring; two were affected, two were probably not affected, three were unaffected, and one parent was affected. Family 3 (Figure 6C) had six offspring; two were affected, one was

possibly affected, two were unaffected, and one was probably not affected. The parents in this family were unavailable for examination. Family 4 (Figure 6D) had ten offspring and two cousins in the third generation. Six individuals in the third generation were examined; three were affected. The parents and grandparents in this family were unavailable for examination, but an uncle in the second generation was affected.

Multipoint linkage analysis

A total of 112 individuals in 33 MacTel families were screened on Illumina 1M Duo arrays. Seventeen informative families were analyzed by multi-point, affected-only parametric linkage analysis under an autosomal dominant model using a subset of SNPs from the Illumina 1M chip. Family members diagnosed as possibly affected were coded as affected. One significant peak was observed on chromosome 1, with a LOD score of 3.45 and HLOD of 3.54 (alpha=0.93) over an interval of approximately 15 Mb (Figures 7 and 8). Only one other region, on chromosome 5, yielded a positive LOD score over an interval of approximately 3.4 Mb (LOD=1.52, HLOD=2.43, alpha=0.76). Ten regions, totaling 13.2 Mb, had negative LOD scores between -2 and 0—not sufficiently negative for exclusion (Table 5.). The remainder of the genome yielded LOD scores below -2, sufficiently negative for exclusion. One large family (family 1) was split into two smaller families because of the large number of missing family members between the two branches. All families were linked to the peak on chromosome 1, with the exception of one branch of this large family. The individuals coded as affected in the unlinked branch all had a diagnosis of "possibly affected".

Recomination mapping by determination of identical by descent sharing

IBD status was inferred along each chromosome with MERLIN (Abecasis et al., 2002), and only those positions where all affected individuals within a pedigree shared the same chromosomal segment were deemed consistent with the hypothesis of a necessary rare variant being located in a given genomic region. The full genome was analyzed for one trio of siblings and seven affected sib pairs, of which two had both parents (one affected in each family) genotyped. The results from the 8 families included in this analysis were then combined and the full genome was analyzed to map regions where exclusion was declared for at least one family; these regions were marked as excluded (Figure 9). Twelve chromosomes (chromosomes 3, 5, 6, 8, 10, 13, 15, 18, 19, 20, 21, and 22) were entirely excluded (Table 5). Chromosomes 1, 2, 4, 7, 9, 11, 12, 14, and 16 were partially excluded. In total, 153.81 Mb remained as potentially able to harbor a causal allele for MacTel under the assumption that a single copy of a single variant was necessary (but not sufficient) for disease (Table 5).

Comparison of identical by descent to linkage exclusion

Comparison of regions that were not excluded by either linkage analysis or by recombination breakpoint analysis revealed four regions that were not excluded by either analysis, including the significant linkage peak on chromosome 1. Three additional regions with LOD scores between 0 and -2 under autosomal dominant linkage analysis were not excluded by breakpoint analysis in strictly affected siblings: chromosome 7, 125.92-145.41 Mb; chromosome 12, 5.38-7.03 Mb; and chromosome 14, 102.52-106.38 Mb (positions based on recombination analysis).

Reviewing regions of exclusion in these two different ways serves to clarify whether a chromosomal segment is excluded based on information from definitely affected individuals or possibly affected family members where the phenotype is less strictly defined. A total of 32.378 Mb were not excluded by parametric linkage analysis, including the two regions of positive linkage. A total of 153.81 Mb were not excluded by the more stringent recombination breakpoint analysis based only on definitely affected siblings.

Gene sequencing

A 1.8 Mb region of the 15.6 Mb chromosome 1 linkage interval from 221,168,406-222,994,872, corresponding to the maximum LOD score, was selected for sequencing, which included thirteen genes and one micro-RNA. All exons and flanking intronic regions of *DISP1, TLR5, SUSD4, BEND5, CAPN8, CAPN2, TP53BP2, FBXO28, DEGS1, NVL, CNIH4, WDR26,* and *CNIH3* genes and the micro-RNA, *MIR320B2* were screened by Sanger sequencing in two affected family members, one from family 1 (individual IV4), the other from family 2 (individual III2). Each of these families consists of an affected sib pair, unaffected siblings, and two parents, one of which is affected in each family. Altogether, only sixteen variants (confirmed by bi-directional sequencing) were detected in coding regions of the 13 genes (Table 6) Six of these were synonymous variants; of the 10 non-synonymous variants, 5 were frequent polymorphisms, 2 were known variants of low or undetermined minor allele frequency that failed to segregate with the disease, and 2 were previously unknown missense variants that failed to segregate with the disease. One known variant with MAF=0.033, p.Val404Ile in NVL,

was detected in only one family, in which it is present in all affected family members, one possibly affected sibling, and no unaffected family members. However, this variant is more frequent than would be expected given the prevalence of MacTel and, therefore, not deemed disease-associated.

Discussion

This study presents the first genome-wide linkage analysis of MacTel. The identification of multiplex families with the disease supports the hypothesis of a genetic component to the disease. Previously, we investigated candidate genes by Sanger sequencing in the same cohort; however, no variants associated with MacTel were detected (Parmalee et al., 2010).

We examined 17 multiplex families in a genome-wide linkage scan and detected significant linkage to 1q41-42, with a multipoint LOD score of 3.45. Assuming an autosomal dominant pattern of inheritance, and analyzing only affected and possibly affected individuals, all families were linked to this locus, with the exception of one branch of a large family that was analyzed as a separate family. Two relatives in the unlinked branch of family 1 were diagnosed as "possibly affected." One possibility is that this family is segregating a risk allele at a different locus; another possibility is that the two relatives are not actually affected with the disease. The detection of a significant linkage peak provides the first evidence of a possible susceptibility locus for this disease.

Analysis of recombination breakpoints in strictly affected siblings defined chromosomal segments that were incompatible with inheritance of a rare disease allele. The results from individual families were combined to assess which regions across the genome are not excluded in any family, as would be expected if MacTel is a monogenic disease, caused by variants in the same region in all families. Two regions were identified on chromosome 1 that were compatible

with monogenic, autosomal dominant inheritance: one region corresponds to the region of the linkage peak; the other, smaller region is 5.3 Mb centromeric to the boundary of the linkage region (Figure 9). A second region of positive linkage on chromosome 5, with alpha=0.76, yielded LOD scores below the threshold for significance. This region was excluded in one family based on the observation of 0 alleles shared between affected siblings. This analysis was limited to affected sib pairs and one sibling trio that were diagnosed as definitively affected. Family members diagnosed as "possibly affected" were excluded from these analyses.

MacTel was previously believed to be a disease with no discernable pattern of inheritance; however, in many families identified in this study, the disease appears to segregate as a monogenic, dominant trait. This does not preclude the possibility of genetic heterogeneity, but rather provides a starting point for genetic dissection of the trait in families. Intensive efforts have been undertaken by collaborators within the MacTel project to refine the definition of the phenotype and gain insight into the progression of MacTel. It is unknown at this point whether family members diagnosed as "possibly affected" are in the early stages of the disease and will eventually manifest full signs of MacTel, or whether these individuals carry modifiers that lessen the expression of the phenotype. It is noted that the median age of relatives diagnosed as possibly affected is younger than those relatives with a definite diagnosis. For the linkage analysis part of this study, we categorized "possibly affected" relatives as affected, given that, in the view of the Reading Center adjudicator, they exhibit signs of the diseases that are not typically seen in the unaffected population. While the ratios of affected to unaffected individuals in large families correspond well to the expected ratios for a Mendelian trait with

autosomal dominant inheritance, most families in the cohort are not large and, in most cases, parents are deceased or unavailable for screening due to the late age of onset of the disease. The observation that in some pedigrees parents of affected offspring are unaffected has suggested that MacTel is not fully penetrant. Whether this is due to locus heterogeneity, the presence of modifying alleles segregating in families, or environmental influences acting on an underlying genetic predisposition is unknown at this time. The chromosome 1 linkage region with a significant LOD score is the primary region of interest for a causative variant for MacTel. The second region of positive linkage, on chromosome 5, was excluded in the combined IBD analysis of affected siblings. Future work will include complete sequencing of the linkage region on chromosome 1,as well as other regions not excluded by IBD analyses, and analysis of additional families as they become available.

Materials and Methods

Study cohort

Patients, relatives, and controls were recruited at 23 participating clinical centers in 7 countries (Australia, Germany, France, The United Kingdom, Switzerland, Israel, and the United States). Informed consent was obtained at each participating clinical center in accordance with ethics protocols for human subjects approved by the appropriate governing body at each site in accordance with the Declaration of Helsinki. All protocols and records of consent were centrally managed by the EMMES Corporation (Rockville, Maryland).

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. (Clemons et al., 2010) based on Gass and Blodi (Gass and Blodi, 1993). 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. Each sample was assigned to one of four diagnostic categories: affected, possibly affected, probably not affected, or unaffected. Participants are re-evaluated at regular intervals over the course of the study.

Peripheral venous blood was drawn from each participant and used to isolate DNA (Qiagen blood maxi kit 51194, Valencia, CA). DNA concentration was determined using the NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE). DNA samples of low purity were subjected to column purification (Qiagen blood and tissue kit 69504).

Genotyping and marker selection

Samples were genotyped on the Illumina 1M chip. The linkage marker set was selected by pruning stringently for genotype quality in GenomeStudio using the following parameters: GenTrain threshold ≥ 0.50, cluster separation ≥ 0.16, number of no calls = 0. The remaining markers were pruned for LD in PLINK (Purcell; Purcell et al., 2007) to remove markers with r2 greater than 0.17. The final marker set consisted of 11,676 independent markers. An identity by state-based relatedness analysis was performed in PLINK using genome-wide marker sets to confirm family structures.

Linkage analysis

Parametric multipoint linkage analysis was carried out using MERLIN (Abecasis et al., 2002), with risk allele frequency of 0.001, penetrance of 0.9 and phenocopy rate of 0 under the assumption that for a rare disease, affected individuals in a multiplex family are most likely affected for the same genetic reason. Affected only analysis was performed, with unaffected individuals coded as unknown. Possibly affected family members were coded as affected, under the rationale that they exhibited specific phenotypic signs consistent with MacTel, though not

sufficient for a definitive clinical diagnosis. Allele frequencies were estimated from a total of 112 individuals. Heterogeneity LOD scores and percentage of families linked to a locus (alpha) were calculated by MERLIN. As a confirmatory analysis that is less sensitive to model misspecification, we also performed a non-parametric linkage (NPL) analysis using MERLIN.

Recombinations were mapped by determining IBD allele sharing using the NPL algorithm in MERLIN. Only definitively affected sib pairs and trios were included in the initial analysis to mitigate the possibility of excluding chromosomal regions based on incorrect diagnoses. Allele sharing was determined by comparing NPL scores from siblings to scores from a simulated dataset, to determine the values associated with each allele sharing state. Where parental genotypes were missing, segments with NPL scores of -0.3 (0 alleles shared) were excluded. Where parents were genotyped and one parent was affected, segments with NPL scores of -0.3 (0 alleles shared) and -0.12 (1 allele shared from the unaffected parent) were excluded. Chromosomal segments with intermediate values were classified as ambiguous. The results of this analysis were aggregated to compile a genome-wide map of chromosomal segments where at least one allele was shared IBD in siblings in all families. This result was compared to parametric linkage results from the entire cohort to examine allele sharing in the linkage interval on chromosome 1, to validate exclusion by negative parametric LOD score, and to search for regions that could be prioritized for gene screening by Sanger sequencing.

Candidate gene sequencing

Genomic DNA from two unrelated affected individuals was amplified using primers specific for exons in the genes of interest. PCR amplification was carried out using 2 ng of genomic DNA in a total volume of 25 µg, with 25 pmol each of forward and reverse primer, 200 µM dATP, dCTP, dGTP, dTTP, 2.5 mM $MgCl_2$, 1.5 U Taq Polymerase (Hot Fire DNA Polymerase, Solis Biodyne, Tartu, Estonia, or AmpliTaq Gold, Applied Biosystems, Carlsbad, CA), and 10× buffer supplied by the manufacturer. Thermocycling was performed using either the Stepdown protocol, or the Touchdown (68–55 °C) protocol. Stepdown: an initial 12 min denaturation step at 95 °C was followed by 12 cycles of 95 °C for 12 s, 65 °C for 20 s (with a 0.5 °C reduction in temperature for each cycle), and 72 °C for 55 s. This was followed by 30 cycles of 95 °C for 12 s, 50 °C for 20 s, and 72 °C for 55 s, with a final 7 min extension at 72 °C. Touchdown (68–55 °C): an initial 12 min denaturation step at 95 °C was followed by 26 cycles of 95 °C for 15 s, 68 °C for 20 s (with a 0.5 °C reduction in temperature for each cycle), and 72 °C for 45 s. This was followed by 15 cycles of 95 °C for 15 s, 55 °C for 20 s, and 72 °C for 45 s, with a final 7 min extension at 72 °C. Sequencing was performed by Genewiz (South Plainfield, NJ). Primers were selected using Primer3 (http://frodo.wi.mit.edu/primer3/). Sequences were compared to the hg19 reference sequence.

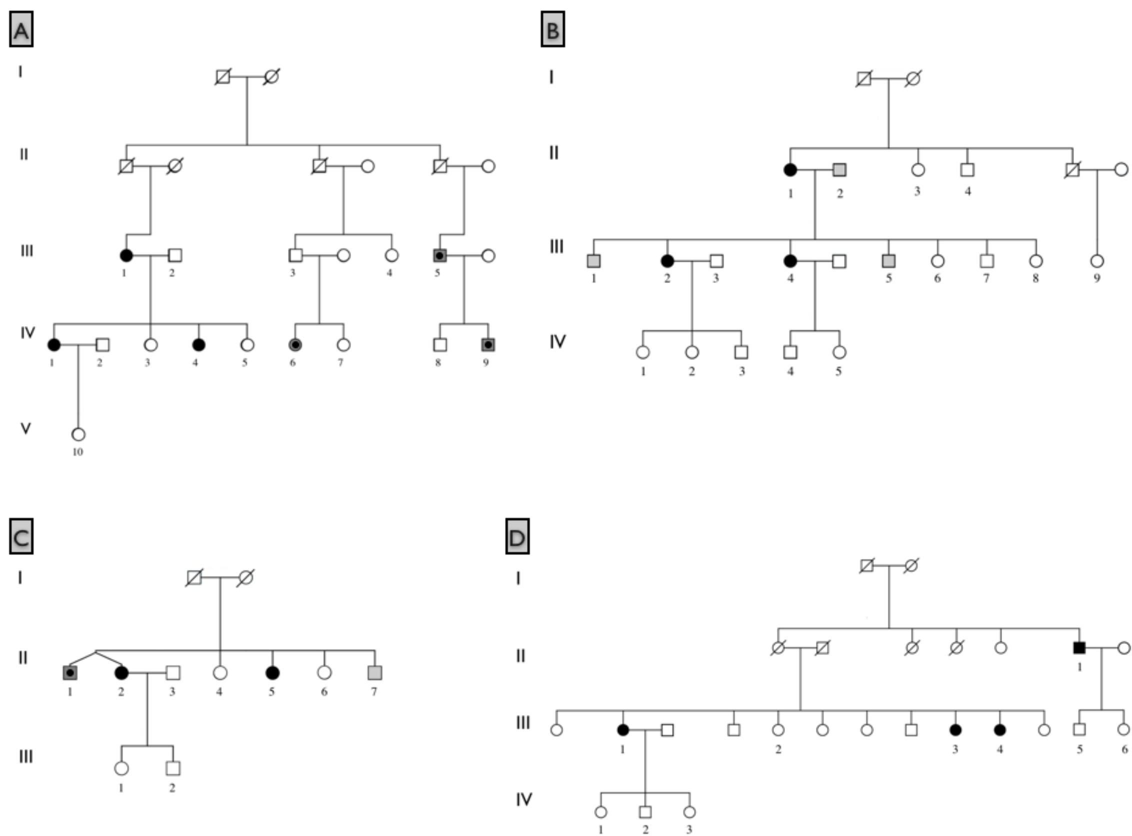
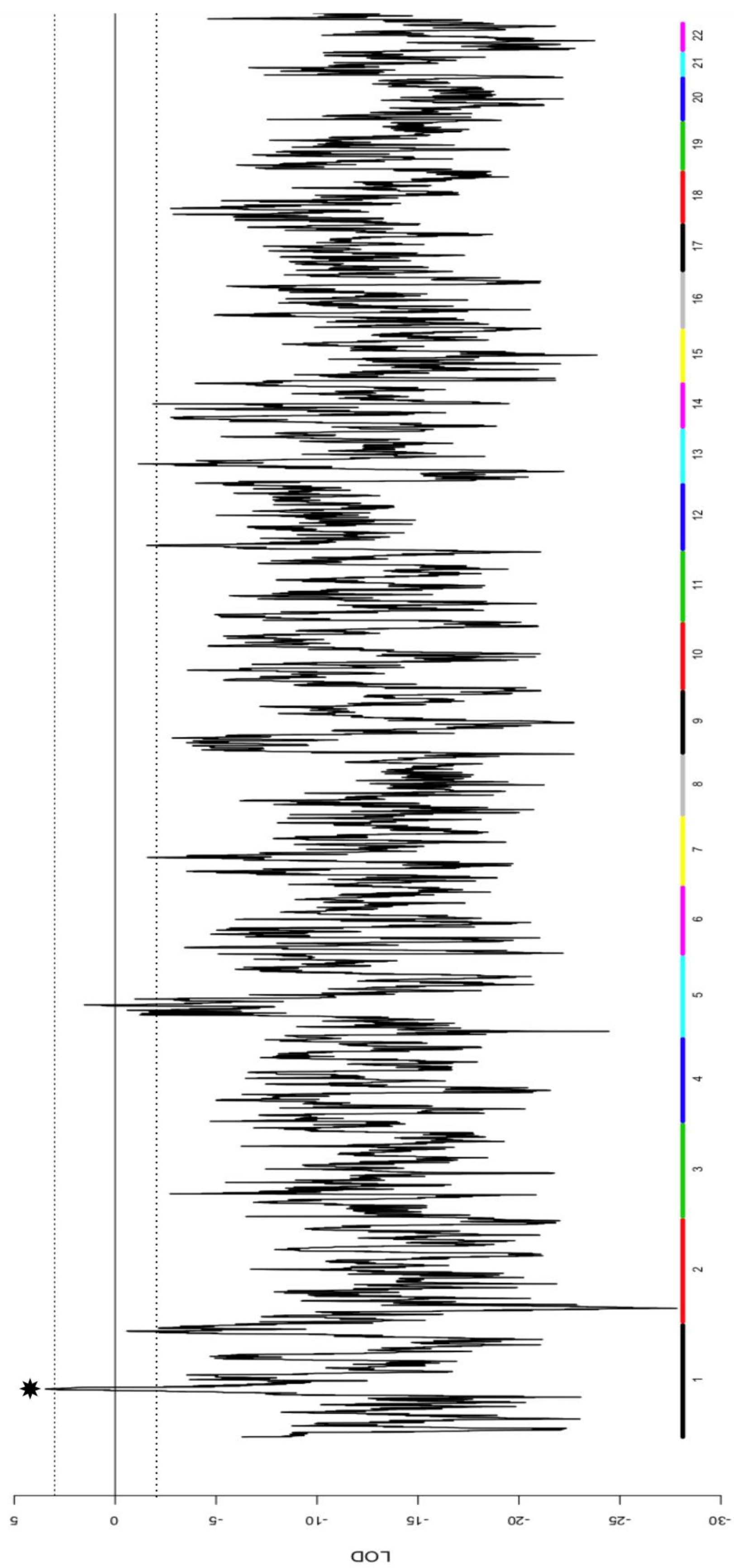


Figure 6. Four families with multiple relatives affected with MacTel.

Figure 7. Genome-wide affected only linkage scan in 23 families with 88 individuals.



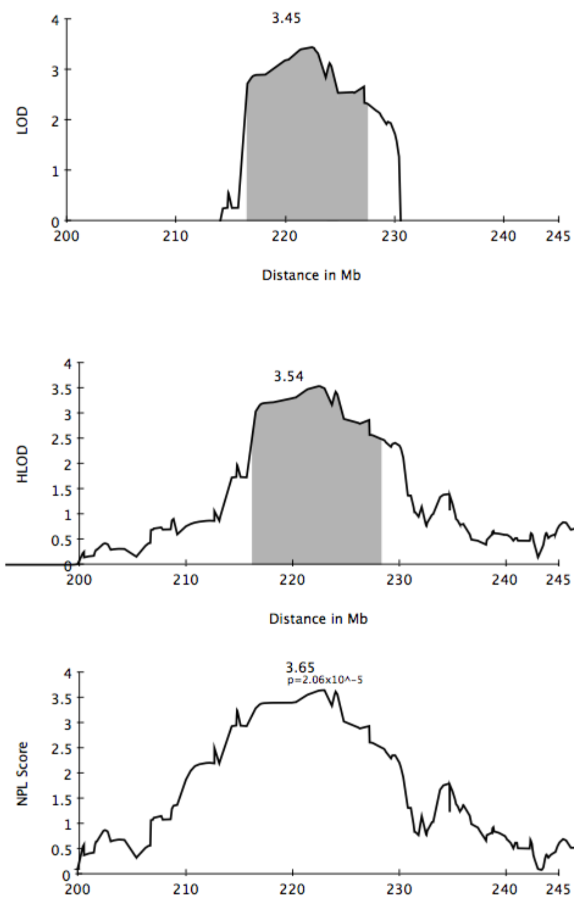


Figure 8. LOD, HLOD, and NPL scores for chromosome 1. The 1-LOD support interval is shaded in gray.

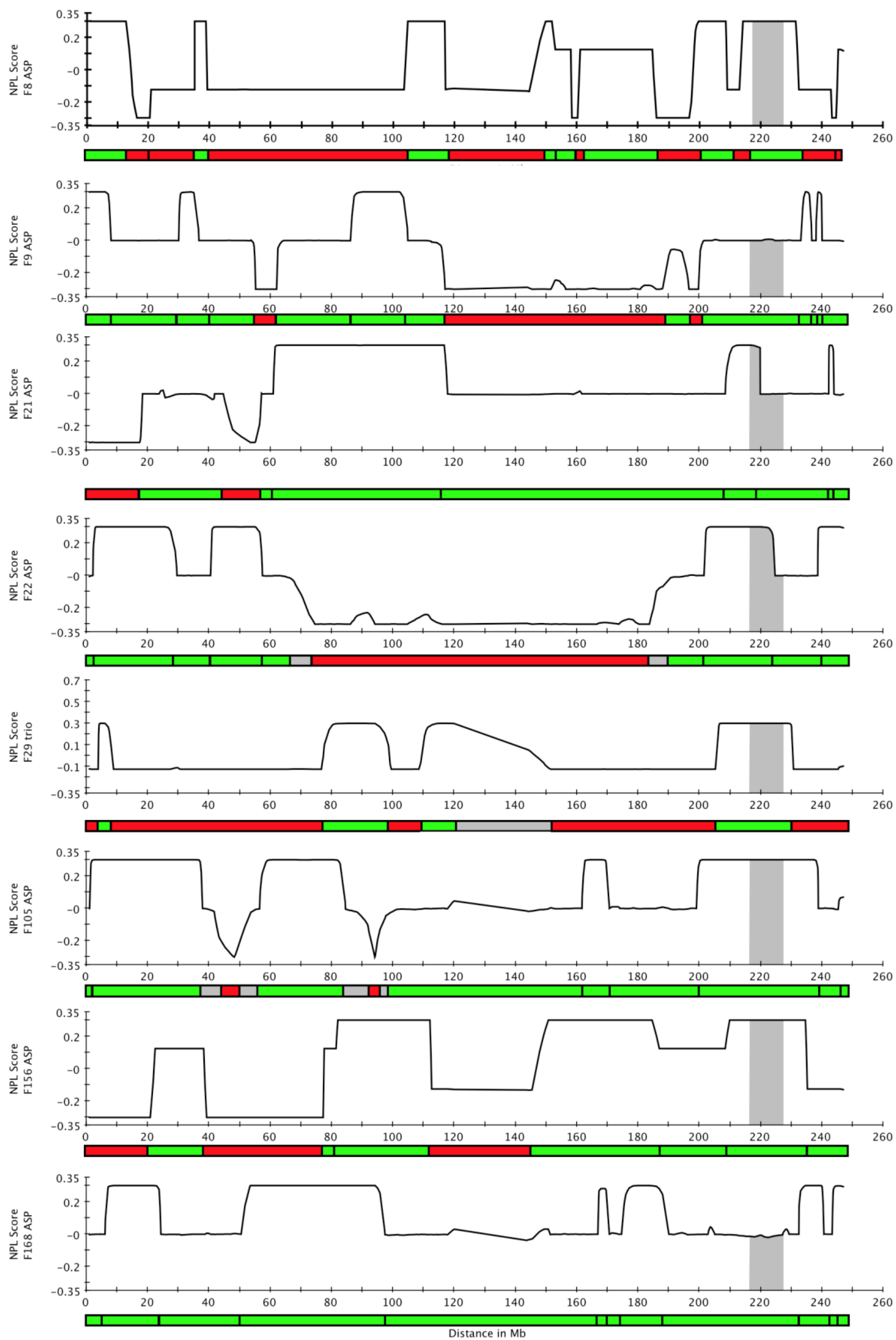


Figure 9. IBD allele sharing on chromosome 1 in 8 families.

Chr	Regions not excluded by parametric linkage (Mb)	Regions not excluded by allele sharing (Mb)
1	213.2 (2 markers) 214.27-230.038 (LOD=3.43, 15.768 Mb)	205.89-208.95 214.32-229.92
2	-	134.49-149.84 240.24-243.12
3	-	-
4	-	86.68-99.52
5	87.1-88.46 (1.36) 96.85-97.71 (0.86) 109.26-111.17 (1.91) 116.81-119.55 (2.74) 146.41-149.78 (1.52 alpha=.76; HLOD=2.43, 3.37 Mb) 150.46-152.95 (2.49) 172.63-172.69 (0.06)	
6	-	-
7	136.74-137.33(0.59)	33.75-77.87 125.92-145.41
8	-	-
9	-	8.10-11.76 78.0-86.35
10	-	-
11	-	7.0-16.89
12	6.62-7.16 (0.54)	5.38-7.03
13	80.56-83.03 (2.47)	-
14	106.14-106.36 (0.22)	82.66-94.26 102.52-106.38
15	-	-
16	-	10.31-11.74
17	-	-
18	-	-
19	-	-
20	-	-
21	-	-
22	-	-

Table 5. Exclusion based on parametric LOD score less than -2 under an autosomal dominant model.

Gene	Number of exons	Variants detected	MAF	Notes
DISP1	7	Unknown 75 g>a P25P Unknown 2835 a>g K945K rs9441941 3822 a>c P1274P	 .292	Syn Syn Syn
TLR5	1	rs5744174 1846 T>C F616L rs2072494 1775 A>G N592S	.375 .139	FV FV
SUSD4	9	None		
BEND5	1	Unknown 986 g>a G329D		DNS
CAPN8	19	rs35539373 734 c>a S245Y rs61823553 1775 c>t T592M	.472 .389	FV FV
CAPN2	21	rs17599 1702 a>c K568Q Unknown 582 g>a A194A	.292	FV Syn
TP53BP2 (1, 12)	18	rs61749337 566 c>t A189V rs34683843 685 c>a Q229K	ND .058	DNS DNS
FBXO28	5	None		
DEGS1 (1)	3	None		
NVL (15, 22, 23)	23	rs7534447 456 g>a R152R rs3754090 738 g>a L246L rs34631151 1210 g>a V404I	.058 .058 .033	Syn Syn
CNIH4 (3)	5	Unknown 48 t>g het F6V		DNS
WDR26	13	None		
MIR320B2	1	None		
CNIH3	6	None		

Table 6. Variants detected by Sanger sequencing in two affected individuals.

Syn = synonymous

FV = frequent variant

DNS = does not segregate with disease

( ) indicates exons that could not be sequenced

Summary of Relevance

The results described in this chapter represented the first indication of a potential susceptibility locus for MacTel. At the time these experiments were performed enrollment in the MacTel cohort was ongoing. Additional families with multiple affected relatives were soon identified, allowing us to repeat the linkage analysis with additional families. The findings from those studies are described in the following chapter.

Chapter Five – Whole exome sequencing, addition of families to the linkage cohort, and haplotype analysis

Whole exome analysis of four MacTel patients

Introduction

Recent advances in massively parallel sequencing have made it possible to interrogate large regions of the genome at a cost, and in a timeframe, that was impossible even a short time ago. Over the course of this project, the methods chosen for sequencing have been determined by the research question at hand, but also by the cost and availability of sequencing technologies. This project has been carried out at a time of enormous technological change in the methods of genetic analysis. This is reflected in the approaches we have employed in the study of this disease.

Exome sequencing represents a step between Sanger sequencing of small chromosomal segments and full genome sequencing. Exome sequencing is a type of targeting sequencing that uses large scale, high-throughput methods to interrogate selected regions of the genome, in this case, the protein coding portion of the genome. The advantage to exome sequencing at the time of this project was that a large number of bases could be interrogated at a cost far less than that of whole genome sequencing. Selecting regions of interest also reduces the information processing requirements necessary to interpret the large datasets that result from large scale sequencing projects.

When this project was undertaken, sequencing the exome, using commercially available capture methods, was more cost effective than creating a custom library of probes to target a region of interest, and whole genome sequencing was still in the purview of large facilities, rather than independent laboratories.

Exome screening as a technique for discovering disease causing variants rests on several assumptions. Foremost, this approach assumes that the disease under investigation is caused by a protein coding variant. Of equal importance, it assumes that a functional, disease causing variant can be discerned from a large number of variants not relevant to the disease phenotype (Ng et al., 2010b).

The exome is defined as the totality of all protein coding regions in the genome. The intention of whole exome sequencing is to utilize high-throughput next generation sequencing to detect all variants in all exons in a genome. Many Mendelian genetic diseases for which genetic causes have been determined are attributed to variants in the coding sequence of a gene. Potentially functional coding variants are more readily interpretable in comparison to a reference sequence than non-coding variants. Variants in noncoding regions may be equally prevalent and significant in their effect, but these are far more difficult to interpret.

In 2009, Ng et al published a report in which the causal variant for Miller syndrome, a rare recessive disease, was detected by sequencing the exomes of two affected siblings and two additional unrelated patients (Ng et al., 2010a). The detected variants were filtered to remove synonymous variants, frequent variants, and variants not predicted to be damaging, resulting in a short list of candidates for further analysis. At the time of that publication, sequencing costs were becoming far less prohibitive than they had been previously, however whole genome sequencing was still beyond the reach of most laboratories. This situation would change within a year—an indication of the current rate of change in both technology and pricing. At the beginning of 2010, the cost to sequence one exome was $20,000, compared to a year later when the cost to sequence an exome is approximately $2,000, and the cost to sequence a whole genome is approximately $10,000, and dropping.

Following the strategy of Ng et al., we selected two affected sisters, and two additional unrelated probands for whole exome sequencing, with the expectation that a causal variant would be inherited identical by descent in the two sisters. In the unrelated probands, a causal variant could be the same variant, identical by state—though we considered this to be unlikely given that MacTel is not a common disease—or another coding variant in the same gene, assuming no genetic heterogeneity.

METHODS

Sequencing and alignment was performed by Roche-454 (Connecticut). DNA quality was assessed by NanoDrop spectral analysis, OD 260/280 ratio, and molecular weight as assessed by gel eletrophoresis. Libraries were created using NimbleGen Sequence capture. Target sequences were amplified utilizing probes designed to capture 180,000 exons covering 34 Mb of the genome. Exon capture and sequencing was performed by Roche-454 using the GS FLX Titanium protocol. Alignment and mapping was performed using the GS FLX software package. Sequences were aligned against human genome assembly GRCh36 hg18. Known variants were identified based on dpSNP130. Final processed output was delivered in two forms: all variants, and high quality variants, as well as raw sequence reads. A variant was called if a base pair differed from the reference sequence on at least two non-duplicate reads, did not occur at the beginning or end of the read where the highest proportion of error is expected, and few other bases differed from the reference sequence for that read. Variants declared as “high quality” met all these criteria, and additionally were detected on both forward and reverse oriented reads.

The number of variants reported per sample was compared to expectations for the total number of variants as compared to published whole exome projects. Selected variants were validated against Illumina 1M genotypes and Sanger sequence data for the same samples. Sanger sequencing was performed to confirm variants of interest.

RESULTS

The total number of variants reported for each sample is shown in Table 7. The number of total variants detected ranged from 89,208 in sample A98 to 107,866 in sample A39. The number of high confidence, unknown coding variants (the variant pool of interest) ranged from 194 for sample A98 to 303 for sample A39. The average sample depth was 8.4x, indicating that, on average, each target was sequenced with approximately eight reads. Sequence was reported for between 83.8% and 91.1% of the targeted exome. The distribution of coverage across the genome was highly variable, with a small proportion of regions covered with many reads, but with much of the exome covered by only 3-5 reads (Figure 10).

Sample	A39	A98	A117	A180
Primary target region coverage	89.1%	83.8%	84.1%	91.1%
Average depth	9.4x	7.5x	7.8x	8.9x
All variants: all regions	107,866	89,208	94,025	133,450
All variants: target region	17,750	14,891	15,538	17,621
High confidence variants: all regions	52,877	40,226	40,309	58,402
High confidence variants: target region	9,211	6,969	6,887	8,773
High confidence missense	3,862	3,060	3,049	3,796
High confidence unknown missense	303	194	269	230

Table 7. Summary of the variants reported by sample, the percentage of the exome covered and the average number of reads (depth) by sample.

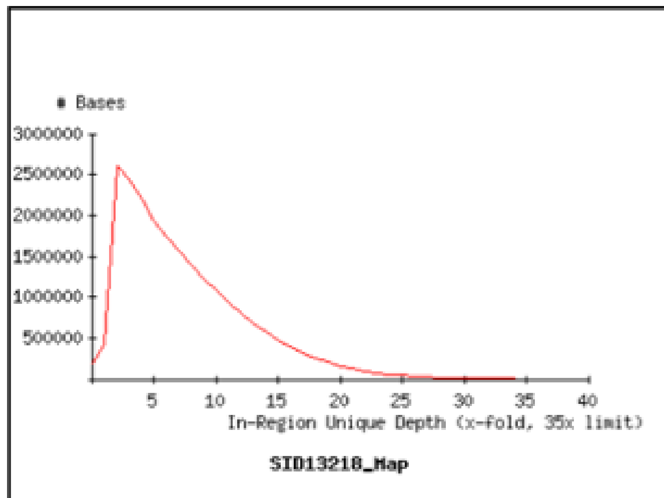


Figure 10. Plot of the number of bases (y-axis) covered by multiple reads (x-axis).

To assess data quality, whole exome results were compared to heterozygotes detected on the Illumina 1M chip for these samples, and to results from Sanger sequencing. Approximately 30% of the heterozygous loci detected in coding regions on the 1M chip were also detected in the whole exome sequence. Similarly, approximately 30% of the variants detected by Sanger sequencing were also detected in the whole exome sequence, indicating that coverage of the exome in these samples was incomplete. In light of this, it was not possible to evaluate the hypothesis of a shared variant between the affected siblings.

Whole-exome versus Illumina 1M					
Sample	**Coding dbSNP hets**	**Detected by 454 HC**	**Percent detected by 454 HC**	**Detected and called as hets**	**Percent detected and called as hets**
A39	6,386	2,454	38%	2,375	37%
A98	6,439	1,773	28%	1,697	26%
A117	6,377	1,653	26%	1,591	25%
A180	6,369	2,217	35%	2,146	34%
Mean	**6,393**	**2,024**	**32%**	**1,952**	**31%**

Table 8. Comparison of heterozygous coding variants detecting by exome sequencing and by Illumina 1M genotyping in four samples.

Whole-exome versus Sanger sequencing			
Coding variants		**Variants in flanking regions**	
Detected in Sanger	23	Detected in Sanger	18
Detected in 454 HC	7	Detected in 454 HC	5
Detected in 454 ALL	9	Detected in 454 ALL	8
Percent HC	30%	Percent HC	28%
Percent ALL	39%	Percent ALL	45%

Table 9. Comparison of heterozygous coding variants detecting by exome sequencing and by Sanger sequencing in four samples.

In the chromosome 1 linkage interval, no variants reported in the exome results were shared between the two affected sisters in family 8. By IBD analysis, it was determined that the two sisters share 2 alleles IBD in this region. Exome sequencing data indicated 1 private, unknown missense variant in A98 and 3 in A117 in this region. Analysis of all variants reported by exome sequencing in this region was similarly inconsistent with the previously established IBD allele sharing for these siblings. This inconsistency provided further indication that inadequate sequencing coverage resulted in a high percentage of false negatives in this experiment.

Gene	Sample	Amino acid change	Position	Read frequency	Number of reads
LYST	A98 A117	N > D	233923344	50% 36%	6
PPP1R12B	A117	NK > TK	200676467-200676469	23%	13
ZC3H11A	A117	F > L	202052855-202052856	87%	8
MAPKAPK2	A117	K > D	204972603-204972604	100%	3
CENPF	A117	I > F	212858621-212858622	36%	11
TGFB2	A117	N > T	216676001	89%	18
CDC42BPA	A117	KI > K*	225346236-225346238	23%	13
SPHAR	A117	FY > LC	227507564-227507565	50%	10
C1orf107	A98	K > N	208078957	67%	6
USH2A	A98	VL > VY	214661919-214661921	18%	17
FMN2	A180	R > Q M > R	238436936 238437392	56% 75%	16 4
OPN3	A180	R > Q	239824612	57%	7
OR2M3	A180	M > V	246432993	56%	9
PGBD2	A180	K > R	247178503-247178504	45%	11
LAMB3	A39	R > W	207869741	33%	9
OR2T8	A39	H > R	246151094	75%	4
OR2T34	A39	S > F V > A	246803942 246804218	33% 50%	9 10

Table. 10 Unknown missense variants detected by exome sequencing on chromosome 1 from 200 to 247 Mb (the 1q41-42 linkage interval and flanking regions). Allele sharing between the siblings A98 and A117 in the linkage interval from 214-299 Mb was inconsistent with allele sharing determined from the Illumina 1M SNP chip genotypes, which indicated that the siblings share 2 alleles IBD in this region.

One unknown missense variant was detected in sample A117 in TGFB2, which is known to be involved in blood vessel regulation (Zhang et al., 2003). In the same gene, a different, known, missense variant was detected in A98, the sibling of A117, and a third known missense variant

was detected in A39, an unrelated proband. The 8 exons in this gene were sequenced by Sanger sequencing in all four of the samples submitted for exome sequencing. None of the variants reported by exome sequencing were confirmed by Sanger sequencing, indicating that false positives were also present in the data set.

Discussion

This experiment revealed several practical limitations of exome sequencing. It was immediately apparent in the number of variants reported per sample that the overall number of variants was less than had been reported in similar studies. Our results significantly underreported the expected number of variants as compared to published reports and personal communication with other researchers. Additionally, the number of variants reported per sample was inconsistent. In sample A180, with an average of 8.9 fold coverage over 91.1% of the target region, 133,450 variants were detected. In sample A98, with an average of only 7.5 fold coverage over 83.8% of the target region, 89,208 variants were detected. These data strongly suggested that the detection of variants was incomplete, and was a function of poor coverage. Comparisons to genotypes from Illumina 1M SNP data and Sanger sequencing confirmed that only about 30% of the variants detected by those methods were reported in the whole exome dataset. Efforts to rectify this situation with the vendor were unsuccessful.

While the coverage delivered was in the range of the project specifications as stated by Roche-454, this level of coverage was inadequate to assess co-inheritance of alleles in the siblings A98 and A117. Ng et al. (Ng et al., 2010a) and Choi et al.(Choi et al., 2009) achieved approximately 40x coverage over 96 to 97% of the targeted exome. Other studies have sequenced to a fold coverage of at least 20x (Zhou et al., 2011) and as high as 100x (Yan et al., 2011). Choi et al. reported that the sensitivity to detect variants rose sharply between 5x and 20x coverage, and plateaued at 50x. While the sensitivity may vary depending on the read length and error rate of different sequencing technologies, published reports corroborate our results, indicating that the claims of sufficient sensitivity at 7-9x coverage for this technology were unfounded.

In practice, identifying all portions of the genome that are transcribed and translated is not a trivial undertaking. During the course of this experiment, we became aware of limitations in the definition of the exome as applied in commercial sequence capture methods (personal communication from Michael Dean, National Eye Institute, and Jason Corneveux, Translational Genomics Research Institute). The target regions defined as the “exome” for both capture methods available at the time of this experiment, Nimblegen, and Agilent, were based on the consensus coding sequence database (CCDS). The goal of the CCDS project is “to identify a core set of protein coding regions that are consistently annotated and of high quality.”(Yan et al., 2011) However, the CCDS lacks annotation for many transcribed genes, and annotation for isoforms of genes that are included in the CCDS is incomplete (Coffey et al., 2011). Thousands of known genes were not targeted at all by these capture methods, such that even with

complete coverage at the highest achievable sensitivity, any dataset produced by these methods would be necessarily incomplete.

As these issues have been brought to light by researchers in the field, an expanded definition of the exome has been established in an effort to improve on these limitations. Lessons learned from inadequate coverage of target regions can be used to improve experimental design as the field moves rapidly toward whole genome sequencing.

Addition of families to the linkage cohort and indications of genetic heterogeneity

Introduction

The analysis of additional families in a linkage study can reinforce a preliminary result, and narrow the region of interest. Additional evidence for the presence of a disease allele in the region of a linkage signal can be achieved by testing the effect on the LOD score when new families are added to the cohort. If a region identified by previous analysis co-segregates with a true susceptibility locus, the expectation is that the LOD score at the previously observed peak will increase. If an observed peak arose from the chance segregation of a region with the disease trait, the addition of more families decreases the probability of observing the same result by chance alone. If a region of positive linkage does not contain a disease allele, or if there is locus heterogeneity in the cohort, the expectation is that the LOD score will be reduced when additional families are added. The region of interest may be narrowed if one or more of the families added to the cohort have regions of positive linkage over a more narrow interval than the larger region of interest. If locus heterogeneity is present, additional regions of interest may be identified.

A linkage interval with a maximum LOD score of 3.45 was previously identified at chromosome 1q41-42 by analyzing 17 families with multiple relatives diagnosed with MacTel. Following that analysis, DNA was obtained from seven additional multiplex families. These families were

genotyped and the cohort was reanalyzed for linkage. For two siblings of probands in the original cohort, diagnoses were updated from possibly affected to unaffected after images from follow-up examinations were analyzed by the Reading Center. These two families consisted of affected sibling pairs, and thus were excluded from further analysis as simplex cases. One new sibling pair was excluded from analysis after it was noted that the observed genotypes were inconsistent with the expected proportion of allele sharing for siblings. A total of 21 families were analyzed for linkage in the new cohort.

METHODS

Five affected sibling pairs, one affected trio of siblings, and a pair of affected cousins were genotyped on the Illumina 1M SNP chip, as described in chapter 4. Three of the families (F38, F43, and F184) were recruited in Australia, two were recruited in the United States (F159 and F204), and one (F188) was recruited in Germany. All individuals in these families reported European ancestry, and all relatives that were analyzed were diagnosed as strictly affected.

All samples passed a genotyping quality control threshold with call rates greater than 99%. Genotypes from probands and relatives in the full cohort were reclustered, and the genotypes were inspected for quality metrics. The marker set of approximately 11,000 SNPs used in previous rounds of linkage analysis was applied to this cohort. These markers were screened for quality control in the reclustered genotypes as in previous analysis. Markers retained for

analysis were required to have a GenTrain score greater than or equal to 0.50, cluster separation greater than or equal to 0.16, and no missing genotypes. Markers that failed to pass any of these criteria after reclustering were excluded. Additional SNPs were added to replace those that were excluded. The resulting marker set consisted of 11,677 independent markers.

Genotypes were analyzed in MERLIN as in previous experiments, using model parameters of penetrance = 0.9, risk allele frequency = 0.001, and phenocopy rate = 0. Both two point and multipoint analysis were performed. Allele frequencies were supplied from 128 individuals genotyped in the MacTel cohort. Affected and possibly affected individuals in the cohort were coded as affected. Unaffected individuals and relatives with a diagnosis of probably not affected were coded as unknown. The analysis was carried out as an affected only analysis to prevent miscoding of relatives as unaffected who may carry a disease allele but not express the MacTel phenotype given that our hypothesis is that MacTel is not fully penetrant. Additionally, many of the genotyped relatives are not yet of the age that they would be expected to manifest signs of the disease. Additional tests were performed with 1) only strictly affected individuals coded as affected (all others coded as unknown), 2) affected and possibly affected individuals coded as affected, unaffected individuals coded as such, and those with a diagnosis of probably not affected coded as unknown, and 3) affected and possibly affected individuals coded as affected, unaffected individuals under the age of 55 coded as unknown, and unaffected individuals over the age of 55 coded as unaffected. After observing the change in LOD score from that observed in the smaller linkage cohort, additional tests were performed excluding

individual families to determine the contribution of individual families to the changes in the LOD score.

Recombinations were mapped in the newly added families in the region of the previously observed linkage peak at chromosome 1q41-42 using nonparametric analysis in each family to detect allele sharing states. This output was compared to previous results from affected sibling pairs and trios as described in chapter four to determine whether the minimal candidate region could be narrowed based on these data. Per family analysis was carried out to determine the contribution of each new family to the new LOD score.

Results

Nonparametric analysis of individual families revealed that the sibling pair in family 159 showed no allele sharing identical by descent, indicating that these two samples did not originate from siblings. This family was excluded from further analysis. The maximum multipoint LOD score obtained in affected only analysis with family 156 excluded (21 families analyzed) was 2.19, with HLOD = 2.92 and alpha = 0.81 at chromosome 1q41 (Figure 11). The peak spanned approximately 5 Mb, from 215 to 220 Mb. Previous linkage analysis had indicated a minimal candidate region of 15.6 Mb from approximately 214 Mb to 229 Mb. A second region of positive linkage was observed on chromosome 9p24-22. The maximum LOD score for this region was observed at rs10046838 (LOD = 1.25, HLOD 2.42, alpha = 0.72). The maximum HLOD

for this region was observed at markers rs10810584 and rs10115913 (LOD = 0.65, HLOD = 2.92, alpha = 0.75). All other regions of the genome produced negative LOD scores. Two additional regions produced HLOD scores above 2, though the LOD scores for these regions were weakly negative: one on chromosome 4 (rs7435692, HLOD = 2.43, alpha = 0.64), the other on chromosome 13 (rs182977, HLOD = 2.39, alpha = 0.73).

The cohort was reanalyzed for linkage with only relatives diagnosed as strictly affected coded as affected. Relatives with a diagnosis of possibly affected were coded as unknown to control for the possibility of misdiagnoses. A maximum LOD score of 2.49 (HLOD = 2.61, alpha = 0.90) was observed at chromosome 1q41, at the same position observed when possibly affected relatives were coded as affected. Two additional regions of positive linkage were observed on chromosomes 9 (LOD = 1.74, HLOD = 2.76, alpha = 0.85) and 4 (LOD = 1.20, HLOD = 2.22, alpha = 0.82), in the same regions as noted above.

When the same analysis was performed with unaffected relatives coded as such, no evidence for linkage was observed; LOD scores were negative at all loci. This result was not surprising, given that many relatives in the samples are much younger than the typical age of onset for MacTel. Relatives who are too young to be expected to manifest the MacTel phenotype yet carry a risk allele would reduce the LOD score when they are declared to be unaffected.

To control for miscoding of relatives resulting from age of onset effects, the unaffected relatives in the cohort were recoded as unaffected only if they were over age 55; unaffected relatives under age 55 were coded as unknown. This resulted in a maximum LOD score of 1.19 (HLOD = 1.86, alpha = 0.74) at the same position on chromosome 1 as observed in the affected only analysis. All other linkage scores across the genome were negative. This result was consistent with our hypothesis of reduced penetrance in the MacTel phenotype under the presumption that the linkage signal on chromosome 1 in the affected only analysis reflects a true susceptibility locus.

Inspection of allele sharing states in individual families across the previously identified linkage interval indicated that in four of the newly added families, siblings shared either one or two alleles identical by descent across the 1q41-42 region, compatible with a shared autosomal dominant risk allele in this region (Figure 12). Family 188, a pair of cousins, showed 0 alleles shared IBD on chromosome 1. Inspection of allele sharing across the genome in this family showed 1 allele shared IBD across ~13% of the genome, consistent with expected proportion of allele sharing for cousins. Family 171, an affected sibling pair, showed 0 alleles shared across most of the linkage region. A recombination was mapped to the region between 214 and 218 Mb near the centromeric boundary of the previously established minimal candidate region. This recombination failed to resolve more precisely by analyzing additional markers. From approximately 206 Mb to approximately 214 Mb the siblings shared 1 allele IBD, compatible with a co-inherited susceptibility locus. Allele sharing between 214 and 218 Mb remained ambiguous. From 218 Mb to the telomeric boundary of the linkage interval, the siblings shared

0 alleles IBD, indicating that the rest of the linkage interval is incompatible with a co-inherited susceptibility locus in this family.

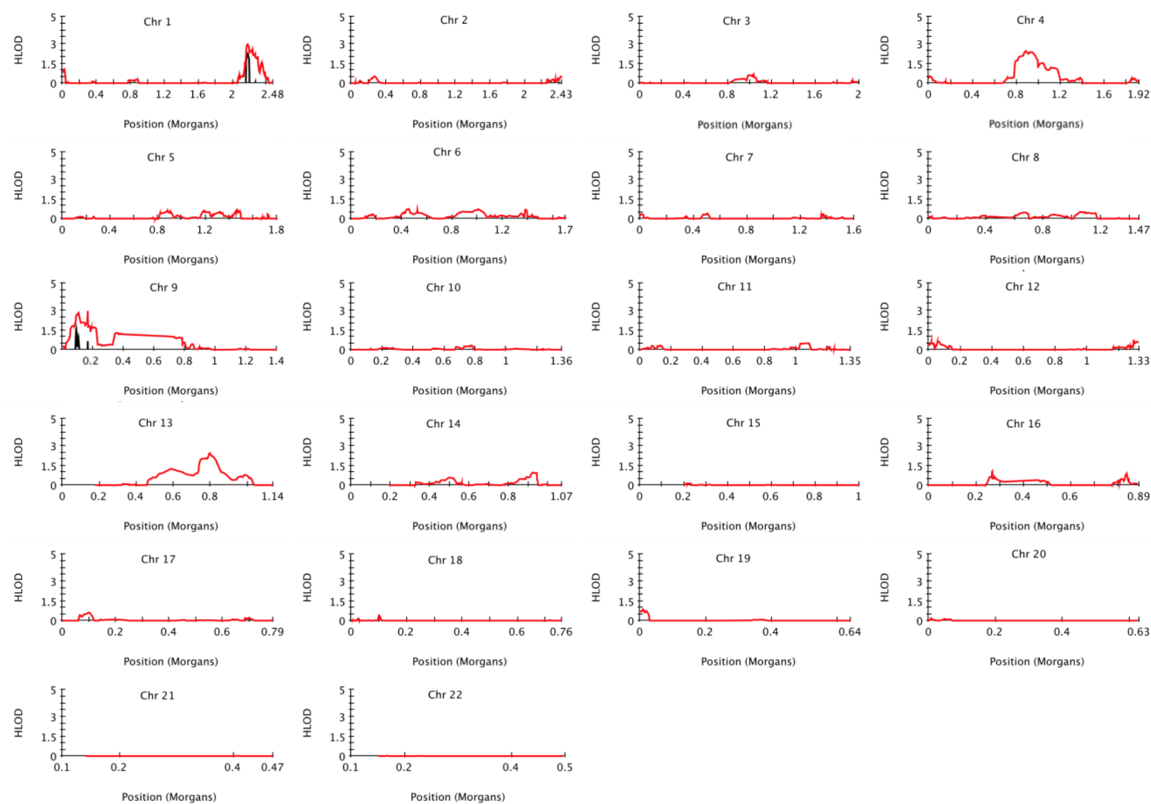


Figure 11. Linkage scores for the full cohort, analyzed with affected and possibly affected relatives coded as affected, and all other relatives coded as unknown. HLOD scores are shown in red, LOD scores are shown in black. Negative LOD scores are not shown. Positive LOD scores were observed on chromosomes 1 and 9. Regions of suggestive linkage based on heterogeneity LOD scores were observed on chromosomes 9, 4, and 13.

To determine which families were driving the reduction in LOD score from the previous linkage result, the analysis was repeated excluding the families 171 and 188 that were not linked to the 1q41-42 locus as determined by evaluation of IBD allele sharing. Excluding these two families,

the multi-point LOD score was 5.46 at chromosome 1q41-42, with HLOD = 5.46 and alpha = 1.0, indicating that all remaining families are linked to this locus. The previous result, before the addition of these families, resulted in a LOD score of 3.45 in the same region, with HLOD = 3.54 and alpha = 0.93. In that cohort, one large family was analyzed as two separate families because of the large number of missing individuals between the two branches of the family. One branch of that large family was not linked to the 1q41-42 region; the affected relatives in that branch were diagnosed as possibly affected. When only F171 was excluded, the maximum observed LOD score was 4.27 (HLOD = 4.29, alpha = 0.97), at 1q41, confirming that the narrowing of the linkage interval was due to the contribution from this family, consistent with the observed allele sharing in this region. When only F188 was excluded, the maximum LOD score observed was 5.12 (HLOD = 5.12, alpha = 1) at 1q41, indicating that all remaining families were linked to the narrow interval.

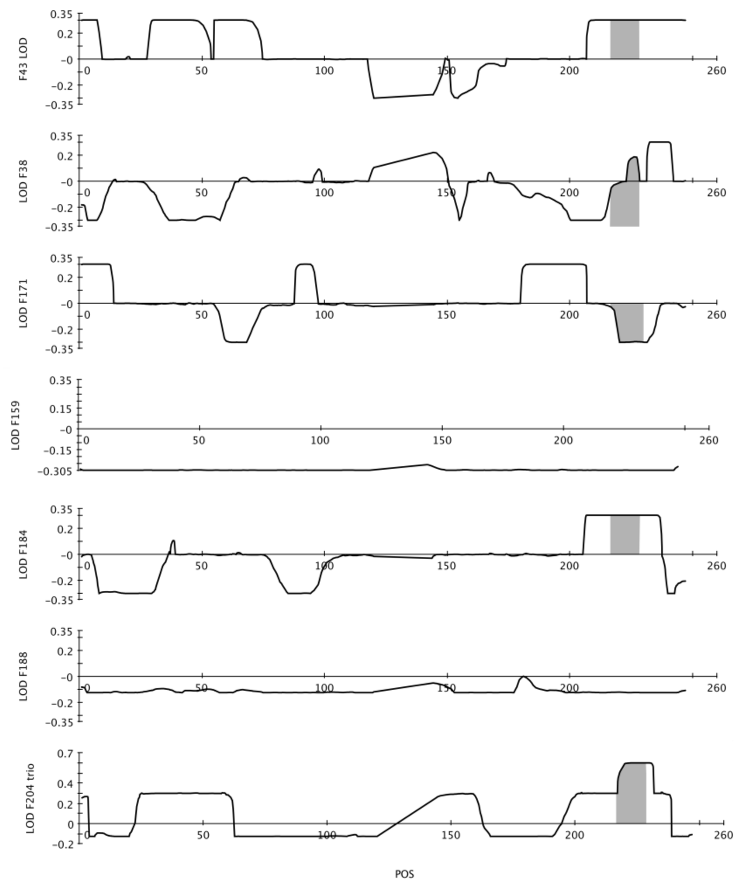


Figure 12. IBD allele sharing in seven new multiplex MacTel families. Family 159 showed no alleles shared IBD across the genome, indicating that the two individuals are not siblings. This family was excluded. Family 188 (cousins) showed no allele sharing IBD on chromosome 1. Family 171 showed a region near the centromeric boundary of the 1q41-42 where IBD allele sharing is ambiguous. The remaining families shared at least one allele IBD in the 1q41-42 region.

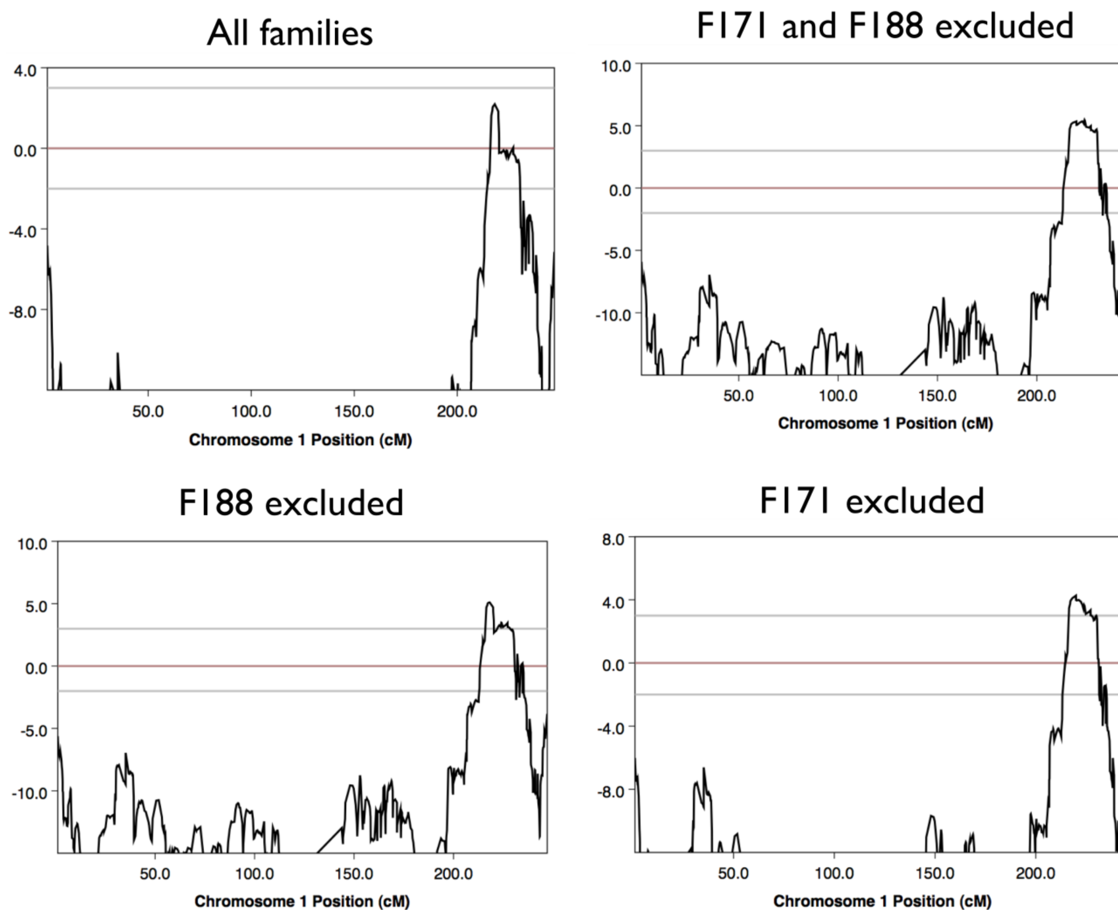


Figure 13. Linkage tests excluding families that did not exhibited allele sharing IBD in the region of interest on chromosome 1q41-42. Graphical output was generated by MERLIN. LOD score is indicated on the x-axis (the scale of the x-axis varies between graphs). The gray line at LOD 3 indicates the consensus value for rejection of the null hypothesis of no linkage. The gray line at LOD -2 indicates the value below which linkage is considered to be significantly rejected. The two left panels show the region of maximum linkage to be narrowed when family 171 is included in the analysis. The bottom two and the top right panels show that the maximum LOD score is increased when either family 171, family 188, or both are excluded from analysis. This result shows that the position of the linkage interval is unchanged from the previous linkage result when family 171 is excluded, and the reduction in LOD score from that result is due to the contributions of families 171 and 188.

Discussion

The observation of families unlinked to the primary MacTel susceptibility locus on chromosome 1q41-42 provided the first evidence of locus heterogeneity contributing to the MacTel phenotype. Locus heterogeneity is well known in other genetic ocular diseases, for example Leber congenital amaurosis (den Hollander et al., 2008), familial exudative vitreoretinopathy (FEVR) (Shastry et al., 1993), and retinitis pigmentosa (Daiger et al., 2007). Because the recombination in this region in family 171 is mapped ambiguously, it remains uncertain whether this family is linked to only a narrow region of the 1q41-42 locus, or whether the entire region is excluded in this family. The observation that family 188 is not linked to any portion of this region provides the first evidence of locus heterogeneity in the MacTel cohort. Given this observation, and the fact that in any one family 75% of the genome will be compatible with an autosomal dominant inherited allele by coincidence, exclusion of any portion of the linkage interval based on the inheritance pattern in only one family could be spurious. That is to say, if family 171 is linked to only a narrow region of the larger interval, that linkage could arise because a true susceptibility locus exists in the narrow region, or it could arise by chance. There is insufficient evidence at this time to discriminate between these two possibilities. Thus, given that additional families have been identified with demonstrated linkage to the 1q41-42 locus, this broad region remains of interest as possibly harboring a risk allele for MacTel. However, the observation of unlinked families indicates that other loci may to contribute to the phenotype, under the assumption that the phenotype is sufficiently well defined and these cases are not phenocopies.

PHENOTYPIC HETEROGENEITY IN MACTEL

Evaluation of the MacTel phenotype has identified variable expressivity in affected individuals even in families where presumably all affected relatives have inherited the same causative allele. This observation has persistently raised the question of whether MacTel is caused by alleles at one locus, whether more than one contributing locus is necessary to generate the phenotype, or whether modifying alleles and/or environmental factors affect the expression of the phenotype. Segregation of the disease in the few large families that have been evaluated is consistent with the inheritance pattern expected for a single segregating allele, thus, we have referred to MacTel as a Mendelian disorder in these families. This does not preclude the possibility that entirely different alleles, at different loci, could be segregating in different families, as in retinitis pigmentosa (Daiger et al., 2007) for example, where many distinct loci are known to cause the disorder, which segregates in a Mendelian fashion in those families.

The observation of genetic heterogeneity in the MacTel family cohort raises additional issues that must be taken into consideration in dissecting the susceptibility locus. When considering any given pair of siblings, on average, 2 alleles will be shared IBD across 25% of the genome, 1 allele will be shared across 50% of the genome, and 0 alleles with be shared IBD across 25% of the genome. Thus, 75% of the genome is compatible with inheritance of an autosomal dominant disease allele in any given affected pair of siblings. The power of this, or any, study comes from the analysis of many informative meioses, and from large families where the

expected proportion of the genome inherited identical by descent is much smaller due to the combined probabilities of any given region segregating to a gamete in multiple meioses. Linkage between a marker locus of known position and a disease locus of unknown position is declared when the observed data indicate that it is statistically improbable that the same region would be inherited identical by descent in many families, or in all affected relatives in a large family by chance alone.

When all families are linked to a presumed susceptibility locus with a significant LOD score, it is reasonable to assume for the purposes of further study that there is a true disease locus in the region of interest in all families. However, the possibility remains that such a presumption is an over simplification, and if more families were sampled, evidence for additional loci would emerge.

Given the observation of families not linked to the 1q41-42 locus, each family under analysis has a probability of demonstrating linkage to the region of interest by chance alone. It remains questionable whether family 171 is linked to the most centromeric region of the 1q41-42 locus given the inability to accurately resolve the recombination in the region in that family. There are three possible interpretations of the result: the recombination is centromeric to the linkage interval, in which case the family is not linked to the region of interest; the recombination occurred within the interval, and the siblings carry an inherited disease allele in the narrowed linkage interval; the recombination occurred within the interval but the siblings do not carry an

inherited disease allele in this region, rather allele sharing consistent with autosomal dominant inheritance occurred by coincidence. Given the evidence, it is not possible to determine which of these possibilities is the true condition. Parental genotypes are not available for these siblings. If additional siblings became available for genotyping, this added information would assist in reconstructing the parental genotypes such that the recombination might be accurately resolved. This could answer the question of whether the family is linked to the 1q41-42 locus, but would not indicate whether or not a true disease allele is present in the interval.

Depending on the true linkage status of family 171, either one or two families out of 21 are unlinked to the 1q41-42 locus. This observation enables an estimation of the proportion of families linked to the region, with the caveat that sampling error is likely in this small number of families. The estimation can serve as a rough guideline for how much confidence should be placed on the contribution of any single family. If 1 in 10 families do not carry a disease allele in the 1q41-42 region, any given family has a 90% chance of carrying a true risk allele in that region. Further genetic dissection of the interval must be interpreted with this caveat in mind.

The observation of suggestive heterogeneity LOD scores on chromosomes 9, 4, and 13 provides some evidence, albeit weak, for additional loci that could contribute to the MacTel phenotype. Enrollment in the MacTel study is ongoing; if additional multiplex families are identified, further linkage analysis may be warranted. As more families are sampled, the possibility increases of ascertaining families that harbor risk alleles in other regions, if these exist. While this would

partition the power of the cohort overall, it would increase the power to detect significant linkage at other loci.

Haplotyping of affected siblings

Introduction

A haplotype is the set of genetic material inherited in a single gamete. The defining feature of a haplotype is that a group of alleles together without recombination. For the purpose of searching for a disease allele, it is of interest to identify chromosomal segments inherited identical by descent in a pair of relatives. Sequencing or genotyping reveals which alleles are present in a genome, but provides no information as to the phase of the alleles, that is to say whether an allele is present on the maternally or the paternally contributed chromosome. Determining the phasing of alleles allows identification of the alleles on a putative risk haplotype. These alleles can then be compared across families.

The prediction that a shared risk haplotype might be present in unrelated families rests on the premise that "unrelated" families share a distant common ancestor, and that a risk allele could exist on a haplotype coinherited with the disease. The alternative hypothesis is that disease causing alleles have arisen de novo in multiple lineages as separate events. If the first scenario were true, the expectation would be that the risk allele would be identical across families. If the second scenario were true, the expectation would be that either locus heterogeneity, allelic heterogeneity, or both would be observed across families. Cystic fibrosis (CF) provides an example in which elements of both are observed. Over 1,800 variants in the CFTR gene have been identified, but approximately 70% of Caucasian patients are homozygous for the

p.Phe508del variant (Cutting, 2010). The presence of a common risk haplotype was exploited to narrow the region of interest in order to identify CFTR as the gene responsible for CF (Rommens et al., 1988).

The goal of this analysis was to identify phased risk haplotypes in affected siblings, compare the genotypes of the risk haplotypes across families, and determine whether a common haplotype was enriched in families with inherited MacTel in the region of the linkage peak on chromosome 1q41. Recombinations had been previously mapped using the NPL algorithm in the analysis of IBD allele sharing, so this effort was not expected to narrow the minimal candidate region based on allele sharing, rather to discern phase information in the siblings in order to be able to compare genotypes across families in an attempt to detect a shared haplotype, which could serve to prioritize a region of interest for further study.

METHODS

The genotypes of all probands and relatives in the linkage cohort were analyzed in MERLIN using the haplotype function. All available relatives were included in the analysis, without regard for diagnostic status. The analysis was performed using the set of chromosome 1 SNPs employed in linkage analysis, as well as a dense marker set in the region of interest consisting of all SNPs that passed the minimum quality control threshold as described in previous sections. To address the potentially confounding issue of ethnic heterogeneity in the study cohort, an

additional marker set was created by removing SNPs with MAFs that are very different between HapMap populations in an effort to reduce potential bias arising from the use of mismatched population allele frequencies. Population allele frequencies for SNPs in the region of interest were compared across 11 populations for which data was available in dbSNP (YRI, TSI, MKK, MEX, LWK, JPT, GIH, CHD, CHB, ASW, CEU (International HapMap Project, 2011). SNPs with allele frequencies that vary across these populations were identified by determining the standard deviation of the minor allele frequencies across all these populations. SNPs with a standard deviation greater than .10 were excluded. Markers were analyzed from 214-221 Mb on chromosome 1, flanking the 1q41 linkage peak described earlier in this chapter.

The output of the analysis consisted of two datasets for each family, one indicating the grandparental origin of each locus (designated A, B, C, or D), the other giving the corresponding genotypes. The phased haplotypes for affected relatives were compared to IBD allele sharing states as previously determined using nonparametric analysis as a control for consistency of recombination mapping with previous findings.

Applying a dense marker set consisting of all high quality SNPs in the region from 215-221 Mb, phased haplotypes and the corresponding genotyped were successfully resolved for families 8 and 156, where both parents were genotyped in each family (Figure 14). Comparison of these data to the previously mapped IBD allele sharing states showed agreement between the two methods.

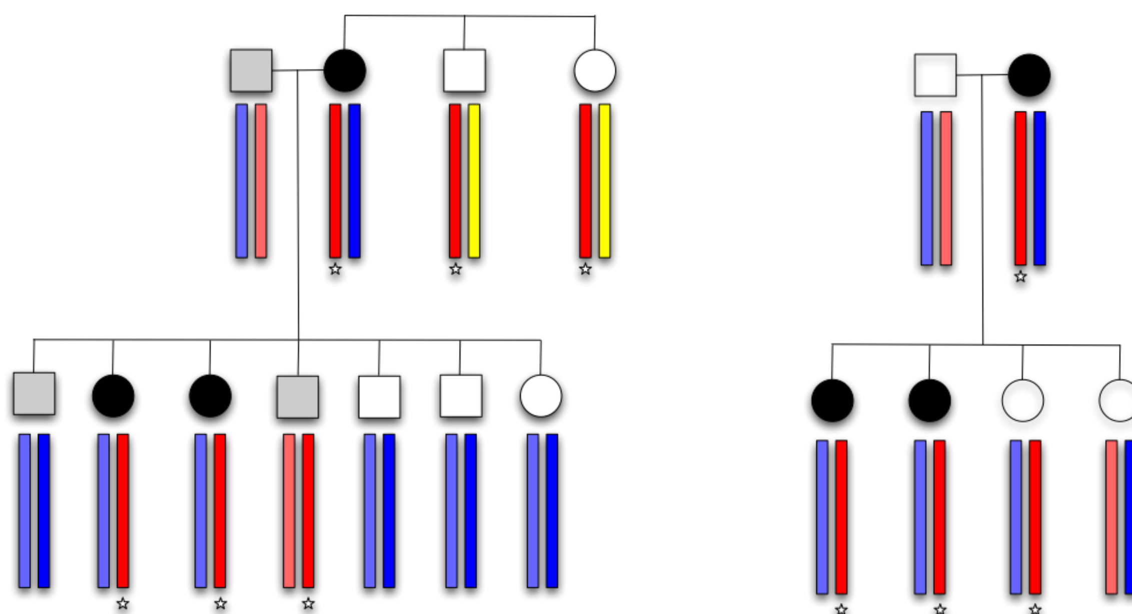


Figure 14. Pedigrees of family 156 (left) and family 8 (right) showing phased haplotypes in the region of the linkage interval on chromosome 1q41. Black pedigree symbols indicate affected individuals, open symbols indicate unaffected individuals, and gray symbols indicate individuals who had some abnormal characteristics on examination but were not diagnosed as affected with MacTel (designated "probably not affected"). The risk haplotypes are showed in red and starred. In offspring, the paternally contributed chromosome is shown on the left, the maternally contributed chromosome on the right.

A comparison of the alleles on the putative risk haplotypes in families 156 and 8 was of limited utility (Figure 14). An unwieldy number of regions showed exact matching of genotypes between the two families. The markers compared between the two families were identical between siblings within each family, to control for genotyping error. However, in most instances, runs of identical genotypes were interrupted by only one or two discordant markers. The largest region of identical genotypes occurred between 218.3 and 219 Mb, interrupted by two discordant markers at 218.6 Mb. The position of maximum LOD was observed at 218.2 Mb.

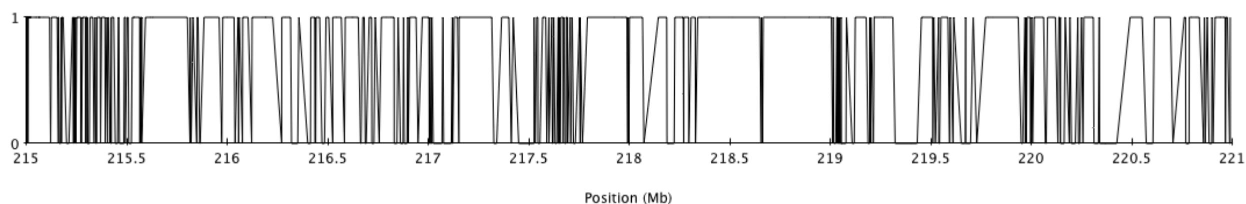


Figure 14. Comparison of genotypes in F8 and F156. A value of 1 indicates the genotypes are the same, and a value of 0 indicates that they differ.

All other families analyzed consisted of sibling pairs or trios with parental genotypes missing. In these families, the haplotype output was inconsistent with IBD allele sharing states as previously determined. The data indicated as many as 9 crossover events in a span of 30 Mb, inconsistent with the expected recombination frequency of 1 per 100 Mb, indicating an error in the method. Output from a sparse set of high quality markers mapped fewer recombinations, but still indicated false recombinations based on the recombination rate that would be required to produce such a result. Low resolution mapping also resulted in output where the phase was ambiguous over large portions of the region of interest.

In an attempt to resolve the issue, a marker set of 981 SNPs, filtered to remove SNPs with MAFs with a standard deviation greater than 0.10 between HapMap populations was applied, with similar results.

Discussion

The results achieved using this method identified many regions where alleles on the putative risk haplotype were shared between affected relatives in families 156 and 8. The largest region was observed from 218.3 to 219 Mb. However, genotypes were shared at numerous positions. The results in other families indicated a serious deficiency in the method as applied to the families in this cohort. Examination of population LD blocks using the CEU analysis panel in Haploview showed 26 LD blocks in the region between 218-219 Mb, spanning the longest segment of genotype sharing in families 156 and 8 (Figure 16). This region contains nine genes, none of which have been screened. However, this is only a portion of the linkage interval; there is not strong evidence to prioritize this region over others in the linkage interval.

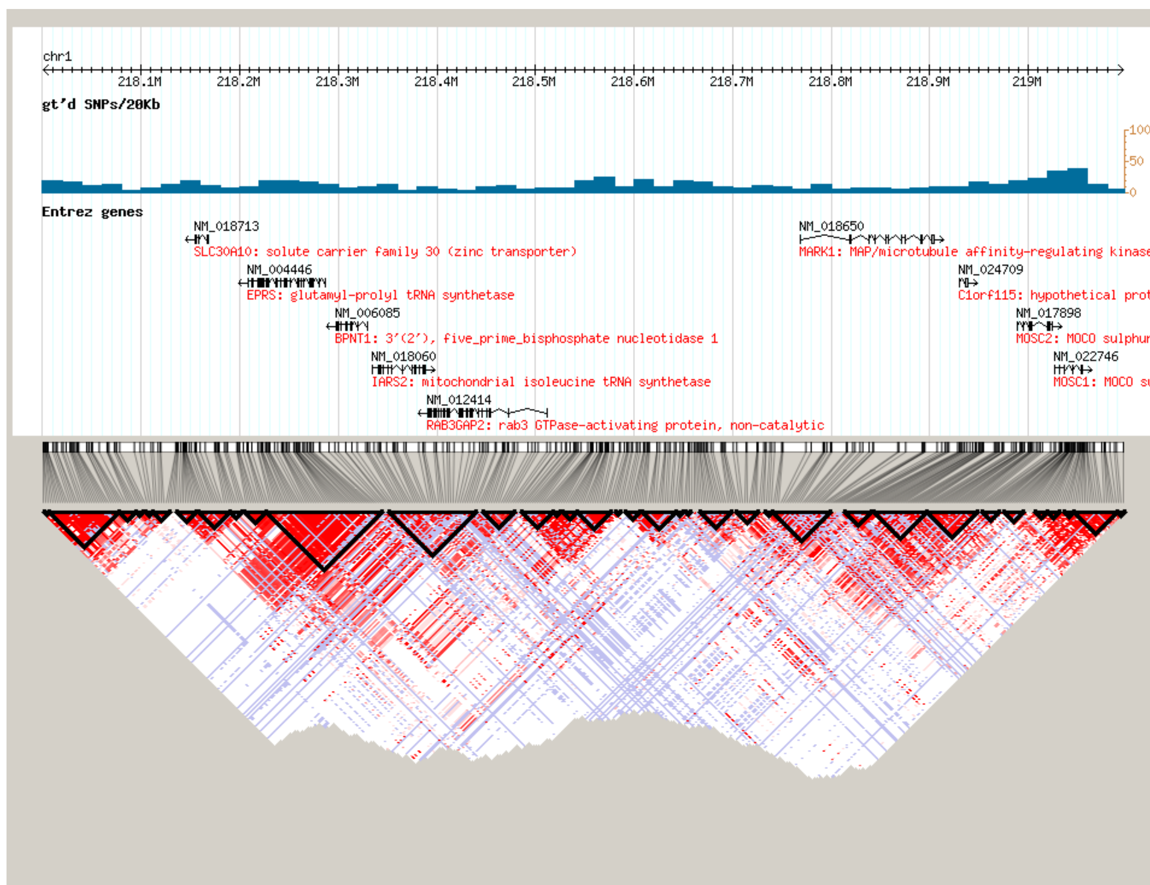


Figure 16. Linkage disequilibrium in the CEU population, visualized in Haploview (HapMap Phase 3, release 2 data).

This analysis was severely limited by the inability to resolve phased haplotypes in sibling pairs and trios where parental genotypes were missing. Niu et al. (Niu, 2004) described several computational methods for inferring haplotypes. Inferring haplotypes from nuclear families with absent parental genotypes, requires the evaluation of all possible phase configurations based on a limited amount of observed data. Niu notes that MERLIN performs well in large pedigrees with highly informative markers where there are fewer ambiguities to resolve, but has been observed to perform less well in the analysis of sibling pairs. Whereas the output of the NPL statistic in MERLIN is a probability of allele sharing, allowing ambiguity to be observed

and interpreted, the output from the phased haplotype analysis does not include a metric to interpret the confidence that should be assigned to the result. The phasing of genotypes is declared in a binary fashion by this method, even when confidence in the result should be low. Results can be compared to expectations of recombination rates, and to other analyses, such as IBD allele sharing, to determine whether the output conforms to reasonable expectations. In the case of most families available for analysis in this study, it did not.

In families 156 and 8, where parental genotypes were available, and haplotypes were well resolved, the putative risk haplotype was inherited by two unaffected and one possibly affected relative in family 156, and by one unaffected sibling in family 8. Taken alone, these data indicate that this haplotype does not segregate with the disease in these families. Given that heterogeneity was observed by linkage analysis, as described in the first section of this chapter, it is entirely possible that the positive linkage scores observed in one or both of these families under affected only analysis were spurious. A second hypothesis is that there is a true risk allele on this haplotype, but it is not sufficient to cause the disease. This hypothesis is consistent with the hypothesis of reduced penetrance in MacTel. The comparison of affected only versus affected-unaffected linkage analysis indicated that, if there is a true risk allele in the chromosome 1q41-42 region, some relatives diagnosed as unaffected are linked to that region, since the linkage signal was abolished when all unaffected were coded as such. Thus, observing unaffected relatives who carry the putative risk haplotype is consistent with other analysis. The linkage score observed at chromosome 1q41-42 remains the strongest evidence the existence

of a susceptibility locus for MacTel. This region continues to be of interest, with the caveat that in any given family, linkage to this region may be observed by chance alone.

Chapter Six – Summary and Conclusions

Summary of results

In this work, I have analyzed the first cohort of MacTel patients and their families in collaboration with other researchers in the MacTel project. During the course of this project, significant work has been done by others to refine the MacTel phenotype, and study the epidemiology of the disease. These efforts support the genetic analysis. Findings related to macular pigment distribution suggest that defects in these pathways may be primary in the etiology of the disease. Retinal vascular effects may occur subsequent to disruptions in macular pigment distribution, but the specificity of the neovascular phenotype still suggests that there may be an underlying genetic predispostion that predisposes the intraretinal vessels to become dysregulated. It is conceivable that a common regulatory pathway might impinge on both macular pigment transport and intraretinal vascular regulation, though this remains speculative pending the identification of a causative gene for the disease. The implication of Mueller glia cells in the MacTel phenotype (Powner et al., 2010) is significant, in that it is supposed that these cells play a role in macular pigment transport and deposition in the retina. Mueller glia cells are also in contact with intraretinal vessels. A single gene expressed in these cells could play a dual role, affecting both macular pigment transport and vascular homeostasis, or multiple genes, coexpressed in the same location could interact to affect both systems. Functional studies will be required to answer these questions if genes involved in the pathogenicity of MacTel are discovered.

At the outset of this project, the hypothesis was made that MacTel is in fact a genetic disease. This hypothesis was tested by examining the family members of probands with full ophthalmic examination and imaging adequate to detect the phenotype. The first major finding of this work, due to the efforts of the clinicians at the MacTel recruitment centers and the Moorfields Reading Center, was to identify families with multiple relatives affected with MacTel. The second significant finding was that the inheritance pattern is consistent with autosomal dominant inheritance, suggesting that the disease could be Mendelian in these families.

Significantly, it has not been possible to classify each family member as either affected or unaffected. The four categories of diagnostic status employed in this study—affected, possibly affected, probably not affected, and unaffected—may represent different stages of the disease, or may represent different presentations of the disease. So while it was a significant step to identify families with inherited disease, we also observed that the phenotype is variably expressed.

At the outset, it was believed that MacTel was a multifactorial trait. This may be the case, but the indication at this point is that, in some families, the expression of the phenotype may be controlled by large effects of a small number of factors, rather than small effects of many factors. In the few large families in the study, the ratio of affected to unaffected offspring is close to the 50:50 ratio that would be expected for a fully penetrant disorder with autosomal dominant inheritance. The question remains as to whether relatives that have been diagnosed

as possibly affected carry a modifying allele that affects the expression of the phenotype, whether environmental factors affect expression, or whether the disease is not significantly different in these patients, but they are in the early stages. The last point will be answered as these patients are followed clinically. The first question may be addressed if a causative gene is identified, but more families would be needed to sufficiently power a linkage study to identify modifying alleles. The question of possible environmental effects will remain difficult to answer with certainty.

In this work, I have described negative results from Sanger sequencing of 40 candidate genes, 27 of which were selected based on functional criteria, and the remainder selected based on their position under the chromosome 1 linkage peak. The exons of these genes were sequenced in probands from families with multiple affected relatives. Rare or unknown missense variants were screened in a cohort of unrelated MacTel patients, and in a large cohort of screened controls. Such variants were also screened in the relatives of the family in which they were detected to determine whether or not the variant co-segregated with the disease. The criteria for declaring a variant to be a putative causative variant for MacTel were that it would be present in all affected relatives in the pedigree in which it was detected, and that it would not be present with similar frequency in controls as in unrelated MacTel patients. No variant met these criteria.

The failure to detect a causative variant by candidate screening was not surprising. That portion of the work was undertaken during the early enrollment portion of the study, before a sufficient number of individuals had been enrolled to provide power for other analysis methods. While there was always a low probability of finding a causative allele with that method, there is extensive literature on the molecular genetics of retinal vascular pathways in the retina, and to a far lesser degree, the pathways that contribute to macular pigment transport. It is entirely possible that the pathways selected for screening may play a role, but in many cases, it is simplistic to use the term “pathways” given that these are actually networks of gene products, with many interacting gene products. In some cases, such as the Norrin/Frizzled-4/LRP5 pathway, I screened several known interacting genes. However, there are many more in this pathway alone that I did not screen. The same can be said for most of the genes that were screened. Candidate genes were selected based on the strongest evidence from the literature for genes with phenotypic similarity to the MacTel phenotype, but many interacting proteins, both known and unknown, could impinge on the same pathways, making and exhaustive screen by direct sequencing next to impossible. As an interim approach during early enrollment phase, it was a reasonable approach, given the low cost of Sanger sequencing, until more powerful methods were available.

Additionally, candidate gene screening only queried the exons and exonic flanking regions of the genes selected for screening. Variants detected in flanking intronic regions were largely uninterpretable. Intronic regions are more variable than coding regions, thus more variants are detected in those regions, but population frequency data is less readily available. Thus a

causative variant would be more difficult to discern from the many benign variants in non-coding regions. Intronic and non-genic variants are supposed, probably correctly, to be more numerous because they have no functional effect on the gene product, and thus are not under selective pressure. However, some number of these variants certainly occur in regions with regulatory function. The difficulty is in discerning which are which. Regulatory elements may be located at some distance from the coding region of the gene, and lack the clear structure of an open reading frame, a start, and a stop, by which to identify them. Work done in model organisms, such as drosophila melanogaster, has clearly shown the importance of regulatory elements, but at this time it is difficult to extend this work to human studies in a meaningful way.

This work identified the first locus with statistically significant linkage in MacTel families. When 17 families with inherited MacTel were analyzed for linkage, all families displayed linkage to chromosome 1q41-42 with a significant lod score. When additional families were added to the analysis, heterogeneity was detected, indicating that there may be more than one causative locus for the disease. This was not an unexpected finding. Observing that two out of 23 families are not linked to the chromosome 1q41-42 locus, each family that does display linkage must be assessed for the possibility that linkage to the region could be by chance alone. Considering an affected sibling pair, it is expected that across 25% of the genome, the siblings will share two alleles identical by descent, across 50% 1 allele with be shared identical by descent, and across 25% of the genome no alleles will be shared identical by descent. Given this expectation, 75% of the genome is consistent with inheritance of an autosomal dominant trait. Power is built up in

linkage studies by combining the findings from many families, or from identifying large families where the proportion of expected allele sharing is lower because of the complexity of the family structure. Extrapolating from the small number of families in this cohort, it is reasonable to estimate that 1 in 10 families does not carry a susceptibility allele in the 1q41-42 region. While this estimate is subject to sampling error given the small number of families, it is an indication that results from any single family may be spurious.

FUTURE DIRECTIONS

This work has been carried out during a time of rapid change in the technology and methods available to individual labs for the study of human genetic disease. Due to the support of the MacTel Project, we have been able to apply very newly available technologies to the study of this disease. The linkage interval detected remains too large to sequence reasonably using Sanger sequencing. Narrowing the interval using methods available previously would require adding additional families to the linkage analysis, in the hopes of observing a recombination between the endpoints of the linkage interval and the true disease locus. The observation of heterogeneity in the cohort complicates this approach, because any such recombination would have to be interpreted in light of the conditional probability that linkage to the region in any given family could be spurious.

Exome sequencing, in theory, could reveal a coding mutation, and filtering the results based on linkage and recombination mapping results adds additional power to such an approach. The inherent limitation of this approach is that non-coding variants will not be detected. Exome sequencing has been widely seen as an interim step in methodology. It became available at a time when genome sequencing was out of reach for individual labs. One advantage to exome sequencing is that it produces a smaller dataset, with more readily interpretable results. As whole genome sequencing becomes more readily available, the bioinformatics support necessary to interpret that data will need to keep pace. This is already proving to be a significant issue. In the case of this project, significant technological limitations of exome sequencing as a very new technology resulted in data that was insufficient to answer the research question. In the short time since, publications and personal communications indicate that groups still employing exome sequencing are taking these risks into account, demanding heavier sequencing coverage of the exome, and expanding the definition of the exome.

Recently, our group sequenced the genome of one MacTel patient using the services of Complete Genomics. The data was compared to genotyping and Sanger sequencing results to determine that both the false positive and false negative rates were acceptably low. We have recently sequenced eight additional genomes from 7 families. Variants detected will be filtered as planned for exome sequencing, to query rare or unknown missense variants. These data will allow inspection of the entire linkage interval on chromosome 1. With the majority of families linked to this region, the hypothesis is that a causal variant would be shared identical by descent in affected relatives, and that the same or another variant will be present in unrelated

affected individuals. As sequencing costs continue to fall rapidly, additional genomes may be reasonably sequenced, adding additional power to detect a true causal variant.

This is the first time in history that large scale sequencing has been available to individual labs to investigate a disease in this manner. This work has been performed at a time when many groups are searching for the best ways to apply these newly available methods to detect functional variants in the genome. This work has drawn on both traditional and very new methods, in new combinations, to attempt to determine the genetic cause of a disease that has not been previously studied for a potential genetic cause.

This work represents my effort, and the efforts of many collaborators, to identify a cause for MacTel. These efforts are ongoing, in the hope that identifying a genetic cause would provide guidance for further research, and potentially identify targets for treatment to improve the lives of those affected with this disease.

Appendix One – A mathematical exploration of model parameters

Introduction

In the previous analyses, we had assumed an autosomal dominant mode of inheritance for MacTel, and that no cases in our sample were caused by non-genetic factors. Only affected and possibly affected individuals were analyzed to prevent potentially declaring non-penetrant risk allele carriers as unaffected. This approach had the drawback of discarding genetic information from the unaffected individuals. Here we explore the effects of some different assumption on the linkage results, testing dominant versus recessive models under three prevalence values and three penetrance values, allowing for a small proportion of sporadic (non-genetic) cases. We also use all individuals in the families, both affected and unaffected. The goal of these analyses was to recover information from the unaffecteds that was not utilized in the previous analyses. Additionally, model misspecification can result in linkage scores that are diminished as compared to the scores that would be observed under a true model, which is to say a model that is consistent with the actual values for mode of inheritance, risk allele frequency, penetrance, and phenocopy rate for the disease (Hodge and Elston, 1994). While the true values for these parameters are unknown, this kind of analysis can reveal something about the true mode of inheritance and associated parameters by maximizing over models (Hodge and Elston, 1994; Greenberg et al., 1998; Abreu et al., 1999).

Methods

The total population prevalence of a disease is the sum of the prevalence due to genetic reasons and the prevalence due to non-genetic reasons (phenocopies). We assume that the majority of MacTel cases are genetic, and a small percentage (1%) are non-genetic. By assuming values for total population prevalence and further assuming that the allele (or alleles) responsible are in Hardy Weinberg equilibrium, that is, not subject to selective pressure (a reasonable assumption for a late-onset disease), we can calculate the allele frequencies that would lead to a given population prevalence. Additionally, we assume reduced penetrance, which is to say that not every individual carrying a risk allele will express the MacTel phenotype. This analysis tests population prevalence values of 0.001, 0.002, and 0.005, and penetrance values of 20%, 50%, and 80%.

We use G to represent the at-risk genotype. In other words, for a dominant disease, where A is the risk allele, and a is the non-risk allele, G = AA or Aa. For a recessive disease where a represents the risk allele, G = aa. Let K represent the population prevalence. Let f represent the penetrance, where penetrance is defined as the probability of being affected given that at individual has a risk genotype. This can be written as P(affected | G). Let g represent the phenocopy frequency, defined as the probability of having the disease given that an individual does *not* have an at-risk genotype. This can be written as P(affected | not G). Then, the total population prevalence can be expressed as the sum of the proportion of individuals in the

population with the at-risk genotype multiplied by the probability they will express the trait, plus the proportion of individuals in the population with the non-risk genotype multiplied by the probability they will express the trait for non-genetic reasons.

$$K = f \cdot P[G] + g \cdot P[not\ G] = fP[G] + g(1 - P[G])$$

The prevalence due to genetic and non-genetic causes can also be expressed graphically, as in Figure 1.

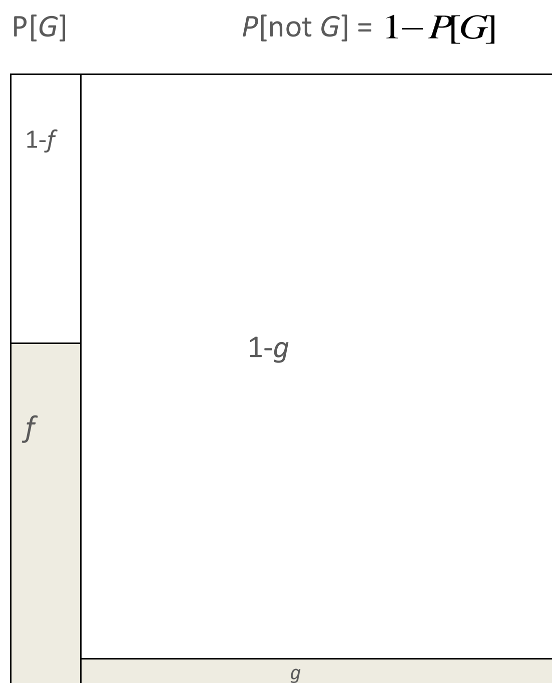


Figure 1. Visual representation of proportions of affected and unaffected people in a hypothetical population. The narrow left bar represents proportion of the at-risk genotype *G* in the population, & the broad right section of the graph represents the proportion of the population without *G*.

For a single biallelic locus, where p is the risk allele and q is the non-risk allele, the sum of the two allele frequencies in the population must add up to 1. This is expressed as p + q = 1. The sum of the genotype frequencies must also add up to 1. This is expressed as $p^2 + 2pq + q^2 = 1$. From this, we can express the proportion of genetic cases in terms of allele frequencies.

$$P[G] = \begin{cases} p^2 + 2pq = 1 - q^2 & \text{for dominant models} \\ p^2 & \text{for recessive models} \end{cases}$$

Referring to figure 1, within the *G*, section the proportion *f* is affected (shaded in), & the proportion $1 - f$ is unaffected. Similarly, within the not-*G* section, the proportion *g* is affected.

The total *shaded* area represents *K*, the population prevalence. From the diagram, we can see that *K* must equal

$$K = f \cdot P[G] + g \cdot P[not\ G] = fP[G] + g(1 - P[G]). \quad (1)$$

We can also see that the proportion *of K* due to nongenetic causes, ϕ, must equal the shaded area showing the *g* (i.e., the narrow in bottom right of diagram) over the total shaded area:

$$\phi = \frac{g \cdot P[notG]}{K} = \frac{g(1 - P[G])}{K},$$

or, equivalently,

$$g(1 - P[G]) = \phi K$$

From these equations, substitute eq. (*b*) into the rightmost term of eq. (a), to get $K = fP[G] + \phi K$. Subtract ϕK from both sides to get $K - \phi K = (1-\phi)K = f\,P[G]$. Then divide both sides by *f* in order to solve for *P*[*G*]:

$$P[G] = \frac{(1-\phi)K}{f}. \tag{2}$$

Another way to think of eq. (2) is this: Of the total shaded area in the figure (which represents population prevalence *K*), the proportion that's genetic is $(1-\phi)K$. If penetrance were 100%, then that proportion would also exactly equal the proportion of the at-risk genotype in the population. But in our model, the proportion of the at-risk genotype must be greater, so that when we multiply by the reduced penetrance, we'll get the right population proportion of affected people with the genetic form.

Once we know *P*[*G*], we can find *g* , because we know from eq. (*b*) that

$$g = \frac{\phi K}{1 - P[G]}.$$

From these relationships, we can solve for the risk allele frequencies under any assumed values for K and g.

Dominant

From eq. (1), $q^2 = 1 - P[G]$, so *q* is the square root of that, & then subtract from unity to get *p*:

$$p = 1 - \sqrt{1 - P[G]}. \tag{3}$$

Recessive

From eq. (1), $p^2 = P[G]$, so

$$p = \sqrt{P[G]}\ .$$

Table 1 gives the calculated values assuming populations prevalences of 1/10,000, 1/5,000, and 1/2,000, assuming penetrance values of 20%, 50%, and 80% for each assumed prevalence, with the assumption that non-genetic cases account for 1% of total prevalence.

				Risk allele frequency (p)	
Prevalence (K)	Penetrance (f)	Phenocopy rate (g)	Pop. prev. of at-risk genotype(s), P[G]	Dom.	Recess.
.0001	0.20	1.0005×10^{-6}	4.950×10^{-4}	2.475×10^{-4}	0.0222
	0.50	1.0002×10^{-6}	1.980×10^{-4}	9.900×10^{-5}	0.0140
	0.80	1.0001×10^{-6}	1.238×10^{-4}	6.188×10^{-5}	0.0111
.0002	0.20	2.0020×10^{-6}	9.900×10^{-4}	4.951×10^{-4}	0.0314
	0.50	2.0008×10^{-6}	3.960×10^{-4}	1.980×10^{-4}	0.0198
	0.80	2.0005×10^{-6}	2.475×10^{-4}	1.238×10^{-4}	0.0157
.0005	0.20	5.0124×10^{-6}	2.475×10^{-3}	1.238×10^{-3}	0.0497
	0.50	5.0050×10^{-6}	9.900×10^{-4}	4.951×10^{-4}	0.0314
	0.80	5.0031×10^{-6}	6.188×10^{-4}	3.094×10^{-4}	0.0248

Table 1. Calculated values for phenocopy rate, population prevalence of at-risk genotypes, and allele frequencies for dominant and recessive models under assumed values for total disease prevalence and penetrance.

Results

Consistent with results from previous analyses, multipoint parametric linkage analysis using the models listed in Table 1 resulted in a maximum LOD score at chromosome 1q41, under all dominant models (Figure 2.). The maximum LOD score of 3.83 (HLOD = 3.86, alpha = .96) was observed under the assumption of 1/5000 prevalence with 20% penetrance. The maximum HLOD was observed at the same marker, under the same model. For all prevalence values, the maximum score was observed under the model of 20% penetrance. No evidence for linkage was observed under any recessive model. One additional region of slightly positive linkage was observed on chromosome 9 (LOD = .42, HLOD = 1.6, alpha = .68) at approximately 9 cM under a model of 1/5000 prevalence and 20% penetrance (Figure 3). This region yielded a slightly positive LOD score in previous analyses. All other regions across the genome yielded negative LOD scores under all models tested.

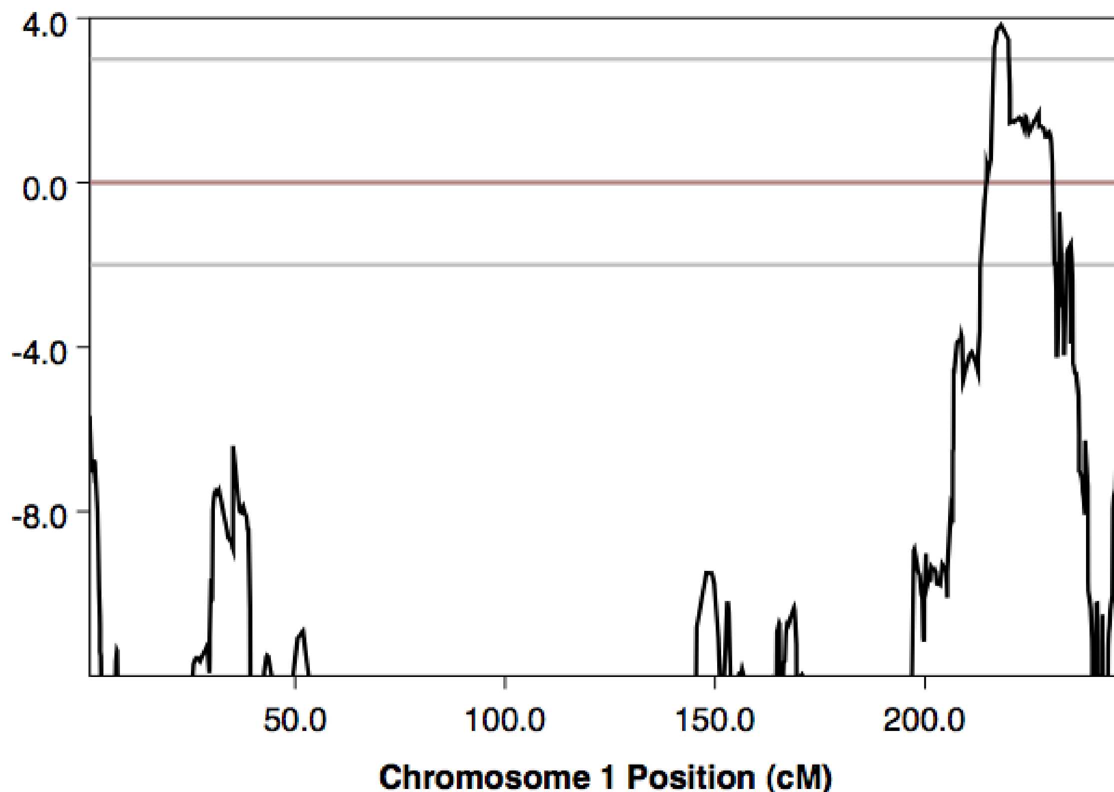


Figure 2. The maximum LOD score observed under all models occurred at chromosome 1q41, under a model of 1/5000 prevalence and 20% penetrance.

Discussion

These analyses resulted in a maximum LOD score observed at the same position on chromosome 1 as observed in previous tests for linkage. The maximum LOD score increased slightly as compared to previous tests when unaffected family members were included in the analysis and reduced penetrance was modeled. The calculated risk allele frequency under the model that yielded the highest LOD score in these tests was 0.0012383; the risk allele frequency

assumed in previous tests was 0.001. This result suggests that MacTel is subject to reduced penetrance, and that there is evidence under additional models for linkage at the region of previous interest at chromosome 1q41. From these analyses, we observed that there is statistical evidence favoring an autosomal dominant mode of inheritance for MacTel.

This work represents an approach to analyze a collection of families for linkage when the mode of inheritance, population prevalence, penetrance, and phenocopy rates are unknown. One advantages of the method used here over the approaches described in previous chapters is that the information from unaffected individuals in the sample is utilized. Additionally, the risk of missing true linkage due to model misspecification is mitigated. When the population parameters of a disease are unknown, as is often the case, one approach is to choose a set of parameters arbitrarily, accepting that the resulting LOD score may be reduced as compared to what would be observed under parameters that better approximate the true population values, but that linkage may be observed nonetheless. The effects of model parameter assumptions on linkage analysis were described by Hodge and Elston (Hodge and Elston, 1994). An alternate approach is to maximize over models, on the premise that model misspecification reduces the resulting LOD score, and that the power to detect true linkage is increased by evaluating a range of parameters. The validity of this method was shown in simulation studies (Greenberg et al., 1998; Abreu et al., 1999).

These analyses provided statistical evidence for reduced penetrance in MacTel, which had long been presumed, but had not been shown. In previous analysis, it was shown that when the

sample was analyzed using all individuals, with unaffected individuals declared as such, the linkage signal on chromosome 1 was greatly reduced under a penetrance value of 0.9. That result suggested that unaffected individuals in the sample do carry genotypes compatible with linkage in the region of interest. Declaring such individuals to be unaffected reduced the evidence for linkage in the region. While it is formally possible that there is in fact no true disease causing allele in the region, that analysis failed to evaluate the alternate hypothesis that the MacTel phenotype does not manifest uniformly or completely in every individual who carries a putative risk allele (or alleles). This is a reasonable and necessary hypothesis to evaluate, given that complete penetrance is likely the exception rather than the rule in inherited disease, and given the fact that variable expressivity has been observed for this disease.

The additional contribution of these analyses beyond the results presented in previous chapters is to present statistical evidence for an autosomal dominant mode of inheritance in the absence of formal segregation analysis, and to present evidence for reduced penetrance in the families analyzed. It is encouraging that the results observed and presented here are consistent with previous observations, bolstering the hypothesis that a true risk allele for MacTel may exist in the region of interest on chromosome 1q41.

REFERENCES

Abecasis, G. R., Cherny, S. S., Cookson, W. O. and Cardon, L. R. (2002) 'Merlin--rapid analysis of dense genetic maps using sparse gene flow trees', *Nat Genet* 30(1): 97-101.

Abreu, P. C., Greenberg, D. A. and Hodge, S. E. (1999) 'Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases', *Am J Hum Genet* 65(3): 847-57.

Albrechtsen, A., Nielsen, F. C. and Nielsen, R. (2010) 'Ascertainment biases in SNP chips affect measures of population divergence', *Mol Biol Evol* 27(11): 2534-47.

Altshuler, D., Daly, M. J. and Lander, E. S. (2008) 'Genetic mapping in human disease', *Science* 322(5903): 881-8.

Antonarakis, S. E. and Beckmann, J. S. (2006) 'Mendelian disorders deserve more attention', *Nat Rev Genet* 7(4): 277-82.

Arjamaa, O. and Nikinmaa, M. (2006) 'Oxygen-dependent diseases in the retina: role of hypoxia-inducible factors', *Exp Eye Res* 83(3): 473-83.

Aung, K. Z., Wickremasinghe, S. S., Makeyeva, G., Robman, L. and Guymer, R. H. (2010) 'The prevalence estimates of macular telangiectasia type 2: the Melbourne Collaborative Cohort Study', *Retina* 30(3): 473-8.

Bhosale, P., Larson, A. J., Frederick, J. M., Southwick, K., Thulin, C. D. and Bernstein, P. S. (2004) 'Identification and characterization of a Pi isoform of glutathione S-transferase (GSTP1) as a zeaxanthin-binding protein in the macula of the human eye', *J Biol Chem* 279(47): 49447-54.

Botstein, D. and Risch, N. (2003) 'Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease', *Nat Genet* 33 Suppl: 228-37.

Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. (1980) 'Construction of a genetic linkage map in man using restriction fragment length polymorphisms', *Am J Hum Genet* 32(3): 314-31.

Campochiaro, P. A. (2000) 'Retinal and choroidal neovascularization', *J Cell Physiol* 184(3): 301-10.

Charbel Issa, P., Berendschot, T. T., Staurenghi, G., Holz, F. G. and Scholl, H. P. (2008a) 'Confocal blue reflectance imaging in type 2 idiopathic macular telangiectasia', *Invest Ophthalmol Vis Sci* 49(3): 1172-7.

Charbel Issa, P., Finger, R. P., Helb, H. M., Holz, F. G. and Scholl, H. P. (2008b) 'A new diagnostic approach in patients with type 2 macular telangiectasia: confocal reflectance imaging', *Acta Ophthalmol* 86(4): 464-5.

Charbel Issa, P., Finger, R. P., Holz, F. G. and Scholl, H. P. (2009a) 'Multimodal imaging including spectral domain OCT and confocal near infrared reflectance for characterization of outer retinal pathology in pseudoxanthoma elasticum', *Invest Ophthalmol Vis Sci* 50(12): 5913-8.

Charbel Issa, P., Helb, H. M., Rohrschneider, K., Holz, F. G. and Scholl, H. P. (2007) 'Microperimetric assessment of patients with type 2 idiopathic macular telangiectasia', *Invest Ophthalmol Vis Sci* 48(8): 3788-95.

Charbel Issa, P., Holz, F. G. and Scholl, H. P. (2009b) 'Metamorphopsia in patients with macular telangiectasia type 2', *Doc Ophthalmol* 119(2): 133-40.

Charbel Issa, P., van der Veen, R. L., Stijfs, A., Holz, F. G., Scholl, H. P. and Berendschot, T. T. (2009c) 'Quantification of reduced macular pigment optical density in the central retina in macular telangiectasia type 2', *Exp Eye Res* 89(1): 25-31.

Chew, E., Gillies, M. and Bird, A. (2006) 'Macular telangiectasia: a simplified classification', *Arch Ophthalmol* 124(4): 573-4.

Chew, E. Y., Murphy, R. P., Newsome, D. A. and Fine, S. L. (1986) 'Parafoveal telangiectasis and diabetic retinopathy', *Arch Ophthalmol* 104(1): 71-5.

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloglu, A., Ozen, S., Sanjad, S. et al. (2009) 'Genetic diagnosis by whole exome capture and massively parallel DNA sequencing', *Proc Natl Acad Sci U S A* 106(45): 19096-101.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K. and Lander, E. S. (2007) 'Distinguishing protein-coding and noncoding genes in the human genome', *Proc Natl Acad Sci U S A* 104(49): 19428-33.

Clark, A. G., Wang, X. and Matise, T. (2010) 'Contrasting methods of quantifying fine structure of human recombination', *Annu Rev Genomics Hum Genet* 11: 45-64.

Clemons, T. E., Gillies, M. C., Chew, E. Y., Bird, A. C., Peto, T., Figueroa, M. J. and Harrington, M. W. (2010) 'Baseline characteristics of participants in the natural history study of macular telangiectasia (MacTel) MacTel Project Report No. 2', *Ophthalmic Epidemiol* 17(1): 66-73.

Coffey, A. J., Kokocinski, F., Calafato, M. S., Scott, C. E., Palta, P., Drury, E., Joyce, C. J., Leproust, E. M., Harrow, J., Hunt, S. et al. (2011) 'The GENCODE exome: sequencing the complete human exome', *Eur J Hum Genet*.

Cutting, G. R. (2010) 'Modifier genes in Mendelian disorders: the example of cystic fibrosis', *Ann N Y Acad Sci* 1214: 57-69.

Daiger, S. P., Bowne, S. J. and Sullivan, L. S. (2007) 'Perspective on genes and mutations causing retinitis pigmentosa', *Arch Ophthalmol* 125(2): 151-8.

den Hollander, A. I., Roepman, R., Koenekoop, R. K. and Cremers, F. P. (2008) 'Leber congenital amaurosis: genes, proteins and disease mechanisms', *Prog Retin Eye Res* 27(4): 391-419.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E. et al. (1996) 'A comprehensive genetic map of the human genome based on 5,264 microsatellites', *Nature* 380(6570): 152-4.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W., Smith, D. R., Lander, E. S. et al. (1987) 'A genetic linkage map of the human genome', *Cell* 51(2): 319-37.

Dowling, John E. (1987) *The Retina - An approachable part of the brain*, Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Finger, R. P., Charbel Issa, P., Fimmers, R., Holz, F. G., Rubin, G. S. and Scholl, H. P. (2009) 'Reading performance is reduced by parafoveal scotomas in patients with macular telangiectasia type 2', *Invest Ophthalmol Vis Sci* 50(3): 1366-70.

Finlay, B. L. (2008) 'The developing and evolving retina: using time to organize form', *Brain Res* 1192: 5-16.

Fisher, R.A. (1922) 'On the mathematical foundations of theoretical statistics', *Philosophical Transactionf of the Royal Society of London. Series A, Containing Papers of a Mathematical of Physical Character* 222: 309-368.

Fulton, A. B., Hansen, R. M., Moskowitz, A. and Akula, J. D. (2009) 'The neurovascular retina in retinopathy of prematurity', *Prog Retin Eye Res* 28(6): 452-82.

Gass, J. D. and Blodi, B. A. (1993) 'Idiopathic juxtafoveolar retinal telangiectasis. Update of classification and follow-up study', *Ophthalmology* 100(10): 1536-46.

Gillies, M. C., Zhu, M., Chew, E., Barthelmes, D., Hughes, E., Ali, H., Holz, F. G., Scholl, H. P. and Charbel Issa, P. (2009) 'Familial asymptomatic macular telangiectasia type 2', *Ophthalmology* 116(12): 2422-9.

Greenberg, D. A., Abreu, P. and Hodge, S. E. (1998) 'The power to detect linkage in complex disease by means of simple LOD-score analyses', *Am J Hum Genet* 63(3): 870-9.

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y. et al. (1983) 'A polymorphic DNA marker genetically linked to Huntington's disease', *Nature* 306(5940): 234-8.

Haldane, J. B. S. (1934) 'Methods for the detection of autosomal linkage in man', *Annals of Eugenics* 6: 26-65.

Hall, W. D., Mathews, R. and Morley, K. I. (2010) 'Being more realistic about the public health impact of genomic medicine', *PLoS Med* 7(10).

Hannan, S. R., Madhusudhana, K. C., Rennie, C. and Lotery, A. J. (2007) 'Idiopathic juxtafoveolar retinal telangiectasis in monozygotic twins', *Br J Ophthalmol* 91(12): 1729-30.

Helb, H. M., Charbel Issa, P., RL, Van Der Veen, Berendschot, T. T., Scholl, H. P. and Holz, F. G. (2008) 'Abnormal macular pigment distribution in type 2 idiopathic macular telangiectasia', *Retina* 28(6): 808-16.

Hodge, S. E. and Elston, R. C. (1994) 'Lods, wrods, and mods: the interpretation of lod scores calculated under different models', *Genet Epidemiol* 11(4): 329-42.

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., Middle, C. M., Rodesch, M. J., Albert, T. J., Hannon, G. J. et al. (2007) 'Genome-wide in situ exon capture for selective resequencing', *Nat Genet* 39(12): 1522-7.

Hutton, W. L., Snyder, W. B., Fuller, D. and Vaiser, A. (1978) 'Focal parafoveal retinal telangiectasis', *Arch Ophthalmol* 96(8): 1362-7.

Iturralde, D., Spaide, R. F., Meyerle, C. B., Klancnik, J. M., Yannuzzi, L. A., Fisher, Y. L., Sorenson, J., Slakter, J. S., Freund, K. B., Cooney, M. et al. (2006) 'Intravitreal bevacizumab (Avastin) treatment of macular edema in central retinal vein occlusion: a short-term study', *Retina* 26(3): 279-84.

Klein, R., Blodi, B. A., Meuer, S. M., Myers, C. E., Chew, E. Y. and Klein, B. E. (2010) 'The prevalence of macular telangiectasia type 2 in the Beaver Dam eye study', *Am J Ophthalmol* 150(1): 55-62 e2.

Lander, E. S. (2011) 'Initial impact of the sequencing of the human genome', *Nature* 470(7333): 187-97.

Li, M., Li, C. and Guan, W. (2008) 'Evaluation of coverage variation of SNP chips for genome-wide association studies', *Eur J Hum Genet* 16(5): 635-43.

Loane, E., Kelliher, C., Beatty, S. and Nolan, J. M. (2008a) 'The rationale and evidence base for a protective role of macular pigment in age-related maculopathy', *Br J Ophthalmol* 92(9): 1163-8.

Loane, E., Nolan, J. M., O'Donovan, O., Bhosale, P., Bernstein, P. S. and Beatty, S. (2008b) 'Transport and retinal capture of lutein and zeaxanthin with reference to age-related macular degeneration', *Surv Ophthalmol* 53(1): 68-81.

McPherson, J. D. Marra, M. Hillier, L. Waterston, R. H. Chinwalla, A. Wallis, J. Sekhon, M. Wylie, K. Mardis, E. R. Wilson, R. K. et al. (2001) 'A physical map of the human genome', *Nature* 409(6822): 934-41.

Menchini, U., Virgili, G., Bandello, F., Malara, C., Rapizzi, E. and Lanzetta, P. (2000) 'Bilateral juxtafoveolar telangiectasis in monozygotic twins', *Am J Ophthalmol* 129(3): 401-3.

Morgan, T. H. (1910) 'Sex Limited Inheritance in Drosophila', *Science* 32(812): 120-2.

Morton, N. E. (1955) 'Sequential tests for the detection of linkage', *Am J Hum Genet* 7(3): 277-318.

Ng, P. C., Levy, S., Huang, J., Stockwell, T. B., Walenz, B. P., Li, K., Axelrod, N., Busam, D. A., Strausberg, R. L. and Venter, J. C. (2008) 'Genetic variation in an individual human exome', *PLoS Genet* 4(8): e1000160.

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A. et al. (2010a) 'Exome sequencing identifies the cause of a mendelian disorder', *Nat Genet* 42(1): 30-5.

Ng, S. B., Nickerson, D. A., Bamshad, M. J. and Shendure, J. (2010b) 'Massively parallel sequencing and rare disease', *Hum Mol Genet* 19(R2): R119-24.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E. et al. (2009) 'Targeted capture and massively parallel sequencing of 12 human exomes', *Nature* 461(7261): 272-6.

Niu, T. (2004) 'Algorithms for inferring haplotypes', *Genet Epidemiol* 27(4): 334-47.

Oh, K. T. and Park, D. W. (1999) 'Bilateral juxtafoveal telangiectasis in a family', *Retina* 19(3): 246-7.

Ott, Jurg (1999) *Analysis of Human Genetic Linkage*: The Johns Hopkins University Press.

Parmalee, N. L., Schubert, C., Merriam, J. E., Allikmets, K., Bird, A. C., Gillies, M. C., Peto, T., Figueroa, M., Friedlander, M., Fruttiger, M. et al. (2010) 'Analysis of candidate genes for macular telangiectasia type 2', *Mol Vis* 16: 2718-26.

Penrose, L. S. (1953) 'The General Purpose Sib-Pair Linkage Test', *Annals of Eugenics* 18(2): 120-124.

Powner, M. B., Gillies, M. C., Tretiach, M., Scott, A., Guymer, R. H., Hageman, G. S. and Fruttiger, M. (2010) 'Perifoveal muller cell depletion in a case of macular telangiectasia type 2', *Ophthalmology* 117(12): 2407-16.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. et al. (2007) 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *Am J Hum Genet* 81(3): 559-75.

Purcell, S. PLINK: Whole genome association analysis tool, version 1.07, http://pngu.mgh.harvard.edu/purcell/plink/. Harvard University; Website accessed Sept. 2011.

Rafiq, S., Anand, S. and Roberts, R. (2010) 'Genome-wide association studies of hypertension: have they been fruitful?', *J Cardiovasc Transl Res* 3(3): 189-96.

Rajappa, M., Saxena, P. and Kaur, J. (2010) 'Ocular angiogenesis: mechanisms and recent advances in therapy', *Adv Clin Chem* 50: 103-21.

Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G. P. and Mariotti, S. P. (2004) 'Global data on visual impairment in the year 2002', *Bull World Health Organ* 82(11): 844-51.

Rommens, J. M., Zengerling, S., Burns, J., Melmer, G., Kerem, B. S., Plavsic, N., Zsiga, M., Kennedy, D., Markiewicz, D., Rozmahel, R. et al. (1988) 'Identification and regional localization of DNA markers on chromosome 7 for the cloning of the cystic fibrosis gene', *Am J Hum Genet* 43(5): 645-63.

Shastry, B. S., Hartzer, M. K. and Trese, M. T. (1993) 'Familial exudative vitreoretinopathy: multiple modes of inheritance', *Clin Genet* 44(5): 275-6.

Siddiqui, N. and Fekrat, S. (2005) 'Group 2A idiopathic juxtafoveolar retinal telangiectasia in monozygotic twins', *Am J Ophthalmol* 139(3): 568-70.

Sturtevant, A.H. (1913) 'The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association', *Journal of Experimental Zoology* 14: 43-59.

Terwilliger, J. D. and Goring, H. H. (2000) 'Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design', *Hum Biol* 72(1): 63-132.

Terwilliger, J. D. and Goring, H. H. (2009) 'Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. 2000', *Hum Biol* 81(5-6): 663-728.

Terwilliger, J. D., Zollner, S., Laan, M. and Paabo, S. (1998) 'Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion', *Hum Hered* 48(3): 138-54.

Venter, J. C. Adams, M. D. Myers, E. W. Li, P. W. Mural, R. J. Sutton, G. G. Smith, H. O. Yandell, M. Evans, C. A. Holt, R. A. et al. (2001) 'The sequence of the human genome', *Science* 291(5507): 1304-51.

Weiss, K. M. and Terwilliger, J. D. (2000) 'How many diseases does it take to map a gene with SNPs?', *Nat Genet* 26(2): 151-7.

Yan, X. J., Xu, J., Gu, Z. H., Pan, C. M., Lu, G., Shen, Y., Shi, J. Y., Zhu, Y. M., Tang, L., Zhang, X. W. et al. (2011) 'Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia', *Nat Genet* 43(4): 309-15.

Yannuzzi, L. A., Bardal, A. M., Freund, K. B., Chen, K. J., Eandi, C. M. and Blodi, B. (2006) 'Idiopathic macular telangiectasia', *Arch Ophthalmol* 124(4): 450-60.

Zhang, H., Akman, H. O., Smith, E. L., Zhao, J., Murphy-Ullrich, J. E. and Batuman, O. A. (2003) 'Cellular response to hypoxia involves signaling via Smad proteins', *Blood* 101(6): 2253-60.

Zhao, L. P., Aragaki, C., Hsu, L. and Quiaoit, F. (1998) 'Mapping of complex traits by single-nucleotide polymorphisms', *Am J Hum Genet* 63(1): 225-40.

Zhou, C., Zang, D., Jin, Y., Wu, H., Liu, Z., Du, J. and Zhang, J. (2011) 'Mutation in ribosomal protein L21 underlies hereditary hypotrichosis simplex', *Hum Mutat*.

Zuchner, S., Dallman, J., Wen, R., Beecham, G., Naj, A., Farooq, A., Kohli, M. A., Whitehead, P. L., Hulme, W., Konidari, I. et al. (2011) 'Whole-exome sequencing links a variant in DHDDS to retinitis pigmentosa', *Am J Hum Genet* 88(2): 201-6.

Bibliography

Research Articles

Macular Telangiectasia Type 2: Linkage to 1q41-42 and Analysis of the Locus. **Parmalee NL**, Schubert C, Figueroa M, Bird AC, Peto T, Gillies MC, Bernstein PS, Kiryluk K, Terwilliger JD, Allikmets R, The MacTel Project. (Submitted to PLoS One).

Analysis of Candidate Genes for Macular Telangiectasia Type 2. **Parmalee NL**, Schubert C, Merriam JE, Allikmets K, Bird AC, Gillies MC, Peto T, Figueroa M, Friedlander M, Fruttiger M, Greenwood L, Moss SE, Smith LE, Toomes C, Inglehearn CF, Allikmets R. (2010) Molecular Vision 14(16): 2718-2726

Wnt5a Signaling Induces Proliferation, Survival and Expression of MMP-1 and Tie-2 in Endothelial Cells. Masckauchán TNH, Agalliu D, Vorontchikhina M, Ahn A, **Parmalee NL**, Li CM, Khoo A, Tycko B, Brown AMC, and Kitajewski J. (2006) Molecular Biology of the Cell 17(12): 5163-5172

Review Articles

Wnt Signaling in Angiogenesis. **Parmalee NL** and Kitajewski J. (2008) Current Drug Targets. 9(7): 558-64 Review. Invited submission.

Abstracts and Presentations

Determining the Genetic Cause of Macular Telangiectasia. **Parmalee NL**, Kiryluk K, Schubert C, Allikmets K, Gillies M, Allikmets R. ARVO 2010.

Genome-wide Linkage Analysis for Genetic Causes of Macular Telangiectasia Type 2. **Parmalee NL**, Kiryluk K, Gillies M, Allikmets R, The MacTel Project. ASHG 2009.

Genetic Studies of Macular Telangiectasia. **Parmalee NL**, Schubert C, Toomes C, Merriam J, Gillies M, Allikmets R, The MacTel Project. ARVO 2009.

Controlled Suppression of RPE65 *in vitro* and *in vivo*: Implications for Treating Stargardt Disease. **Parmalee NL**, Doi K, Kong J, Zernant J, Kim SR, Schorderet DF, Sparrow JR, Allikmets R. ARVO 2008.