Human Mutation

Single base substitutions in the CHM promoter as a cause of choroideremia

Journal:	*Human Mutation*
Manuscript ID	humu-2016-0385.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Radziwon, Alina; University of Alberta, Ophthalmology Arno, Gavin; University College London, Institute of Ophthalmology; Moorfields Eye Hospital Wheaton, Dianna; Retina Foundation of the Southwest, McDonagh, Ellen; Queen Mary University of London, Genomics England Baple, Emma; Queen Mary University of London, Genomics England; University of Exeter, RILD Wellcome Wolfson Centre Webb, Kaylie; Retina Foundation of the Southwest Birch, David; Retina Foundation of the Southwest Webster, Andrew; University College London, Institute of Ophthalmology; Moorfields Eye Hospital MacDonald, Ian; University of Alberta, Ophthalmology
Key Words:	choroideremia, CHM, REP-1, ZNF143, promoter, THAP11

Single base substitutions in the *CHM* promoter as a cause of choroideremia

Alina Radziwon[1*], Gavin Arno[2, 3], Dianna Wheaton[4], Ellen M. McDonagh[5], Emma L. Baple[5,6], Kaylie Webb-Jones[4], David Birch[4], Andrew R. Webster[2, 3], Ian M. MacDonald[1]

[1]Department of Ophthalmology and Visual Sciences, University of Alberta, Edmonton, Canada

[2]UCL Institute of Ophthalmology, University College London, London EC1V 9EL, UK

[3]Moorfields Eye Hospital, London EC1V 2PD, UK

[4] Retina Foundation of the Southwest, Dallas, Texas, United States of America

[5]Genomics England, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK

[6]Medical Research (Level 4), University of Exeter Medical School, RILD Wellcome Wolfson Centre, Royal Devon and Exeter NHS Foundation Trust, Barrack Road, Exeter, EX2 5DW, UK

*Correspondence: Alina Radziwon, Department of Ophthalmology and Visual Sciences, 7-030 Katz Building, University of Alberta, Edmonton, Alberta, Canada, T6G 2E1. Phone: 1-780-492-1141; e-mail: alinar@ualberta.ca

Presented at: The Association for Research in Vision and Ophthalmology Annual Meeting, May 2016, Seattle, WA, USA.

Financial Support: IM: Canadian Institutes of Health Research, Emerging Team Grant: 119190, Foundation Fighting Blindness, Canada, Choroideremia Research Foundation Canada, Inc., and Alberta Innovates-Health Solutions 201201139

DB: National Eye Institute grant EY-09076, Foundation Fighting Blindness

GA, AW: The National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at Moorfields Eye Hospital, and the UCL Institute of Ophthalmology. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health). The 100,000 Genomes Project is funded by the NIHR and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

Abstract

Although over 150 unique mutations affecting the coding sequence of *CHM* have been identified in patients with the X-linked chorioretinal disease choroideremia (CHM), no regulatory mutations have been reported, and indeed the promoter has not been defined. Here we describe two independent families affected by CHM bearing a mutation outside the gene's coding region at position c.-98: C>A and C>T, which segregated with the disease. The male proband of family 1 was found to lack *CHM* mRNA and its gene product Rab escort protein 1 (REP-1), while whole genome sequencing of an affected male in family 2 excluded the involvement of any other known retinal genes. Both mutations abrogated luciferase activity when inserted into a reporter construct, and by further employing the luciferase reporter system to assay sequences 5′ to the gene, we identified the *CHM* promoter as the region encompassing nucleotides c.-119 to c.-76. These findings suggest that the *CHM* promoter region should be examined in patients with choroideremia who lack coding-sequence mutations, and reveals, for the first time, features of the gene's regulation.

Keywords

CHM, choroideremia, promoter, REP-1, ZNF143, THAP11

Introduction

Choroideremia (CHM; MIM# 303100) is an X-linked, recessively inherited chorioretinal dystrophy with an incidence of 1/50,000. Progressive degeneration of photoreceptors, retinal pigment epithelium (RPE), and the choroid causes affected hemizygous males to develop night blindness in the first or second decade of life, followed by a decrease in peripheral visual fields and an eventual loss of central visual acuity in advanced stages of the disease (Coussa and Traboulsi, 2012). Carrier females, while usually asymptomatic, may exhibit signs of retinal degeneration upon fundoscopic examination, and more rarely have reduced dark adaptation and peripheral vision (Karna, 1986; Roberts et al., 2002).

To date, CHM has only been linked to mutations within the *CHM* gene, coding for REP-1 (Cremers et al., 1990). The protein serves as a molecular chaperone for small GTPases from the Rab family, presenting them to Rab geranylgeranyl transferase which modifies them by the covalent attachment of a lipid moiety. The lipid modification, known as prenylation, of target Rab proteins is essential for intracellular vesicular transport (Seabra et al., 1992; Seabra et al., 1993). A homologue, REP-2, encoded by *CHML* or choroideremia-like, functions similarly and appears to compensate for the absence of REP-1 in all tissues except the eye (Cremers et al., 1994). *CHM* spans over 150 kb of Xq21.2 and contains 15 exons (van Bokhoven et al., 1994). A 30 base pair 5'untranslated region, found on exon 1, precedes an open reading frame encoding the 653 amino acid REP-1 protein. At least 163 unique pathogenic mutations have been reported in the LOVD Retinal and Hearing Impairment Genetic Mutation Database [https://grenada.lumc.nl/LOVD2/Usher_montpellier/home.php ; accessed Feb 2017] (Fokkema et al., 2011). The mutation spectrum includes transitions and transversions leading to protein truncation, splice defects, indels and large deletions ranging from a single exon to the full gene.

Missense mutations predicted to alter protein structure or impair function (Sergeev et al., 2009; Esposito et al., 2011), transposon insertions (Van Den Hurk et al., 2003), partial gene duplications (Chi et al., 2013; Simunovic et al., 2016) and other variations are infrequently found, but taken collectively, almost all known pathogenic variants in the *CHM* gene have been loss of function mutations that abolish functional REP-1 (McTaggart et al., 2002; Simunovic et al., 2016). Notably there is no apparent correlation between genotype and phenotype, with the age at onset of symptoms, visual acuity and visual fields being unrelated to mutation type (Freund et al., 2016; Simunovic et al., 2016).

The promoter driving expression of *CHM* has been heretofore unidentified. Analyses of promoter mutations causing inherited diseases can be useful in identifying transcription factors involved in the regulation and expression of a gene of interest, as the activity of RNA polymerase II is mediated by the recruitment of general and sequence specific transcription factors to *cis*-acting regulatory sequences. The core promoter directing basal level transcription is generally found between nucleotides -40 and +50 of the transcription start site (TSS) (de Vooght et al., 2009), while proximal promoter elements typically reside within 1 kilo bases (kb), with additional enhancers, repressors, insulators acting even over distances of mega bases (Maston et al., 2006). The investigation of regulatory elements in the context of human disease is complicated by the fact that in contrast to loss of function mutations, regulatory defects may produce small quantitative changes that are difficult to detect. Mutations in *cis*-acting regulatory sequences are a significant cause of human disease, and according to statistics compiled by the Human Gene Mutation Database [http://www.hgmd.cf.ac.uk ; accessed May 2016] approximately 2% of disease causing point mutations are in non-coding regions of the genome (Stenson et al., 2014). Though promoter defects can cause functionally important consequences

for gene expression, their analysis is often not a regular part of DNA diagnostics as the investigation can be complex, laborious and difficult to perform.

In this study, we report the first known regulatory mutations causing choroideremia. The absence of a genetic defect in a CHM family prompted us to explore the upstream non-coding region, revealing the novel promoter variant c.-98C>A. Subsequently, whole genome sequencing in an unrelated family with CHM identified a second variant of the same residue, c.-98C>T. We demonstrate the mutations' effects on transcription using a luciferase assay, and employ the same system to further characterize the boundary of the *CHM* promoter that spans the location. The results suggest a disruption of a transcription factor binding site and impaired transactivation of the *CHM* promoter by the factor(s). Together, these studies further our understanding of regulation and expression of *CHM* and present a possible explanation for cases of unexplained choroideremia where no causative mutation is found within the gene.

Materials and Methods

Clinical Examination and Study Subjects

This study was approved by the institutions' respective ethics review boards. All procedures adhered to the tenets of the Declaration of Helsinki. Before participation, the purpose and risks of the study were explained, and informed consent was obtained. Blood samples were drawn, and a detailed pedigree and history was recorded.

We studied 4 males affected by CHM and 1 carrier female from two unrelated families. Individuals were examined by fundoscopy and functional ophthalmologic methods including some or all of: visual acuity testing, central visual fields, full field electroretinography, optical

coherence tomography and fundus autofluorescence imaging. Choroideremia was diagnosed by senior clinicians with expertise in inherited retinal disease based on clinical findings alone. Affected males had a history of night vision loss and fundus appearance of peripheral retinopathy with broad areas of RPE and choroidal atrophy.

CHM Sanger Sequencing

DNA was extracted from lymphocytes using conventional methodologies. Proband 5116 from family 1 (C127) was initially screened for *CHM* mutations in the coding sequence and splice sites by direct Sanger sequencing from PCR amplicons (Furgoch et al., 2014). Additionally, 1kb of the 5'-flanking sequence was amplified with primers 5'-CAGGGAAGGCCCACTACTGC -3' and 5'-CTTGTGGAAATGAGATCAAGTTAGG-3' and sequenced with the same primers. With the exception of patient 111 and his parents, of family 2 who underwent whole genome sequencing, the remaining study individuals were genotyped only at position c.-98 to confirm the presence or absence of the respective variant. Annotation is in accordance with GenBank: NM_000390.3, where +1 represents the start of translation.

Cell Culture

Leukocytes from patient 5116 were separated with Ficoll-Histopaque (Sigma-Aldrich, St. Louis, MO, USA) from 10mL of whole human blood collected in acid citrate dextrose (ACD-A) tubes (BD, Franklin Lakes, NJ, USA). A lymphoblastoid cell line was established by Epstein-Barr virus transformation within 48 hours of collection (Anderson and Gusella, 1984). Cells were maintained in RPMI-1640 supplemented with 15% fetal calf serum and penicillin-streptomycin.

mRNA and Protein Analysis

Total RNA was extracted from $5x10^6$ lymphoblastoid cells from patient 5116 and a healthy control, with the NucleoSpin RNA isolation kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. cDNA was transcribed from 5µg of total RNA with the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Rockford, IL, USA) using a random hexamer primer mix. PCR amplification from cDNA with forward primer 5'-TAATAGTCACATGACACGTTTCCCG-3', paired with 5'-TGGATTCAGGCAAACCCGT-3', or 5'-TTTAAAATGAGCAAGTCAATGTGC-3' as a reverse primer was used to detect partial transcript (spanning the junction of exon 1 and 2), or the entire coding sequence of *CHM* encompassed by 15 exons, respectively. GAPDH was also amplified as a positive control with intron-spanning primers.

Protein was extracted from $2x10^6$ lymphoblastoid cells and immunoblot analysis of REP-1 was performed as previously described (Furgoch et al., 2014).

Whole Genome Sequencing

UK proband 111 (GC406) and his unaffected parents 190 and 192 underwent whole genome sequencing (WGS) as part of the 100,000 Genomes Project. Briefly, genomic DNA was processed using the Illumina TruSeq DNA PCR-Free Sample Preparation kit (Illumina Inc) and sequenced using an Illumina HiSeq X Ten, generating minimum coverage of 15X for >97% of the callable autosomal genome. Reads were aligned to build GRCh37 of the human genome using the Isaac aligner (Illumina Inc). SNVs and indels were identified using Platypus v0.8.1 and annotated using Cellbase (https://github.com/opencb/cellbase). Variant filtering was performed using minor allele frequency (MAF) in publicly available and in-house datasets, predicted

protein impact and familial segregation. Surviving variants were prioritized using a pre-specified virtual gene panel from PanelApp (https://panelapp.extge.co.uk/crowdsourcing/PanelApp/ Posterior segment abnormalities v1.7. Allelic state was required to match the curated mode of inheritance for variants in panel genes.

Promoter Analysis

Construction of luciferase reporter plasmids

Primers were initially designed to amplify a 1063 bp fragment from upstream of the *CHM* transcription start site (c.-1093_c.-31) from control and patient 5116's DNA. The c.-98C>T mutation present in family 2 was generated by site directed mutagenesis from the control plasmid. All sequences cloned contained only untranscribed DNA and did not include the 30 base pair 5'UTR directly upstream of the start codon. The fragment was cloned according to standard protocols, with SacI and HindIII restriction sites added to primers allowing digest by the enzymes (Thermo Scientific, Rockford, IL, USA). Ligation into the promoterless firefly luciferase reporter vector, pGL3-Basic (Promega, Madison, WI, USA) generated constructs pGL3-1063_-31 and pGL3-1063_-31 c.-98 C>A. Subsequently, a 4kb fragment, and a series of nested deletions ranging from 2kb to 44bp were cloned into pGL3-Basic with the same restriction sites added and used for subsequent digest, to identify the minimal promoter necessary and sufficient for transcription. pGL3-Basic lacks a eukaryotic promoter sequence upstream of the reporter luciferase gene, and expression of luciferase in transfected cells depends on a functional promoter sequence to be inserted upstream of the *luc+* gene. pGL3-Basic itself served as a negative control. A schematic of all inserts assayed is provided in Figure 1A. All constructs

generated were Sanger sequenced to ensure fidelity. Primer sequences for the generation of plasmid inserts are available upon request.

Luciferase Assay

HEK293T cells cultured with standard reagents and conditions were seeded in 24 well plates and transfected at 70% confluency with polyethylenimine (PEI). Each well was transfected with a total of 650ng of plasmid DNA, including 600ng of pGL3 construct and 50ng of the internal control pRL-CMV, a cytomegalovirus promoter driven *Renilla* luciferase reporter vector. Briefly, plasmid DNA was diluted into 25µL of PBS, and mixed by vortexing with 1µL of 1mg/mL branched PEI (Sigma-Aldrich, St. Louis, MO, USA) in 25µL PBS. After a 15 minute incubation, 50µL of the reaction was added dropwise to each well. Forty-eight hours after transfection, luciferase activity was assayed using the Dual-Luciferase Reporter Assay kit and measured with the Glomax Explorer (both Promega, Madison, WI, USA). Relative luciferase activity was obtained by dividing the relative light units (RLU) produced by the firefly luciferase pGL3 construct by the RLUs produced by *Renilla* luciferase control reaction. Light generated by the reaction can be correlated with the amount of luciferase protein produced which in turn is proportional to promoter activity driving the gene's expression. Values from an n=6 were averaged and normalized to that of the reference construct, pGL3-1093_-31 to obtain a relative measure of activity.

Statistical analysis

All data were expressed as mean ± SD. Data from the luciferase assay represents two independent experiments with triplicate measurements. Differences between groups were examined for statistical significance using Student's *t*-test. A *P*-value <0.01 denoted the presence of a statistically significant difference.

Results

Clinical Characterization of CHM

Family 1

We investigated a progressive retinal degeneration in a Caucasian family of American origin (C127). Remarkably, the proband 5116 was the offspring of a consanguineous union between second cousins, with an affected father and carrier mother. To our knowledge, this has never been reported with choroideremia, and it necessitated a thorough and accurate diagnosis of X-linked CHM and exclusion of an autosomal recessive, or even an unusual male-male transmission as cause for the disease. An investigation of the extended family for which the pedigree is reported in Figure 2A however, clearly demonstrated the X-linked inheritance of the disease.

Patient 5116 was initially seen at age 56 years, when he was referred by a retinal specialist with a diagnosis of choroideremia. At his most recent examination, age 76, best corrected visual acuity (BCVA) was measured as 6/19 OD and 6/15 OS. Visual fields were reduced to less than 5 degrees. The full-field ERG showed non-detectable dark adapted rod-driven responses as well as non-detectable cone responses to a 30 Hz flicker stimulus. OCT imaging indicated loss of the photoreceptor layer across the periphery of the fundus with only a small island of RPE remaining in the macula. Posterior segment examination showed a hypo-pigmented fundus with significant atrophy of the RPE and choroid with areas of bare sclera (Figure 3). These findings are consistent with an advanced state of choroideremia.

The proband's daughter 5113 was examined at 33 years of age and did not report any vision difficulties. BCVA when examined was 6/6, both eyes. Full-field ERG testing of the left eye showed dark adapted rod-driven responses were reduced by 50% in amplitude (42.5 μV) and a normal b-wave implicit time (81.6 msec). Light-adapted cone-driven responses to a 30 Hz flicker stimulus were reduced by 50% in amplitude (27.6 μV) and borderline reduced in implicit time (25.6 msec). These findings were consistent with a classic carrier state of choroideremia.

The proband's father and one brother 5149 were reported to have been previously diagnosed with choroideremia, while a second brother 5147 was reported to be unaffected. Findings are summarized in Table 1.

Family 2:

Family 2 (GC406) was a Caucasian family of British origin with a history of choroideremia. Proband 111 was first examined at age 13, displaying a reduced but not delayed full field ERG with flicker and bright flash responses both within normal limits. He remained asymptomatic until age 32. Confrontational visual field testing at age 35 showed bilateral infratemporal scotoma, and BCVA was 6/5 in both eyes. He is considered symptomatically mild with retained central macular structure.

Of his two maternal uncles, 151 was diagnosed at age 12. Fundus abnormalities typical of choroideremia were noted. 161 was diagnosed at age 8, and upon examination at age 37 exhibited moderately constricted visual fields particularly in the superior field bilaterally. BCVA was reduced to count fingers vision OD and 6/36 OS. Fundus examination was consistent with a clinical diagnosis of choroideremia. The pedigree is reported in Figure 2B, and clinical findings summarized in Table 1.

Genetic Analysis

Family 1 (C127)

Genetic analysis of proband 5116 did not reveal any pathogenic mutation in the coding sequence of the *CHM* gene or splice site boundaries. Yet at position c.-98 relative to the translation start site, a hemizygous C>A transversion was detected (hg38, chrX:g.86047629G>T NM_000390.3, c.-98C>A). The variant was not listed in the latest release of dbSNP [http://www.ncbi.nlm.nih.gov/SNP/; accessed Feb 2017] (Sherry et al., 2001). The proband's affected brother (5149) was also found to harbour the variant while it was absent in the unaffected brother (5147) available for testing. The obligate carrier status was confirmed in the proband's daughter (5113). The location of the variant strongly suggested a regulatory mutation, as evaluation of entries in the HGMD reveals that most promoter mutations are located between +50 and -500 from the TSS of a gene (Stenson et al., 2014).

Family 2 (GC406):

Prior genetic analysis of the coding exons of the *CHM* gene did not reveal a pathogenic mutation. Whole genome sequencing was performed on proband 111 and his parents 190 and 192, as part of the 100,000 Genomes Project. After variant filtering, no causative rare coding variants were identified in any retinal disease gene.

In light of the clinical diagnosis of choroideremia in the family, the complete *CHM* gene was interrogated for rare variants (≤0.001 MAF in 1Kgenome project and internal cohort of over 2000 whole genome sequencing samples) hemizygous in the proband and carried by his mother. One such variant was identified, the transition, c.-98C>T (hg38, chrX:g.86047629G>A NM_000390.3, c.-98C>T). The variant was confirmed in the two affected maternal uncles 151 and 161 by direct Sanger sequencing.

Molecular diagnosis of choroideremia in 5116

Immunoblot analysis of protein harvested from a cell line derived from patient 5116 failed to detect the *CHM* gene product REP-1, providing conclusive confirmation of the clinical diagnosis of choroideremia (Figure 4A). To provide evidence for a regulatory mutation and subsequently reduced transcription, we intended to compare the level of expression between normal and patient samples through qPCR. Endpoint PCR from a cDNA template; however, failed to amplify the 15 exon, 2200 base pair transcript, indicating it was absent or present at a level below the detection threshold, and thus the quantitative assay was not performed. To largely rule out the possibility of a splice defect, we also attempted to amplify a minimal portion of the transcript, a 93 base pair fragment from the 5'-UTR to a region spanning the boundary of exon 1 and 2, and found the patient's cells lacked even this short fragment (Figure 4B). A control housekeeping gene was nevertheless readily detected and both partial and full length transcripts were amplified from normal cDNA.

Effect of the c.-98C>A and c.-98C>T mutations on transcription of CHM

As a starting point, an approximately 1kb fragment upstream of the TSS was assayed for ability to drive gene expression, since the majority of elements necessary for transcription are expected to be found within this region (Rockman and Wray, 2002). Comparing the robust luciferase activity produced by cells transfected with this wild type construct pGL3-1093_-31, to that of pGL3-1093_-31 c.-98C>A and c.-98C>T, we observed complete abrogation of promoter activity (Figure 1B) in the mutants. The drop from 100 ± 9.5 to 2.0 ± 0.3 and 1.3 ± 0.4 respectively, as measured in normalized relative light units (RLU) ± 1 standard deviation is even significantly

lower than that of the negative control pGL3-basic which does not contain a promoter sequence, reading at 5.8 ± 0.9. This startling observation strongly suggested that the mutation spans an element essential to promoter activity.

Based on unpublished reports suggesting other activating elements may exist up to 2.8kb upstream of the TSS (Kaiser NW, *et al.* IOVS 2004;45:ARVO E-Abstract 2451), we additionally compared the activity of two larger constructs approximately 4kb and 2kb long. Typically, regulatory regions upstream of core promoter sequences contain multiple TF specific binding motifs, where several copies of the same factor or cooperation between different factors serve to synergistically stimulate transcriptional activity of a given gene (Maston et al., 2006). Yet in the case of *CHM*, the longer sequences did not stimulate additional luciferase expression. pGL3-3983_-31 and pGL3-2027_-31 did not significantly differ from pGL3-1093_-31, producing 83.7 ± 4.9 and 96.2 ± 7.6 RLUs respectively. The somewhat diminished expression as compared to the reference 1kb construct may be attributed to a lower molar amount of the larger plasmids being delivered during transfection.

Characterization of CHM promoter

To further characterize the promoter, we proceeded to assay progressively shorter constructs, deleting sequences from the -5' and -3' of the 1kb construct to define its boundaries (Figure 1A). Working from the 5' end, the approximately 400, 300, and 100 base pair constructs pGL3-437_-31, pGL3-346_-31, pGL3-119_-31 did not significantly differ from the 1 kb containing reference plasmid pGL3-1093_-31, reading at 100.2 ± 11.4, 96.1 ± 10.6, and 97.7 ± 8.1 RLU respectively. Yet upon deleting a further 11 base pairs, we observed a significant drop to 39.8 ± 7.6 RLU with pGL3-108_-31. Having delineated the -5' boundary of the promoter as extending to no further

than nucleotide c.-119, we subsequently tested pGL3-119_-76 bearing a deletion from the -3' end, and found no significant difference at 93.4 ± 14 RLU. It was the final construct, pGL3-119_-82, which yielded significantly lower activity producing only 27.4 ± 4 RLU. Though further characterization of the promoter region down to the single nucleotide level could be performed, we have defined the borders of the promoter to within the region of c.-119 to c.-76. This short 44 base pair DNA region upstream of the *CHM* coding sequence is able to stimulate transcription in an *in-vitro* assay at a level that is not significantly different from that of a 1kb, or even a 4kb fragment and can be understood to contain essential *cis*-acting elements that positively regulate *CHM* expression.

Discussion

To date, all genetic defects causing choroideremia, which include all major mutation classes, have been observed in exons, introns or intron/exon boundaries of the *CHM* gene (Fokkema et al., 2011). No mutations have been reported in the promoter region of *CHM*, and the regulation of the gene has remained essentially unexplored. In this study, we present the first report of promoter mutations, c.-98C>A and c.-98C>T, causing choroideremia. The critical role of this residue for gene expression is highlighted by the complete abrogation of reporter activity in a promoter assay when the nucleotide is mutated. Though we demonstrate the first examples in a novel mutation class, it is worthwhile noting that the phenotypes observed in affected individuals were typical of choroideremia. In family 1, the genotype c.-98C>A in proband 5116 resulted in severe retinopathy, while characteristic milder signs were found in the carrier female 5113. Males from family 2 bearing the c.-98C>T mutation manifested with chorioretinal disease, but with varying degrees of severity; proband 111 remained symptomatically mild while in his

maternal uncles 151 and 161, disease progression was more typical. CHM is known to exhibit a wide interfamilial, and also intrafamilial variability (Moosajee et al., 2014). As the mutations c.-98C>A and c.-98C>T appear to completely abolish transcription, producing no detectable levels of *CHM* mRNA and REP-1 protein in patient cells, it is not surprising for the disease to present in a classical manner consistent with phenotypes observed for the loss of function mutations. Whole genome sequencing in the parent-offspring trio of family 2 allowed the exclusion of a deep intronic *CHM* mutation, as well as mono or biallelic mutations in any other retinal dystrophy gene as a cause of the observed symptoms. We therefore conclude that the two mutations at residue c.-98 impair *CHM* transcription enough to result in choroideremia. For the several gene therapy trials currently underway employing a "gene replacement" strategy (Dimopoulos et al., 2015), individuals bearing these genotypes would be suitable candidates. The two variants have been submitted and published in the LOVD Retinal and Hearing Impairment Genetic Mutation Database (Fokkema et al., 2011).

Having identified mutations that evidently abolished transcription, we interpreted their location to span a region crucial for gene expression, and set out to define the boundaries of the gene's promoter. Surprisingly, after analyzing fragments as long as 4kb, we found a short 44 base pair DNA fragment to be wholly responsible for driving expression. The region c.-119 to c.-76 comprises the entirety of the *CHM* proximal promoter, and is able to drive robust transcription in an *in-vitro* luciferase assay. The region implicated amounts to less than 3% of the length of the *CHM* coding sequence where the large majority of reported causative mutations have been found (Fokkema et al., 2011). A recent investigation of a large disease cohort (74), found causative mutations in the gene in 94% of cases previously diagnosed with CHM (Simunovic et al., 2016).

The remaining 6% can be understood to be comprised of regulatory mutations, incorrect diagnoses or deep intronic variant causing cryptic splicing (Carss et al., 2017). Promoter defects are therefore expected to be responsible for a small minority of choroideremia cases. The identified region; however, becomes an obvious area for examination in patients in whom no coding sequence mutation is found.

A promoter or regulatory mutation can be expected to either increase or decrease transcriptional activity mediated by the altered binding capacity of trans-acting protein factors specific to a DNA sequence in the promoter region. In this case, the interaction appears to be entirely disrupted based on the null expression of a reporter driven by mutated sequence in our luciferase assay. Impaired transcription due to a 1-bp mutation in a promoter region is unusual, but has been reported previously as in the case of single base mutations stimulating additional transcriptional activity at the *OVOL2* promoter causing autosomal-dominant corneal endothelial dystrophies (Davidson et al., 2016). On the other hand, decreased transactivation of *NMNAT1* due to a single nucleotide change in the promoter was found to cause Leber congenital amaurosis (Coppieters et al., 2015).

While we delineate the *CHM* promoter boundary to this small area upstream of the gene, we cannot exclude the possibility of distant enhancer, repressor, or intronic elements also contributing to regulation; our findings suggest the sequence between c.-119 and c.-76 is essential, but not necessarily sufficient for transcription. The wider upstream sequence of *CHM* lacks the consensus sequences often found in RNA polymerase II promoters, such as CAAT and TATA boxes; as well it is neither GC rich nor associated with any CpG islands. Bioinformatic

analysis with MotifMap, a dataset of computationally predicted transcription factor binding sites based on binding motifs [http://motifmap.ics.uci.edu/] (Daily et al., 2011) identifies a putative binding motif for the transcription factor zinc finger protein 143 (ZNF143) contained within the region of the now revealed *CHM* promoter. ZNF143 participates in the regulation of RNA pol II and III mediated transcription of protein coding, non-coding, and small nuclear RNA (Schaub et al., 1997; Myslinski et al., 1998) and was initially connected with the binding motif TTCCCATTATGCACCGCG (SBS1) (Myslinski et al., 2006). Genome-wide studies revealed the binding site for the factor to be frequently found with an adjacent 5' accessory sequence, forming the ACTACAATTCCCATTATGCACCGCG (SBS2) motif. SBS2 is comprised of both a THAP domain-containing protein 11 (THAP11) and ZNF143 binding site, with the factors believed to act in a competitive manner (Ngondo-Mbongo et al., 2013). The recruitment of THAP11 to its canonical binding site ACTAYRNNNCCCR is most frequently associated with up-regulation genes essential to protein biosynthesis and energy production (Dejosez et al., 2010). More recently, ZNF143 was suggested to cooperatively occupy SBS2 sites with THAP11 and a third factor, the scaffold protein host cell factor 1 (HCFC1) *in-vivo* (Vinckevicius et al., 2015). SBS2 is closely matched by the sequence found at position c.-108 to c.-84 upstream of *CHM*, ACTACAACACCCAGAATGCACTGTT. Notably, ZNF143's binding at promoters was recently implicated in chromatin looping with distal regulatory elements (Bailey et al., 2015), suggesting the involvement of yet other factors in the total regulation of expression of *CHM.*

Of the three transcription factors implicated above, the binding of at least ZNF143 is supported by publicly available chromatin immunoprecipitation sequencing data released as part of the ENCODE project [https://genome.ucsc.edu/ENCODE/] (ENCODE Project Consortium, 2012).

In all cell types assayed: lymphoblastoid, HeLa and K562 cell lines, as well as embryonic stem cells, the promoter region of *CHM* interacted with ZNF143, for which a consensus binding sequence GAACTACAATTCCCAGAAGGC, again is closely matched by GAACTACAACACCCAGAATGC found between position c.-110 toc.-91 relative to the A of the *CHM* start codon. The position-weight matrix for ZNF143 (Figure 5) establishes the relative frequency of the base C at position c.-98 to be 100%, supporting the importance of the residue and the resulting pathogenesis when mutated. Furthermore, surveying the UCSC genome browser (http://genome.ucsc.edu/) multiz alignment of 100 vertabrate genomes tract, the region c.-119 to c.-76 shows a high degree of conservation among mammals, and absolute conservation of residue C at positions corresponding to c.-98 pointing to an important biological role for the sequence (Kuhn et al., 2007; Blanchette et al., 2004). Representatives of birds and amphibian classes, however, lack homology in the region, in while in fish a corresponding region is absent altogether. The promoter, therefore, cannot be considered an ultra-conserved non-coding element (Dimitrieva and Bucher, 2013). Several alignments are listed in Table 2.

Studies of *CHM* mRNA and protein localization have found a broad expression profile for both. In mice, evidence of transcription was found in multiple cell types and in every major layer of retina (Keiser et al., 2005), while immunolabeling of primate retina showed REP-1 localized to both rod and cones (Dimopoulos et al., 2015). Studies in human and primate retina found that mRNA levels did not correspond to the pattern of disease expression; little *CHM* was detected in the RPE and choroid, and there were no marked regional differences in the concentration of *CHM* mRNA apparent with foveal versus mid-peripheral total RNA despite affected males typically exhibiting a preservation of central vision until late in the disease (Bernstein and Wong,

1998). Additionally, REP-1 can be readily detected in human fibroblasts or peripheral blood mononuclear cells (Furgoch et al., 2014; MacDonald et al., 1998). Taken together with ZNF143's characterization as one of the most common and ubiquitously expressed TFs (Myslinski et al., 1998) a picture emerges of widespread and non-specific transcription of *CHM,* despite choroideremia's manifestation as an ocular disease. Indeed, patients' apparent lack of systemic symptoms can be understood to result not from tissue specific expression of REP-1, but from the differing affinities of REP-1 and REP-2 for target Rabs, which may themselves be differentially expressed or possess tissue or cell specific activity. Investigators have implicated Rab27 (Seabra et al., 1995) and Rab38 (Kohnke et al., 2013) as possible contributors.

The study presented here also poses interesting questions, such as whether mutations of other residues less critical to transactivator binding in the *CHM* promoter, that diminish, but not completely abolish mRNA expression, can result in a milder phenotype, or a different rate of progression of choroideremia. Currently, the dbSNP database lists no known SNPs between c.-119 to c.-76 (Sherry et al., 2001). Having shown the *CHM* region responsible for regulation of its expression, described for the first time the features of its promoter, and extended the inventory of molecular changes causing choroideremia, the findings are of clinical and diagnostic interest and present an obvious area of examination for patients with CHM in whom no coding sequence mutation has been found. Further elucidating the roles of ZNF143, THAP11, HCFC1 or other distal factors will prove an important step toward understanding the complete picture of *CHM*'s regulation.

Acknowledgements

The authors have no proprietary or commercial interest in any materials discussed in this article.

Support by Canadian Institutes of Health Research, Emerging Team Grant: 119190, Foundation Fighting Blindness, Canada, Choroideremia Research Foundation Canada, Inc., and Alberta Innovates-Health Solutions 201201139 is acknowledged (IM).

This project was also supported by The National Institute for Health Research (NIHR) and Biomedical Research Centre (BRC) at Moorfields Eye Hospital and the UCL Institute of Ophthalmology (GA, AW).

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health). The 100,000 Genomes Project is funded by the NIHR and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

Support by the National Eye Institute grant EY-09076, and Foundation Fighting Blindness is also acknowledged (DB)

Figure Legends

Figure 1. Functional Analysis of the *CHM* Promoter.

(A) The constructs portrayed on the left were inserted upstream of the luciferase gene in pGL3-basic. Nucleotide +1 represents the translation start site. The promoterless pGL3-basic served as a negative control. **(B)** The mutations c.-98C>A, and c.-98C>T abolish transcription, while the minimal construct c.-119_-76 is sufficient for robust expression of the reporter gene not significantly different from even that of the nearly 4kb construct. All constructs were transiently transfected into HEK293T cells. A dual-luciferase reporter assay was used to assess the potential promoter activity of various sized inserts and the c.-98 mutants. Promoter activity is shown as a ratio of firefly luciferase over *Renilla* luciferase present on the transfection control plasmid pRL-CMV to account for inter-well variation. Activity is normalized to that of the reference construct pGL3-1093_-31 which is artificially set to equal 100. Activity significantly different ($P<0.01$) from the reference construct is denoted by an asterisk. Error bars represent ± 1 SD.

Figure 2. Pedigree Structure of Affected Families with *CHM* Promoter Mutations

(A) Family1. The parents of proband 5116 were second cousins, sharing a set of great grand-parents. The inheritance pattern mimics male-male transmission, but is nevertheless consistent with X-linked inheritance upon examination of the wider family pedigree. **(B)** Family 2. The pedigree showing three generations affected by choroideremia examined in this study. Inheritance follows an X-linked pattern.

Figure 3. Retinal features in patient 5116.

Legend: Left column (OD), right column (OS). **(A)** Fundus photographs of the proband taken at age 76 showing typical choroideremia changes, with atrophy of the choroid and RPE. A small island of preserved RPE remains in the central macula, surrounded by atrophic peripheral areas of apparent bare sclera. **(B)(C)** Fundus autofluorescence image demonstrating areas of residual RPE (L) and the corresponding OCT image (R). Preserved retinal areas with normal

autofluorescence exhibit thicker choroid and preserved retinal lamination. An outer retinal tubulation is seen in the right fundus.

Figure 4. Molecular confirmation of choroideremia in patient 5116.

(A) Patient 5116 lacks REP-1. Western blot results show the absence of a ~100 kDa band corresponding to REP-1 in a lymphoblastoid cell line derived from the patient (lane 2), which is present in a normal control (lane 1). A β-actin antibody was used as a loading control to ensure an adequate protein sample in each lane, with the 42 kDa band present in both samples. **(B)** Patient 5116 lacks *CHM* mRNA. cDNA synthesized from the mRNA harvested from a patient generated lymphoblastoid cell line was used as template for PCR. Lanes 2 and 3 show a 475 bp band resulting from the amplification of the *GAPDH* control housekeeping gene from 5116 and a normal control, indicating cDNA of adequate quality. Lanes 6 and 9 demonstrate an absence of amplification from the patient's cDNA of both partial and full length coding sequence, respectively, as compared to PCR products sized 93 and 2200 base pairs amplified from normal cDNA observed in lanes 5 and 10.

Figure 5. Consensus binding sequences for transcription factor ZNF143.

(A) Partial map of the human *CHM* gene; arrow indicates transcription start site. **(B)** Expanded sequence of the minimal *CHM* promoter from c.-119 to c.-76, as identified through the analysis of progressive deletion constructs. Position c.-98 is marked with an asterisk. **(C)** Sequence logo derived from publically available ChIP-seq data released as part of the ENCODE project, with the position weighted matrix below (ENCODE Project Consortium, 2012). An invariant C is found at position corresponding to c.-98 of *CHM*.

Table 1. Clinical data of genotyped individuals

Table 2. Multiz alignment of the promoter region of 15 *CHM* orthologs

References

Anderson MA, Gusella JF. 1984. Use of cyclosporin A in establishing epstein-barr virus-transformed human lymphoblastoid cell lines. In Vitro 20:856-858.

Bailey SD, Zhang X, Desai K, Aid M, Corradin O, Cowper-Sal Lari R, Akhtar-Zaidi B, Scacheri PC, Haibe-Kains B, Lupien M. 2015. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nat Commun 2:6186.

Bernstein SL, Wong P. 1998. Regional expression of disease-related genes in human and monkey retina. Mol Vis 4:24.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14:708-715.

Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S, Plagnol V, Penkett C, Stirrups K, Rizzo R, Wright G, Josifova D, Bitner-Glindzicz M, Scott RH, Clement E, Allen L, Armstrong R, Brady AF, Carmichael J, Chitre M, Henderson RH, Hurst J, MacLaren RE, Murphy E, Paterson J, Rosser E, Thompson DA, Wakeling E, Ouwehand WH, Michaelides M, Moore AT, NIHR-BioResource Rare Diseases Consortium, Webster AR, Raymond FL. 2017. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am J Hum Genet 100:75-90.

Chi JY, MacDonald IM, Hume S. 2013. Copy number variant analysis in CHM to detect duplications underlying choroideremia. Ophthalmic Genet 34:229-233.

Coppieters F, Todeschini AL, Fujimaki T, Baert A, De Bruyne M, Van Cauwenbergh C, Verdin H, Bauwens M, Ongenaert M, Kondo M, Meire F, Murakami A, Veitia RA, Leroy BP, De Baere E. 2015. Hidden genetic variation in LCA9-associated congenital blindness explained by 5'UTR mutations and copy-number variations of NMNAT1. Hum Mutat 36:1188-1196.

Coussa RG, Traboulsi EI. 2012. Choroideremia: A review of general findings and pathogenesis. Ophthalmic Genet 33:57-65.

Cremers FP, Armstrong SA, Seabra MC, Brown MS, Goldstein JL. 1994. REP-2, a rab escort protein encoded by the choroideremia-like gene. J Biol Chem 269:2111-2117.

Cremers FP, van de Pol DJ, van Kerkhoff LP, Wieringa B, Ropers HH. 1990. Cloning of a gene that is rearranged in patients with choroideraemia. Nature 347:674-677.

Daily K, Patel VR, Rigor P, Xie X, Baldi P. 2011. MotifMap: Integrative genome-wide maps of regulatory motif sites for model species. BMC Bioinformatics 12:495-2105-12-495.

Davidson AE, Liskova P, Evans CJ, Dudakova L, Noskova L, Pontikos N, Hartmannova H, Hodanova K, Stranecky V, Kozmik Z, Levis HJ, Idigo N, Sasai N, Maher GJ, Bellingham J, Veli N, Ebenezer ND, Cheetham ME, Daniels JT, Thaung CM, Jirsova K, Plagnol V, Filipec M, Kmoch S, Tuft SJ, Hardcastle AJ. 2016. Autosomal-dominant corneal endothelial dystrophies CHED1 and PPCD1 are allelic disorders caused by non-coding mutations in the promoter of OVOL2. Am J Hum Genet 98:75-89.

de Vooght KM, van Wijk R, van Solinge WW. 2009. Management of gene promoter mutations in molecular diagnostics. Clin Chem 55:698-708.

Dejosez M, Levine SS, Frampton GM, Whyte WA, Stratton SA, Barton MC, Gunaratne PH, Young RA, Zwaka TP. 2010. Ronin/hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. Genes Dev 24:1479-1484.

Dimitrieva S, Bucher P. 2013. UCNEbase--a database of ultraconserved non-coding elements and genomic regulatory blocks. Nucleic Acids Res 41:D101-9.

Dimopoulos IS, Chan S, MacLaren RE, MacDonald IM. 2015. Pathogenic mechanisms and the prospect of gene therapy for choroideremia. Expert Opin Orphan Drugs 3:787-798.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489:57-74.

Esposito G, De Falco F, Tinto N, Testa F, Vitagliano L, Tandurella IC, Iannone L, Rossi S, Rinaldi E, Simonelli F, Zagari A, Salvatore F. 2011. Comprehensive mutation analysis (20 families) of the choroideremia gene reveals a missense variant that prevents the binding of REP1 with rab geranylgeranyl transferase. Hum Mutat 32:1460-1469.

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. 2011. LOVD v.2.0: The next generation in gene variant databases. Hum Mutat 32:557-563.

Freund PR, Sergeev YV, MacDonald IM. 2016. Analysis of a large choroideremia dataset does not suggest a preference for inclusion of certain genotypes in future trials of gene therapy. Mol Genet Genomic Med 4:344-358.

Furgoch MJB, Mewes-Arès J, Radziwon A, MacDonald IM. 2014. Molecular genetic diagnostic techniques in choroideremia. Mol Vision 20:535-544.

Karna J. 1986. Choroideremia. A clinical and genetic study of 84 finnish patients and 126 female carriers. Acta Ophthalmol Suppl 176:1-68.

Keiser NW, Tang W, Wei Z, Bennett J. 2005. Spatial and temporal expression patterns of the choroideremia gene in the mouse retina. Mol Vision 11:1052-1060.

Kohnke M, Delon C, Hastie ML, Nguyen UT, Wu YW, Waldmann H, Goody RS, Gorman JJ, Alexandrov K. 2013. Rab GTPase prenylation hierarchy and its potential role in choroideremia disease. PLoS One 8:e81758.

Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pedersen JS, Hsu F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baertsch R, Haussler D, Kent WJ. 2007. The UCSC genome browser database: Update 2007. Nucleic Acids Res 35:D668-73.

MacDonald IM, Mah DY, Ho YK, Lewis RA, Seabra MC. 1998. A practical diagnostic test for choroideremia. Ophthalmology 105:1637-1640.

Maston GA, Evans SK, Green MR. 2006. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet 7:29-59.

McTaggart KE, Tran M, Mah DY, Lai SW, Nesslinger NJ, MacDonald IM. 2002. Mutational analysis of patients with the diagnosis of choroideremia. Hum Mutat 20:189-196.

Moosajee M, Ramsden SC, Black GC, Seabra MC, Webster AR. 2014. Clinical utility gene card for: Choroideremia. Eur J Hum Genet 22:10.1038/ejhg.2013.183. Epub 2013 Aug 21.

Myslinski E, Gerard MA, Krol A, Carbon P. 2006. A genome scale location analysis of human staf/ZNF143-binding sites suggests a widespread role for human staf/ZNF143 in mammalian promoters. J Biol Chem 281:39953-39962.

Myslinski E, Krol A, Carbon P. 1998. ZNF76 and ZNF143 are two human homologs of the transcriptional activator staf. J Biol Chem 273:21998-22006.

Ngondo-Mbongo RP, Myslinski E, Aster JC, Carbon P. 2013. Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11. Nucleic Acids Res 41:4000-4014.

Roberts MF, Fishman GA, Roberts DK, Heckenlively JR, Weleber RG, Anderson RJ, Grover S. 2002. Retrospective, longitudinal, and cross sectional study of visual acuity impairment in choroideraemia. Br J Ophthalmol 86:658-662.

Rockman MV, Wray GA. 2002. Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol 19:1991-2004.

Schaub M, Myslinski E, Schuster C, Krol A, Carbon P. 1997. Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. Embo J 16:173-181.

Seabra MC, Brown MS, Goldstein JL. 1993. Retinal degeneration in choroideremia: Deficiency of rab geranylgeranyl transferase. Science 259:377-381.

Seabra MC, Brown MS, Slaughter CA, Sudhof TC, Goldstein JL. 1992. Purification of component A of rab geranylgeranyl transferase: Possible identity with the choroideremia gene product. Cell 70:1049-1057.

Seabra MC, Ho YK, Anant JS. 1995. Deficient geranylgeranylation of ram/Rab27 in choroideremia. J Biol Chem 270:24420-24427.

Sergeev YV, Smaoui N, Sui R, Stiles D, Gordiyenko N, Strunnikova N, MacDonald IM. 2009. The functional effect of pathogenic mutations in rab escort protein 1. Mutat Res 665:44-50.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: The NCBI database of genetic variation. Nucleic Acids Res 29:308-311.

Simunovic MP, Jolly JK, Xue K, Edwards TL, Groppe M, Downes SM, MacLaren RE. 2016. The spectrum of CHM gene mutations in choroideremia and their relationship to clinical phenotype. Invest Ophthalmol Vis Sci 57:6033-6039.

Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. 2014. The human gene mutation database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 133:1-9.

van Bokhoven H, van den Hurk JA, Bogerd L, Philippe C, Gilgenkrantz S, de Jong P, Ropers HH, Cremers FP. 1994. Cloning and characterization of the human choroideremia gene. Hum Mol Genet 3:1041-1046.

Van Den Hurk JAJM, Van De Pol DJR, Wissinger B, Van Driel MA, Hoefsloot LH, De Wijs IJ, Van Den Born I, Heckenlively JR, Brunner HG, Zrenner E, Ropers H-, Cremers FPM. 2003.

667 Novel types of mutation in the choroideremia (CHM) gene: A full-length L1 insertion and an
668 intronic mutation activating a cryptic exon. Hum Genet 113:268-275.

669 Vinckevicius A, Parker JB, Chakravarti D. 2015. Genomic determinants of
670 THAP11/ZNF143/HCFC1 complex recruitment to chromatin. Mol Cell Biol 35:4135-4146.

671

Single base substitutions in the *CHM* promoter as a cause of choroideremia

Alina Radziwon[1*], Gavin Arno[2, 3], Dianna Wheaton[4], Ellen M. McDonagh[5], Emma L. Baple[5,6], Kaylie Webb-Jones[4], David Birch[4], Andrew R. Webster[2, 3], Ian M. MacDonald[1]

[1]Department of Ophthalmology and Visual Sciences, University of Alberta, Edmonton, Canada

[2]UCL Institute of Ophthalmology, University College London, London EC1V 9EL, UK

[3]Moorfields Eye Hospital, London EC1V 2PD, UK

[4] Retina Foundation of the Southwest, Dallas, Texas, United States of America

[5]Genomics England, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK

[6]Medical Research (Level 4), University of Exeter Medical School, RILD Wellcome Wolfson Centre, Royal Devon and Exeter NHS Foundation Trust, Barrack Road, Exeter, EX2 5DW, UK

*Correspondence: Alina Radziwon, Department of Ophthalmology and Visual Sciences, 7-030 Katz Building, University of Alberta, Edmonton, Alberta, Canada, T6G 2E1. Phone: 1-780-492-1141; e-mail: alinar@ualberta.ca

Presented at: The Association for Research in Vision and Ophthalmology Annual Meeting, May 2016, Seattle, WA, USA.

Financial Support: IM: Canadian Institutes of Health Research, Emerging Team Grant: 119190, Foundation Fighting Blindness, Canada, Choroideremia Research Foundation Canada, Inc., and Alberta Innovates-Health Solutions 201201139

DB: National Eye Institute grant EY-09076, Foundation Fighting Blindness

GA, AW: The National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at Moorfields Eye Hospital, and the UCL Institute of Ophthalmology. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health). The 100,000 Genomes Project is funded by the NIHR and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

Abstract

Although over 150 unique mutations affecting the coding sequence of *CHM* have been identified in patients with the X-linked chorioretinal disease choroideremia (CHM), no regulatory mutations have been reported, and indeed the promoter has not been defined. Here we describe two independent families affected by CHM bearing a mutation outside the gene's coding region at position c.-98: C>A and C>T, which segregated with the disease. The male proband of family 1 was found to lack *CHM* mRNA and its gene product Rab escort protein 1 (REP-1), while whole genome sequencing of an affected male in family 2 excluded the involvement of any other known retinal genes. Both mutations abrogated luciferase activity when inserted into a reporter construct, and by further employing the luciferase reporter system to assay sequences 5′ to the gene, we identified the *CHM* promoter as the region encompassing nucleotides c.-119 to c.-76. These findings suggest that the *CHM* promoter region should be examined in patients with choroideremia who lack coding-sequence mutations, and reveals, for the first time, features of the gene's regulation.

Keywords

CHM, choroideremia, promoter, REP-1, ZNF143, THAP11

Introduction

Choroideremia (CHM; MIM# 303100) is an X-linked, recessively inherited chorioretinal dystrophy with an incidence of 1/50,000. Progressive degeneration of photoreceptors, retinal pigment epithelium (RPE), and the choroid causes affected hemizygous males to develop night blindness in the first or second decade of life, followed by a decrease in peripheral visual fields and an eventual loss of central visual acuity in advanced stages of the disease (Coussa and Traboulsi, 2012). Carrier females, while usually asymptomatic, may exhibit signs of retinal degeneration upon fundoscopic examination, and more rarely have reduced dark adaptation and peripheral vision (Karna, 1986; Roberts et al., 2002).

To date, CHM has only been linked to mutations within the *CHM* gene, coding for REP-1 (Cremers et al., 1990). The protein serves as a molecular chaperone for small GTPases from the Rab family, presenting them to Rab geranylgeranyl transferase which modifies them by the covalent attachment of a lipid moiety. The lipid modification, known as prenylation, of target Rab proteins is essential for intracellular vesicular transport (Seabra et al., 1992; Seabra et al., 1993). A homologue, REP-2, encoded by *CHML* or choroideremia-like, functions similarly and appears to compensate for the absence of REP-1 in all tissues except the eye (Cremers et al., 1994). *CHM* spans over 150 kb of Xq21.2 and contains 15 exons (van Bokhoven et al., 1994). A 30 base pair 5'untranslated region, found on exon 1, precedes an open reading frame encoding the 653 amino acid REP-1 protein. At least 163 unique pathogenic mutations have been reported in the LOVD Retinal and Hearing Impairment Genetic Mutation Database [https://grenada.lumc.nl/LOVD2/Usher_montpellier/home.php ; accessed Feb 2017] (Fokkema et al., 2011). The mutation spectrum includes transitions and transversions leading to protein truncation, splice defects, indels and large deletions ranging from a single exon to the full gene.

Missense mutations predicted to alter protein structure or impair function (Sergeev et al., 2009; Esposito et al., 2011), transposon insertions (Van Den Hurk et al., 2003), partial gene duplications (Chi et al., 2013; Simunovic et al., 2016) and other variations are infrequently found, but taken collectively, almost all known pathogenic variants in the *CHM* gene have been loss of function mutations that abolish functional REP-1 (McTaggart et al., 2002; Simunovic et al., 2016). Notably there is no apparent correlation between genotype and phenotype, with the age at onset of symptoms, visual acuity and visual fields being unrelated to mutation type (Freund et al., 2016; Simunovic et al., 2016).

The promoter driving expression of *CHM* has been heretofore unidentified. Analyses of promoter mutations causing inherited diseases can be useful in identifying transcription factors involved in the regulation and expression of a gene of interest, as the activity of RNA polymerase II is mediated by the recruitment of general and sequence specific transcription factors to *cis*-acting regulatory sequences. The core promoter directing basal level transcription is generally found between nucleotides -40 and +50 of the transcription start site (TSS) (de Vooght et al., 2009), while proximal promoter elements typically reside within 1 kilo bases (kb), with additional enhancers, repressors, insulators acting even over distances of mega bases (Maston et al., 2006). The investigation of regulatory elements in the context of human disease is complicated by the fact that in contrast to loss of function mutations, regulatory defects may produce small quantitative changes that are difficult to detect. Mutations in *cis*-acting regulatory sequences are a significant cause of human disease, and according to statistics compiled by the Human Gene Mutation Database [http://www.hgmd.cf.ac.uk ; accessed May 2016] approximately 2% of disease causing point mutations are in non-coding regions of the genome (Stenson et al., 2014). Though promoter defects can cause functionally important consequences

for gene expression, their analysis is often not a regular part of DNA diagnostics as the investigation can be complex, laborious and difficult to perform.

In this study, we report the first known regulatory mutations causing choroideremia. The absence of a genetic defect in a CHM family prompted us to explore the upstream non-coding region, revealing the novel promoter variant c.-98C>A. Subsequently, whole genome sequencing in an unrelated family with CHM identified a second variant of the same residue, c.-98C>T. We demonstrate the mutations' effects on transcription using a luciferase assay, and employ the same system to further characterize the boundary of the *CHM* promoter that spans the location. The results suggest a disruption of a transcription factor binding site and impaired transactivation of the *CHM* promoter by the factor(s). Together, these studies further our understanding of regulation and expression of *CHM* and present a possible explanation for cases of unexplained choroideremia where no causative mutation is found within the gene.

Materials and Methods

Clinical Examination and Study Subjects

This study was approved by the institutions' respective ethics review boards. All procedures adhered to the tenets of the Declaration of Helsinki. Before participation, the purpose and risks of the study were explained, and informed consent was obtained. Blood samples were drawn, and a detailed pedigree and history was recorded.

We studied 4 males affected by CHM and 1 carrier female from two unrelated families. Individuals were examined by fundoscopy and functional ophthalmologic methods including some or all of: visual acuity testing, central visual fields, full field electroretinography, optical

coherence tomography and fundus autofluorescence imaging. Choroideremia was diagnosed by senior clinicians with expertise in inherited retinal disease based on clinical findings alone. Affected males had a history of night vision loss and fundus appearance of peripheral retinopathy with broad areas of RPE and choroidal atrophy.

CHM Sanger Sequencing

DNA was extracted from lymphocytes using conventional methodologies. Proband 5116 from family 1 (C127) was initially screened for *CHM* mutations in the coding sequence and splice sites by direct Sanger sequencing from PCR amplicons (Furgoch et al., 2014). Additionally, 1kb of the 5'-flanking sequence was amplified with primers 5'-CAGGGAAGGCCCACTACTGC -3' and 5'-CTTGTGGAAATGAGATCAAGTTAGG-3' and sequenced with the same primers. With the exception of patient 111 and his parents, of family 2 who underwent whole genome sequencing, the remaining study individuals were genotyped only at position c.-98 to confirm the presence or absence of the respective variant. Annotation is in accordance with GenBank: NM_000390.3, where +1 represents the start of translation.

Cell Culture

Leukocytes from patient 5116 were separated with Ficoll-Histopaque (Sigma-Aldrich, St. Louis, MO, USA) from 10mL of whole human blood collected in acid citrate dextrose (ACD-A) tubes (BD, Franklin Lakes, NJ, USA). A lymphoblastoid cell line was established by Epstein-Barr virus transformation within 48 hours of collection (Anderson and Gusella, 1984). Cells were maintained in RPMI-1640 supplemented with 15% fetal calf serum and penicillin-streptomycin.

mRNA and Protein Analysis

Total RNA was extracted from $5x10^6$ lymphoblastoid cells from patient 5116 and a healthy control, with the NucleoSpin RNA isolation kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. cDNA was transcribed from 5µg of total RNA with the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Rockford, IL, USA) using a random hexamer primer mix. PCR amplification from cDNA with forward primer 5'-TAATAGTCACATGACACGTTTCCCG-3', paired with 5'-TGGATTCAGGCAAACCCGT-3', or 5'-TTTAAAATGAGCAAGTCAATGTGC-3' as a reverse primer was used to detect partial transcript (spanning the junction of exon 1 and 2), or the entire coding sequence of *CHM* encompassed by 15 exons, respectively. GAPDH was also amplified as a positive control with intron-spanning primers.

Protein was extracted from $2x10^6$ lymphoblastoid cells and immunoblot analysis of REP-1 was performed as previously described (Furgoch et al., 2014).

Whole Genome Sequencing

UK proband 111 (GC406) and his unaffected parents 190 and 192 underwent whole genome sequencing (WGS) as part of the 100,000 Genomes Project. Briefly, genomic DNA was processed using the Illumina TruSeq DNA PCR-Free Sample Preparation kit (Illumina Inc) and sequenced using an Illumina HiSeq X Ten, generating minimum coverage of 15X for >97% of the callable autosomal genome. Reads were aligned to build GRCh37 of the human genome using the Isaac aligner (Illumina Inc). SNVs and indels were identified using Platypus v0.8.1 and annotated using Cellbase (https://github.com/opencb/cellbase). Variant filtering was performed using minor allele frequency (MAF) in publicly available and in-house datasets, predicted

protein impact and familial segregation. Surviving variants were prioritized using a pre-specified virtual gene panel from PanelApp (https://panelapp.extge.co.uk/crowdsourcing/PanelApp/ Posterior segment abnormalities v1.7. Allelic state was required to match the curated mode of inheritance for variants in panel genes.

Promoter Analysis

Construction of luciferase reporter plasmids

Primers were initially designed to amplify a 1063 bp fragment from upstream of the *CHM* transcription start site (c.-1093_c.-31) from control and patient 5116's DNA. The c.-98C>T mutation present in family 2 was generated by site directed mutagenesis from the control plasmid. All sequences cloned contained only untranscribed DNA and did not include the 30 base pair 5'UTR directly upstream of the start codon. The fragment was cloned according to standard protocols, with SacI and HindIII restriction sites added to primers allowing digest by the enzymes (Thermo Scientific, Rockford, IL, USA). Ligation into the promoterless firefly luciferase reporter vector, pGL3-Basic (Promega, Madison, WI, USA) generated constructs pGL3-1063_-31 and pGL3-1063_-31 c.-98 C>A. Subsequently, a 4kb fragment, and a series of nested deletions ranging from 2kb to 44bp were cloned into pGL3-Basic with the same restriction sites added and used for subsequent digest, to identify the minimal promoter necessary and sufficient for transcription. pGL3-Basic lacks a eukaryotic promoter sequence upstream of the reporter luciferase gene, and expression of luciferase in transfected cells depends on a functional promoter sequence to be inserted upstream of the *luc+* gene. pGL3-Basic itself served as a negative control. A schematic of all inserts assayed is provided in Figure 1A. All constructs

generated were Sanger sequenced to ensure fidelity. Primer sequences for the generation of plasmid inserts are available upon request.

Luciferase Assay

HEK293T cells cultured with standard reagents and conditions were seeded in 24 well plates and transfected at 70% confluency with polyethylenimine (PEI). Each well was transfected with a total of 650ng of plasmid DNA, including 600ng of pGL3 construct and 50ng of the internal control pRL-CMV, a cytomegalovirus promoter driven *Renilla* luciferase reporter vector. Briefly, plasmid DNA was diluted into 25µL of PBS, and mixed by vortexing with 1µL of 1mg/mL branched PEI (Sigma-Aldrich, St. Louis, MO, USA) in 25µL PBS. After a 15 minute incubation, 50µL of the reaction was added dropwise to each well. Forty-eight hours after transfection, luciferase activity was assayed using the Dual-Luciferase Reporter Assay kit and measured with the Glomax Explorer (both Promega, Madison, WI, USA). Relative luciferase activity was obtained by dividing the relative light units (RLU) produced by the firefly luciferase pGL3 construct by the RLUs produced by *Renilla* luciferase control reaction. Light generated by the reaction can be correlated with the amount of luciferase protein produced which in turn is proportional to promoter activity driving the gene's expression. Values from an n=6 were averaged and normalized to that of the reference construct, pGL3-1093_-31 to obtain a relative measure of activity.

Statistical analysis

All data were expressed as mean ± SD. Data from the luciferase assay represents two independent experiments with triplicate measurements. Differences between groups were examined for statistical significance using Student's *t*-test. A *P*-value <0.01 denoted the presence of a statistically significant difference.

Results

Clinical Characterization of CHM

Family 1

We investigated a progressive retinal degeneration in a Caucasian family of American origin (C127). Remarkably, the proband 5116 was the offspring of a consanguineous union between second cousins, with an affected father and carrier mother. To our knowledge, this has never been reported with choroideremia, and it necessitated a thorough and accurate diagnosis of X-linked CHM and exclusion of an autosomal recessive, or even an unusual male-male transmission as cause for the disease. An investigation of the extended family for which the pedigree is reported in Figure 2A however, clearly demonstrated the X-linked inheritance of the disease.

Patient 5116 was initially seen at age 56 years, when he was referred by a retinal specialist with a diagnosis of choroideremia. At his most recent examination, age 76, best corrected visual acuity (BCVA) was measured as 6/19 OD and 6/15 OS. Visual fields were reduced to less than 5 degrees. The full-field ERG showed non-detectable dark adapted rod-driven responses as well as non-detectable cone responses to a 30 Hz flicker stimulus. OCT imaging indicated loss of the photoreceptor layer across the periphery of the fundus with only a small island of RPE remaining in the macula. Posterior segment examination showed a hypo-pigmented fundus with significant atrophy of the RPE and choroid with areas of bare sclera (Figure 3). These findings are consistent with an advanced state of choroideremia.

The proband's daughter 5113 was examined at 33 years of age and did not report any vision difficulties. BCVA when examined was 6/6, both eyes. Full-field ERG testing of the left eye showed dark adapted rod-driven responses were reduced by 50% in amplitude (42.5 μV) and a normal b-wave implicit time (81.6 msec). Light-adapted cone-driven responses to a 30 Hz flicker stimulus were reduced by 50% in amplitude (27.6 μV) and borderline reduced in implicit time (25.6 msec). These findings were consistent with a classic carrier state of choroideremia.

The proband's father and one brother 5149 were reported to have been previously diagnosed with choroideremia, while a second brother 5147 was reported to be unaffected. Findings are summarized in Table 1.

Family 2:

Family 2 (GC406) was a Caucasian family of British origin with a history of choroideremia. Proband 111 was first examined at age 13, displaying a reduced but not delayed full field ERG with flicker and bright flash responses both within normal limits. He remained asymptomatic until age 32. Confrontational visual field testing at age 35 showed bilateral infratemporal scotoma, and BCVA was 6/5 in both eyes. He is considered symptomatically mild with retained central macular structure.

Of his two maternal uncles, 151 was diagnosed at age 12. Fundus abnormalities typical of choroideremia were noted. 161 was diagnosed at age 8, and upon examination at age 37 exhibited moderately constricted visual fields particularly in the superior field bilaterally. BCVA was reduced to count fingers vision OD and 6/36 OS. Fundus examination was consistent with a clinical diagnosis of choroideremia. The pedigree is reported in Figure 2B, and clinical findings summarized in Table 1.

Genetic Analysis

Family 1 (C127)

Genetic analysis of proband 5116 did not reveal any pathogenic mutation in the coding sequence of the *CHM* gene or splice site boundaries. Yet at position c.-98 relative to the translation start site, a hemizygous C>A transversion was detected (hg38, chrX:g.86047629G>T NM_000390.3, c.-98C>A). The variant was not listed in the latest release of dbSNP [http://www.ncbi.nlm.nih.gov/SNP/; accessed Feb 2017] (Sherry et al., 2001). The proband's affected brother (5149) was also found to harbour the variant while it was absent in the unaffected brother (5147) available for testing. The obligate carrier status was confirmed in the proband's daughter (5113). The location of the variant strongly suggested a regulatory mutation, as evaluation of entries in the HGMD reveals that most promoter mutations are located between +50 and -500 from the TSS of a gene (Stenson et al., 2014).

Family 2 (GC406):

Prior genetic analysis of the coding exons of the *CHM* gene did not reveal a pathogenic mutation. Whole genome sequencing was performed on proband 111 and his parents 190 and 192, as part of the 100,000 Genomes Project. After variant filtering, no causative rare coding variants were identified in any retinal disease gene.

In light of the clinical diagnosis of choroideremia in the family, the complete *CHM* gene was interrogated for rare variants (≤0.001 MAF in 1Kgenome project and internal cohort of over 2000 whole genome sequencing samples) hemizygous in the proband and carried by his mother. One such variant was identified, the transition, c.-98C>T (hg38, chrX:g.86047629G>A NM_000390.3, c.-98C>T). The variant was confirmed in the two affected maternal uncles 151 and 161 by direct Sanger sequencing.

Molecular diagnosis of choroideremia in 5116

Immunoblot analysis of protein harvested from a cell line derived from patient 5116 failed to detect the *CHM* gene product REP-1, providing conclusive confirmation of the clinical diagnosis of choroideremia (Figure 4A). To provide evidence for a regulatory mutation and subsequently reduced transcription, we intended to compare the level of expression between normal and patient samples through qPCR. Endpoint PCR from a cDNA template; however, failed to amplify the 15 exon, 2200 base pair transcript, indicating it was absent or present at a level below the detection threshold, and thus the quantitative assay was not performed. To largely rule out the possibility of a splice defect, we also attempted to amplify a minimal portion of the transcript, a 93 base pair fragment from the 5'-UTR to a region spanning the boundary of exon 1 and 2, and found the patient's cells lacked even this short fragment (Figure 4B). A control housekeeping gene was nevertheless readily detected and both partial and full length transcripts were amplified from normal cDNA.

Effect of the c.-98C>A and c.-98C>T mutations on transcription of CHM

As a starting point, an approximately 1kb fragment upstream of the TSS was assayed for ability to drive gene expression, since the majority of elements necessary for transcription are expected to be found within this region (Rockman and Wray, 2002). Comparing the robust luciferase activity produced by cells transfected with this wild type construct pGL3-1093_-31, to that of pGL3-1093_-31 c.-98C>A and c.-98C>T, we observed complete abrogation of promoter activity (Figure 1B) in the mutants. The drop from 100 ± 9.5 to 2.0 ± 0.3 and 1.3 ± 0.4 respectively, as measured in normalized relative light units (RLU) ± 1 standard deviation is even significantly

lower than that of the negative control pGL3-basic which does not contain a promoter sequence, reading at 5.8 ± 0.9. This startling observation strongly suggested that the mutation spans an element essential to promoter activity.

Based on unpublished reports suggesting other activating elements may exist up to 2.8kb upstream of the TSS (Kaiser NW, *et al.* IOVS 2004;45:ARVO E-Abstract 2451), we additionally compared the activity of two larger constructs approximately 4kb and 2kb long. Typically, regulatory regions upstream of core promoter sequences contain multiple TF specific binding motifs, where several copies of the same factor or cooperation between different factors serve to synergistically stimulate transcriptional activity of a given gene (Maston et al., 2006). Yet in the case of *CHM*, the longer sequences did not stimulate additional luciferase expression. pGL3-3983_-31 and pGL3-2027_-31 did not significantly differ from pGL3-1093_-31, producing 83.7 ± 4.9 and 96.2 ± 7.6 RLUs respectively. The somewhat diminished expression as compared to the reference 1kb construct may be attributed to a lower molar amount of the larger plasmids being delivered during transfection.

Characterization of CHM promoter

To further characterize the promoter, we proceeded to assay progressively shorter constructs, deleting sequences from the -5' and -3' of the 1kb construct to define its boundaries (Figure 1A). Working from the 5' end, the approximately 400, 300, and 100 base pair constructs pGL3-437_-31, pGL3-346_-31, pGL3-119_-31 did not significantly differ from the 1 kb containing reference plasmid pGL3-1093_-31, reading at 100.2 ± 11.4, 96.1 ± 10.6, and 97.7 ± 8.1 RLU respectively. Yet upon deleting a further 11 base pairs, we observed a significant drop to 39.8 ± 7.6 RLU with pGL3-108_-31. Having delineated the -5' boundary of the promoter as extending to no further

than nucleotide c.-119, we subsequently tested pGL3-119_-76 bearing a deletion from the -3' end, and found no significant difference at 93.4 ± 14 RLU. It was the final construct, pGL3-119_-82, which yielded significantly lower activity producing only 27.4 ± 4 RLU. Though further characterization of the promoter region down to the single nucleotide level could be performed, we have defined the borders of the promoter to within the region of c.-119 to c.-76. This short 44 base pair DNA region upstream of the *CHM* coding sequence is able to stimulate transcription in an *in-vitro* assay at a level that is not significantly different from that of a 1kb, or even a 4kb fragment and can be understood to contain essential *cis*-acting elements that positively regulate *CHM* expression.

Discussion

To date, all genetic defects causing choroideremia, which include all major mutation classes, have been observed in exons, introns or intron/exon boundaries of the *CHM* gene (Fokkema et al., 2011). No mutations have been reported in the promoter region of *CHM*, and the regulation of the gene has remained essentially unexplored. In this study, we present the first report of promoter mutations, c.-98C>A and c.-98C>T, causing choroideremia. The critical role of this residue for gene expression is highlighted by the complete abrogation of reporter activity in a promoter assay when the nucleotide is mutated. Though we demonstrate the first examples in a novel mutation class, it is worthwhile noting that the phenotypes observed in affected individuals were typical of choroideremia. In family 1, the genotype c.-98C>A in proband 5116 resulted in severe retinopathy, while characteristic milder signs were found in the carrier female 5113. Males from family 2 bearing the c.-98C>T mutation manifested with chorioretinal disease, but with varying degrees of severity; proband 111 remained symptomatically mild while in his

maternal uncles 151 and 161, disease progression was more typical. CHM is known to exhibit a wide interfamilial, and also intrafamilial variability (Moosajee et al., 2014). As the mutations c.-98C>A and c.-98C>T appear to completely abolish transcription, producing no detectable levels of *CHM* mRNA and REP-1 protein in patient cells, it is not surprising for the disease to present in a classical manner consistent with phenotypes observed for the loss of function mutations. Whole genome sequencing in the parent-offspring trio of family 2 allowed the exclusion of a deep intronic *CHM* mutation, as well as mono or biallelic mutations in any other retinal dystrophy gene as a cause of the observed symptoms. We therefore conclude that the two mutations at residue c.-98 impair *CHM* transcription enough to result in choroideremia. For the several gene therapy trials currently underway employing a "gene replacement" strategy (Dimopoulos et al., 2015), individuals bearing these genotypes would be suitable candidates. The two variants have been submitted and published in the LOVD Retinal and Hearing Impairment Genetic Mutation Database (Fokkema et al., 2011).

Having identified mutations that evidently abolished transcription, we interpreted their location to span a region crucial for gene expression, and set out to define the boundaries of the gene's promoter. Surprisingly, after analyzing fragments as long as 4kb, we found a short 44 base pair DNA fragment to be wholly responsible for driving expression. The region c.-119 to c.-76 comprises the entirety of the *CHM* proximal promoter, and is able to drive robust transcription in an *in-vitro* luciferase assay. The region implicated amounts to less than 3% of the length of the *CHM* coding sequence where the large majority of reported causative mutations have been found (Fokkema et al., 2011). A recent investigation of a large disease cohort (74), found causative mutations in the gene in 94% of cases previously diagnosed with CHM (Simunovic et al., 2016).

The remaining 6% can be understood to be comprised of regulatory mutations, incorrect diagnoses or deep intronic variant causing cryptic splicing (Carss et al., 2017). Promoter defects are therefore expected to be responsible for a small minority of choroideremia cases. The identified region; however, becomes an obvious area for examination in patients in whom no coding sequence mutation is found.

A promoter or regulatory mutation can be expected to either increase or decrease transcriptional activity mediated by the altered binding capacity of trans-acting protein factors specific to a DNA sequence in the promoter region. In this case, the interaction appears to be entirely disrupted based on the null expression of a reporter driven by mutated sequence in our luciferase assay. Impaired transcription due to a 1-bp mutation in a promoter region is unusual, but has been reported previously as in the case of single base mutations stimulating additional transcriptional activity at the *OVOL2* promoter causing autosomal-dominant corneal endothelial dystrophies (Davidson et al., 2016). On the other hand, decreased transactivation of *NMNAT1* due to a single nucleotide change in the promoter was found to cause Leber congenital amaurosis (Coppieters et al., 2015).

While we delineate the *CHM* promoter boundary to this small area upstream of the gene, we cannot exclude the possibility of distant enhancer, repressor, or intronic elements also contributing to regulation; our findings suggest the sequence between c.-119 and c.-76 is essential, but not necessarily sufficient for transcription. The wider upstream sequence of *CHM* lacks the consensus sequences often found in RNA polymerase II promoters, such as CAAT and TATA boxes; as well it is neither GC rich nor associated with any CpG islands. Bioinformatic

analysis with MotifMap, a dataset of computationally predicted transcription factor binding sites based on binding motifs [http://motifmap.ics.uci.edu/] (Daily et al., 2011) identifies a putative binding motif for the transcription factor zinc finger protein 143 (ZNF143) contained within the region of the now revealed *CHM* promoter. ZNF143 participates in the regulation of RNA pol II and III mediated transcription of protein coding, non-coding, and small nuclear RNA (Schaub et al., 1997; Myslinski et al., 1998) and was initially connected with the binding motif TTCCCATTATGCACCGCG (SBS1) (Myslinski et al., 2006). Genome-wide studies revealed the binding site for the factor to be frequently found with an adjacent 5' accessory sequence, forming the ACTACAATTCCCATTATGCACCGCG (SBS2) motif. SBS2 is comprised of both a THAP domain-containing protein 11 (THAP11) and ZNF143 binding site, with the factors believed to act in a competitive manner (Ngondo-Mbongo et al., 2013). The recruitment of THAP11 to its canonical binding site ACTAYRNNNCCCR is most frequently associated with up-regulation genes essential to protein biosynthesis and energy production (Dejosez et al., 2010). More recently, ZNF143 was suggested to cooperatively occupy SBS2 sites with THAP11 and a third factor, the scaffold protein host cell factor 1 (HCFC1) *in-vivo* (Vinckevicius et al., 2015). SBS2 is closely matched by the sequence found at position c.-108 to c.-84 upstream of *CHM*, ACTACAACACCCAGAATGCACTGTT. Notably, ZNF143's binding at promoters was recently implicated in chromatin looping with distal regulatory elements (Bailey et al., 2015), suggesting the involvement of yet other factors in the total regulation of expression of *CHM.*

Of the three transcription factors implicated above, the binding of at least ZNF143 is supported by publicly available chromatin immunoprecipitation sequencing data released as part of the ENCODE project [https://genome.ucsc.edu/ENCODE/] (ENCODE Project Consortium, 2012).

In all cell types assayed: lymphoblastoid, HeLa and K562 cell lines, as well as embryonic stem cells, the promoter region of *CHM* interacted with ZNF143, for which a consensus binding sequence GAACTACAATTCCCAGAAGGC, again is closely matched by GAACTACAACACCCAGAATGC found between position c.-110 toc.-91 relative to the A of the *CHM* start codon. The position-weight matrix for ZNF143 (Figure 5) establishes the relative frequency of the base C at position c.-98 to be 100%, supporting the importance of the residue and the resulting pathogenesis when mutated. Furthermore, surveying the UCSC genome browser (http://genome.ucsc.edu/) multiz alignment of 100 vertabrate genomes tract, the region c.-119 to c.-76 shows a high degree of conservation among mammals, and absolute conservation of residue C at positions corresponding to c.-98 pointing to an important biological role for the sequence (Kuhn et al., 2007; Blanchette et al., 2004). Representatives of birds and amphibian classes, however, lack homology in the region, in while in fish a corresponding region is absent altogether. The promoter, therefore, cannot be considered an ultra-conserved non-coding element (Dimitrieva and Bucher, 2013). Several alignments are listed in Table 2.

Studies of *CHM* mRNA and protein localization have found a broad expression profile for both. In mice, evidence of transcription was found in multiple cell types and in every major layer of retina (Keiser et al., 2005), while immunolabeling of primate retina showed REP-1 localized to both rod and cones (Dimopoulos et al., 2015). Studies in human and primate retina found that mRNA levels did not correspond to the pattern of disease expression; little *CHM* was detected in the RPE and choroid, and there were no marked regional differences in the concentration of *CHM* mRNA apparent with foveal versus mid-peripheral total RNA despite affected males typically exhibiting a preservation of central vision until late in the disease (Bernstein and Wong,

1998). Additionally, REP-1 can be readily detected in human fibroblasts or peripheral blood mononuclear cells (Furgoch et al., 2014; MacDonald et al., 1998). Taken together with ZNF143's characterization as one of the most common and ubiquitously expressed TFs (Myslinski et al., 1998) a picture emerges of widespread and non-specific transcription of *CHM,* despite choroideremia's manifestation as an ocular disease. Indeed, patients' apparent lack of systemic symptoms can be understood to result not from tissue specific expression of REP-1, but from the differing affinities of REP-1 and REP-2 for target Rabs, which may themselves be differentially expressed or possess tissue or cell specific activity. Investigators have implicated Rab27 (Seabra et al., 1995) and Rab38 (Kohnke et al., 2013) as possible contributors.

The study presented here also poses interesting questions, such as whether mutations of other residues less critical to transactivator binding in the *CHM* promoter, that diminish, but not completely abolish mRNA expression, can result in a milder phenotype, or a different rate of progression of choroideremia. Currently, the dbSNP database lists no known SNPs between c.-119 to c.-76 (Sherry et al., 2001). Having shown the *CHM* region responsible for regulation of its expression, described for the first time the features of its promoter, and extended the inventory of molecular changes causing choroideremia, the findings are of clinical and diagnostic interest and present an obvious area of examination for patients with CHM in whom no coding sequence mutation has been found. Further elucidating the roles of ZNF143, THAP11, HCFC1 or other distal factors will prove an important step toward understanding the complete picture of *CHM*'s regulation.

Acknowledgements

The authors have no proprietary or commercial interest in any materials discussed in this article.

Support by Canadian Institutes of Health Research, Emerging Team Grant: 119190, Foundation Fighting Blindness, Canada, Choroideremia Research Foundation Canada, Inc., and Alberta Innovates-Health Solutions 201201139 is acknowledged (IM).

This project was also supported by The National Institute for Health Research (NIHR) and Biomedical Research Centre (BRC) at Moorfields Eye Hospital and the UCL Institute of Ophthalmology (GA, AW).

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health). The 100,000 Genomes Project is funded by the NIHR and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

Support by the National Eye Institute grant EY-09076, and Foundation Fighting Blindness is also acknowledged (DB)

Figure Legends

Figure 1. Functional Analysis of the *CHM* Promoter.

(A) The constructs portrayed on the left were inserted upstream of the luciferase gene in pGL3-basic. Nucleotide +1 represents the translation start site. The promoterless pGL3-basic served as a negative control. **(B)** The mutations c.-98C>A, and c.-98C>T abolish transcription, while the minimal construct c.-119_-76 is sufficient for robust expression of the reporter gene not significantly different from even that of the nearly 4kb construct. All constructs were transiently transfected into HEK293T cells. A dual-luciferase reporter assay was used to assess the potential promoter activity of various sized inserts and the c.-98 mutants. Promoter activity is shown as a ratio of firefly luciferase over *Renilla* luciferase present on the transfection control plasmid pRL-CMV to account for inter-well variation. Activity is normalized to that of the reference construct pGL3-1093_-31 which is artificially set to equal 100. Activity significantly different ($P<0.01$) from the reference construct is denoted by an asterisk. Error bars represent ± 1 SD.

Figure 2. Pedigree Structure of Affected Families with *CHM* Promoter Mutations

(A) Family1. The parents of proband 5116 were second cousins, sharing a set of great grand-parents. The inheritance pattern mimics male-male transmission, but is nevertheless consistent with X-linked inheritance upon examination of the wider family pedigree. **(B)** Family 2. The pedigree showing three generations affected by choroideremia examined in this study. Inheritance follows an X-linked pattern.

Figure 3. Retinal features in patient 5116.

Legend: Left column (OD), right column (OS). **(A)** Fundus photographs of the proband taken at age 76 showing typical choroideremia changes, with atrophy of the choroid and RPE. A small island of preserved RPE remains in the central macula, surrounded by atrophic peripheral areas of apparent bare sclera. **(B)(C)** Fundus autofluorescence image demonstrating areas of residual RPE (L) and the corresponding OCT image (R). Preserved retinal areas with normal

autofluorescence exhibit thicker choroid and preserved retinal lamination. An outer retinal tubulation is seen in the right fundus.

Figure 4. Molecular confirmation of choroideremia in patient 5116.

(A) Patient 5116 lacks REP-1. Western blot results show the absence of a ~100 kDa band corresponding to REP-1 in a lymphoblastoid cell line derived from the patient (lane 2), which is present in a normal control (lane 1). A β-actin antibody was used as a loading control to ensure an adequate protein sample in each lane, with the 42 kDa band present in both samples. **(B)** Patient 5116 lacks *CHM* mRNA. cDNA synthesized from the mRNA harvested from a patient generated lymphoblastoid cell line was used as template for PCR. Lanes 2 and 3 show a 475 bp band resulting from the amplification of the *GAPDH* control housekeeping gene from 5116 and a normal control, indicating cDNA of adequate quality. Lanes 6 and 9 demonstrate an absence of amplification from the patient's cDNA of both partial and full length coding sequence, respectively, as compared to PCR products sized 93 and 2200 base pairs amplified from normal cDNA observed in lanes 5 and 10.

Figure 5. Consensus binding sequences for transcription factor ZNF143.

(A) Partial map of the human *CHM* gene; arrow indicates transcription start site. **(B)** Expanded sequence of the minimal *CHM* promoter from c.-119 to c.-76, as identified through the analysis of progressive deletion constructs. Position c.-98 is marked with an asterisk. **(C)** Sequence logo derived from publically available ChIP-seq data released as part of the ENCODE project, with the position weighted matrix below (ENCODE Project Consortium, 2012). An invariant C is found at position corresponding to c.-98 of *CHM*.

Table 1. Clinical data of genotyped individuals

Table 2. Multiz alignment of the promoter region of 15 *CHM* orthologs

References

Anderson MA, Gusella JF. 1984. Use of cyclosporin A in establishing epstein-barr virus-transformed human lymphoblastoid cell lines. In Vitro 20:856-858.

Bailey SD, Zhang X, Desai K, Aid M, Corradin O, Cowper-Sal Lari R, Akhtar-Zaidi B, Scacheri PC, Haibe-Kains B, Lupien M. 2015. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nat Commun 2:6186.

Bernstein SL, Wong P. 1998. Regional expression of disease-related genes in human and monkey retina. Mol Vis 4:24.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14:708-715.

Carss KJ, Arno G, Erwood M, Stephens J, Sanchis-Juan A, Hull S, Megy K, Grozeva D, Dewhurst E, Malka S, Plagnol V, Penkett C, Stirrups K, Rizzo R, Wright G, Josifova D, Bitner-Glindzicz M, Scott RH, Clement E, Allen L, Armstrong R, Brady AF, Carmichael J, Chitre M, Henderson RH, Hurst J, MacLaren RE, Murphy E, Paterson J, Rosser E, Thompson DA, Wakeling E, Ouwehand WH, Michaelides M, Moore AT, NIHR-BioResource Rare Diseases Consortium, Webster AR, Raymond FL. 2017. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am J Hum Genet 100:75-90.

Chi JY, MacDonald IM, Hume S. 2013. Copy number variant analysis in CHM to detect duplications underlying choroideremia. Ophthalmic Genet 34:229-233.

Coppieters F, Todeschini AL, Fujimaki T, Baert A, De Bruyne M, Van Cauwenbergh C, Verdin H, Bauwens M, Ongenaert M, Kondo M, Meire F, Murakami A, Veitia RA, Leroy BP, De Baere E. 2015. Hidden genetic variation in LCA9-associated congenital blindness explained by 5'UTR mutations and copy-number variations of NMNAT1. Hum Mutat 36:1188-1196.

Coussa RG, Traboulsi EI. 2012. Choroideremia: A review of general findings and pathogenesis. Ophthalmic Genet 33:57-65.

Cremers FP, Armstrong SA, Seabra MC, Brown MS, Goldstein JL. 1994. REP-2, a rab escort protein encoded by the choroideremia-like gene. J Biol Chem 269:2111-2117.

Cremers FP, van de Pol DJ, van Kerkhoff LP, Wieringa B, Ropers HH. 1990. Cloning of a gene that is rearranged in patients with choroideraemia. Nature 347:674-677.

Daily K, Patel VR, Rigor P, Xie X, Baldi P. 2011. MotifMap: Integrative genome-wide maps of regulatory motif sites for model species. BMC Bioinformatics 12:495-2105-12-495.

Davidson AE, Liskova P, Evans CJ, Dudakova L, Noskova L, Pontikos N, Hartmannova H, Hodanova K, Stranecky V, Kozmik Z, Levis HJ, Idigo N, Sasai N, Maher GJ, Bellingham J, Veli N, Ebenezer ND, Cheetham ME, Daniels JT, Thaung CM, Jirsova K, Plagnol V, Filipec M, Kmoch S, Tuft SJ, Hardcastle AJ. 2016. Autosomal-dominant corneal endothelial dystrophies CHED1 and PPCD1 are allelic disorders caused by non-coding mutations in the promoter of OVOL2. Am J Hum Genet 98:75-89.

de Vooght KM, van Wijk R, van Solinge WW. 2009. Management of gene promoter mutations in molecular diagnostics. Clin Chem 55:698-708.

Dejosez M, Levine SS, Frampton GM, Whyte WA, Stratton SA, Barton MC, Gunaratne PH, Young RA, Zwaka TP. 2010. Ronin/hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. Genes Dev 24:1479-1484.

Dimitrieva S, Bucher P. 2013. UCNEbase--a database of ultraconserved non-coding elements and genomic regulatory blocks. Nucleic Acids Res 41:D101-9.

Dimopoulos IS, Chan S, MacLaren RE, MacDonald IM. 2015. Pathogenic mechanisms and the prospect of gene therapy for choroideremia. Expert Opin Orphan Drugs 3:787-798.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489:57-74.

Esposito G, De Falco F, Tinto N, Testa F, Vitagliano L, Tandurella IC, Iannone L, Rossi S, Rinaldi E, Simonelli F, Zagari A, Salvatore F. 2011. Comprehensive mutation analysis (20 families) of the choroideremia gene reveals a missense variant that prevents the binding of REP1 with rab geranylgeranyl transferase. Hum Mutat 32:1460-1469.

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. 2011. LOVD v.2.0: The next generation in gene variant databases. Hum Mutat 32:557-563.

Freund PR, Sergeev YV, MacDonald IM. 2016. Analysis of a large choroideremia dataset does not suggest a preference for inclusion of certain genotypes in future trials of gene therapy. Mol Genet Genomic Med 4:344-358.

608 Furgoch MJB, Mewes-Arès J, Radziwon A, MacDonald IM. 2014. Molecular genetic diagnostic
609 techniques in choroideremia. Mol Vision 20:535-544.

610 Karna J. 1986. Choroideremia. A clinical and genetic study of 84 finnish patients and 126 female
611 carriers. Acta Ophthalmol Suppl 176:1-68.

612 Keiser NW, Tang W, Wei Z, Bennett J. 2005. Spatial and temporal expression patterns of the
613 choroideremia gene in the mouse retina. Mol Vision 11:1052-1060.

614 Kohnke M, Delon C, Hastie ML, Nguyen UT, Wu YW, Waldmann H, Goody RS, Gorman JJ,
615 Alexandrov K. 2013. Rab GTPase prenylation hierarchy and its potential role in choroideremia
616 disease. PLoS One 8:e81758.

617 Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW,
618 Stanke M, Smith KE, Siepel A, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pedersen JS, Hsu
619 F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baertsch R,
620 Haussler D, Kent WJ. 2007. The UCSC genome browser database: Update 2007. Nucleic Acids
621 Res 35:D668-73.

622 MacDonald IM, Mah DY, Ho YK, Lewis RA, Seabra MC. 1998. A practical diagnostic test for
623 choroideremia. Ophthalmology 105:1637-1640.

624 Maston GA, Evans SK, Green MR. 2006. Transcriptional regulatory elements in the human
625 genome. Annu Rev Genomics Hum Genet 7:29-59.

626 McTaggart KE, Tran M, Mah DY, Lai SW, Nesslinger NJ, MacDonald IM. 2002. Mutational
627 analysis of patients with the diagnosis of choroideremia. Hum Mutat 20:189-196.

Moosajee M, Ramsden SC, Black GC, Seabra MC, Webster AR. 2014. Clinical utility gene card for: Choroideremia. Eur J Hum Genet 22:10.1038/ejhg.2013.183. Epub 2013 Aug 21.

Myslinski E, Gerard MA, Krol A, Carbon P. 2006. A genome scale location analysis of human staf/ZNF143-binding sites suggests a widespread role for human staf/ZNF143 in mammalian promoters. J Biol Chem 281:39953-39962.

Myslinski E, Krol A, Carbon P. 1998. ZNF76 and ZNF143 are two human homologs of the transcriptional activator staf. J Biol Chem 273:21998-22006.

Ngondo-Mbongo RP, Myslinski E, Aster JC, Carbon P. 2013. Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11. Nucleic Acids Res 41:4000-4014.

Roberts MF, Fishman GA, Roberts DK, Heckenlively JR, Weleber RG, Anderson RJ, Grover S. 2002. Retrospective, longitudinal, and cross sectional study of visual acuity impairment in choroideraemia. Br J Ophthalmol 86:658-662.

Rockman MV, Wray GA. 2002. Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol 19:1991-2004.

Schaub M, Myslinski E, Schuster C, Krol A, Carbon P. 1997. Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III. Embo J 16:173-181.

Seabra MC, Brown MS, Goldstein JL. 1993. Retinal degeneration in choroideremia: Deficiency of rab geranylgeranyl transferase. Science 259:377-381.

Seabra MC, Brown MS, Slaughter CA, Sudhof TC, Goldstein JL. 1992. Purification of component A of rab geranylgeranyl transferase: Possible identity with the choroideremia gene product. Cell 70:1049-1057.

Seabra MC, Ho YK, Anant JS. 1995. Deficient geranylgeranylation of ram/Rab27 in choroideremia. J Biol Chem 270:24420-24427.

Sergeev YV, Smaoui N, Sui R, Stiles D, Gordiyenko N, Strunnikova N, MacDonald IM. 2009. The functional effect of pathogenic mutations in rab escort protein 1. Mutat Res 665:44-50.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: The NCBI database of genetic variation. Nucleic Acids Res 29:308-311.

Simunovic MP, Jolly JK, Xue K, Edwards TL, Groppe M, Downes SM, MacLaren RE. 2016. The spectrum of CHM gene mutations in choroideremia and their relationship to clinical phenotype. Invest Ophthalmol Vis Sci 57:6033-6039.

Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. 2014. The human gene mutation database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 133:1-9.

van Bokhoven H, van den Hurk JA, Bogerd L, Philippe C, Gilgenkrantz S, de Jong P, Ropers HH, Cremers FP. 1994. Cloning and characterization of the human choroideremia gene. Hum Mol Genet 3:1041-1046.

Van Den Hurk JAJM, Van De Pol DJR, Wissinger B, Van Driel MA, Hoefsloot LH, De Wijs IJ, Van Den Born I, Heckenlively JR, Brunner HG, Zrenner E, Ropers H-, Cremers FPM. 2003.

Novel types of mutation in the choroideremia (CHM) gene: A full-length L1 insertion and an intronic mutation activating a cryptic exon. Hum Genet 113:268-275.

Vinckevicius A, Parker JB, Chakravarti D. 2015. Genomic determinants of THAP11/ZNF143/HCFC1 complex recruitment to chromatin. Mol Cell Biol 35:4135-4146.

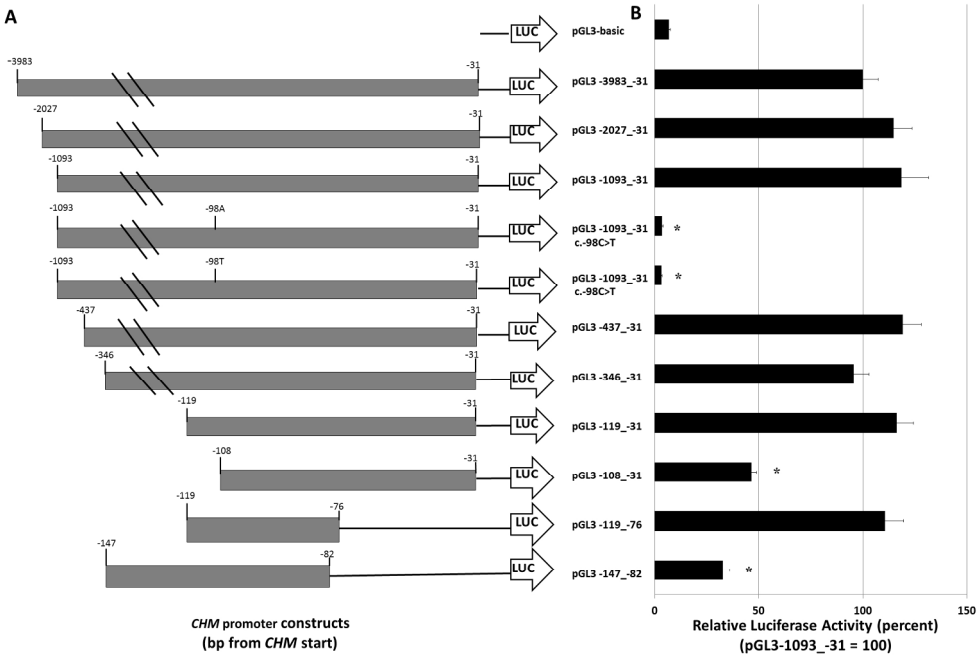


Figure 1. Functional Analysis of the CHM Promoter.
(A) The constructs portrayed on the left were inserted upstream of the luciferase gene in pGL3-basic. Nucleotide +1 represents the translation start site. The promoterless pGL3-basic served as a negative control. (B) The mutations c.-98C>A, and c.-98C>T abolish transcription, while the minimal construct c.-119_-76 is sufficient for robust expression of the reporter gene not significantly different from even that of the nearly 4kb construct. All constructs were transiently transfected into HEK293T cells. A dual-luciferase reporter assay was used to assess the potential promoter activity of various sized inserts and the c.-98 mutants. Promoter activity is shown as a ratio of firefly luciferase over Renilla luciferase present on the transfection control plasmid pRL-CMV to account for inter-well variation. Activity is normalized to that of the reference construct pGL3-1093_-31 which is artificially set to equal 100. Activity significantly different ($P<0.01$) from the reference construct is denoted by an asterisk. Error bars represent ± 1 SD.
Figure 1
609x406mm (96 x 96 DPI)

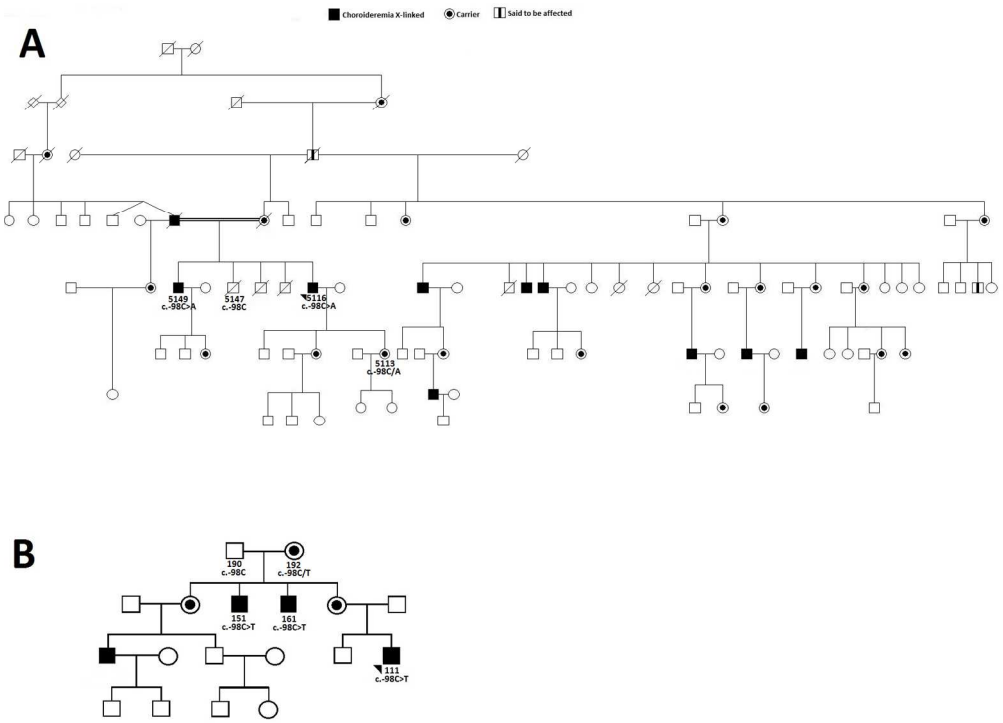


Figure 2. Pedigree Structure of Affected Families with CHM Promoter Mutations
(A) Family1. The parents of proband 5116 were second cousins, sharing a set of great grand-parents. The inheritance pattern mimics male-male transmission, but is nevertheless consistent with X-linked inheritance upon examination of the wider family pedigree. (B) Family 2. The pedigree showing three generations affected by choroideremia examined in this study. Inheritance follows an X-linked pattern.
Figure 2
478x349mm (96 x 96 DPI)

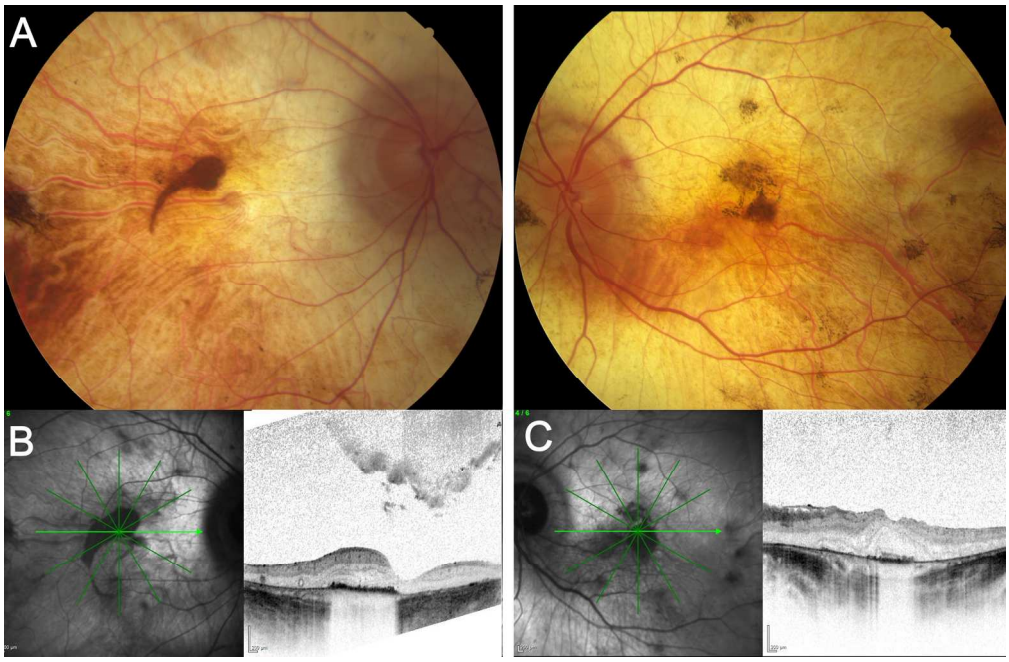


Figure 3. Retinal features in patient 5116.
Legend: Left column (OD), right column (OS). (A) Fundus photographs of the proband taken at age 76 showing typical choroideremia changes, with atrophy of the choroid and RPE. A small island of preserved RPE remains in the central macula, surrounded by atrophic peripheral areas of apparent bare sclera. (B)(C) Fundus autofluorescence image demonstrating areas of residual RPE (L) and the corresponding OCT image (R). Preserved retinal areas with normal autofluorescence exhibit thicker choroid and preserved retinal lamination. An outer retinal tubulation is seen in the right fundus.
Figure 3

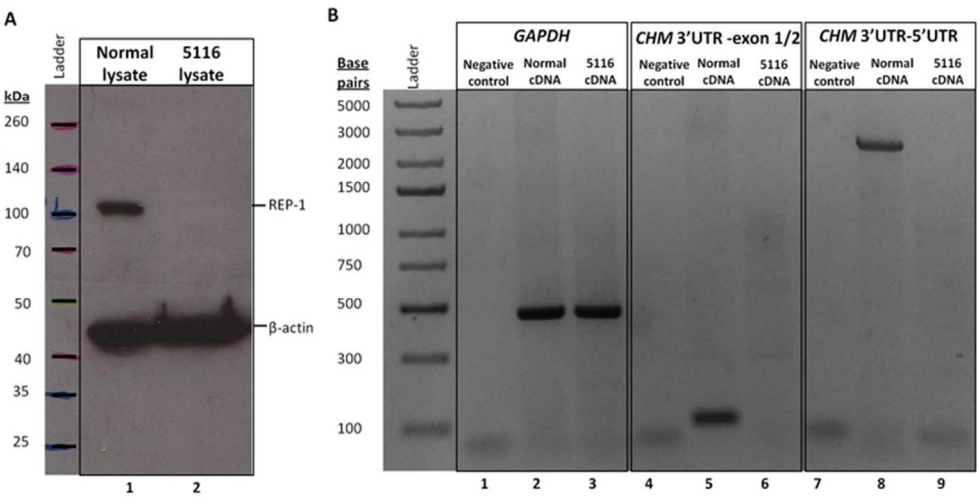
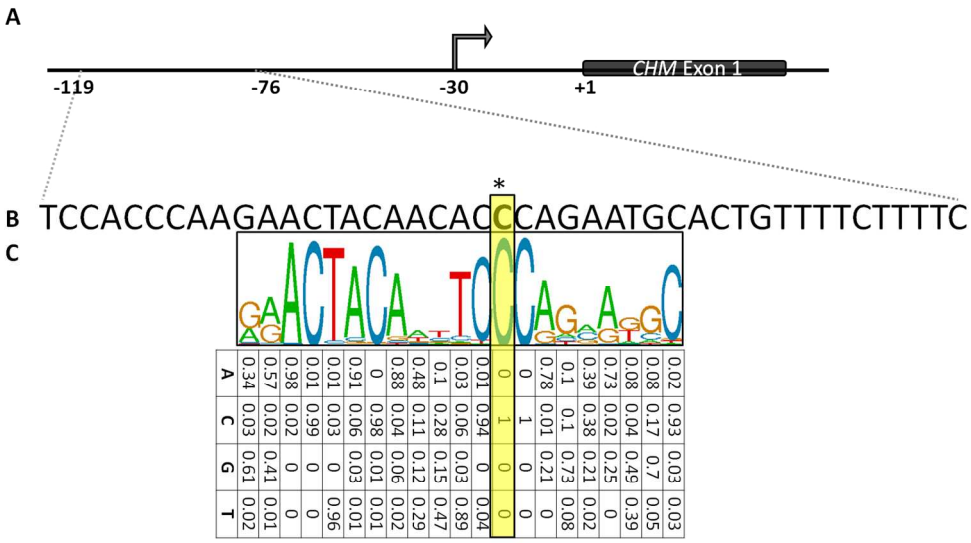


Figure 4. Molecular confirmation of choroideremia in patient 5116.
(A) Patient 5116 lacks REP-1. Western blot results show the absence of a ~100 kDa band corresponding to REP-1 in a lymphoblastoid cell line derived from the patient (lane 2), which is present in a normal control (lane 1). A β-actin antibody was used as a loading control to ensure an adequate protein sample in each lane, with the 42 kDa band present in both samples.
(B) Patient 5116 lacks CHM mRNA. cDNA synthesized from the mRNA harvested from a patient generated lymphoblastoid cell line was used as template for PCR. Lanes 2 and 3 show a 475 bp band resulting from the amplification of the GAPDH control housekeeping gene from 5116 and a normal control, indicating cDNA of adequate quality. Lanes 6 and 9 demonstrate an absence of amplification from the patient's cDNA of both partial and full length coding sequence, respectively, as compared to PCR products sized 93 and 2200 base pairs amplified from normal cDNA observed in lanes 5 and 10.
Figure 4
32x16mm (600 x 600 DPI)



A	0.34	0.57	0.98	0.01	0.01	0.91	0	0.88	0.48	0.1	0.03	0.01	0	0.78	0.1	0.39	0.73	0.08	0.08	0.02
C	0.03	0.02	0.02	0.99	0.03	0.06	0.98	0.04	0.11	0.28	0.06	0.94	1	0.01	0.1	0.38	0.02	0.04	0.17	0.93
G	0.61	0.41	0	0	0	0.03	0.01	0.06	0.12	0.15	0.03	0	0	0.21	0.73	0.21	0.25	0.49	0.7	0.03
T	0.02	0.01	0	0	0.96	0.01	0.01	0.02	0.29	0.47	0.89	0.04	0	0	0.08	0.02	0	0.39	0.05	0.03

Figure 5. Consensus binding sequences for transcription factor ZNF143.
(A) Partial map of the human CHM gene; arrow indicates transcription start site. (B) Expanded sequence of the minimal CHM promoter from c.-119 to c.-76, as identified through the analysis of progressive deletion constructs. Position c.-98 is marked with an asterisk. (C) Sequence logo derived from publically available ChIP-seq data released as part of the ENCODE project, with the position weighted matrix below (ENCODE Project Consortium, 2012). An invariant C is found at position corresponding to c.-98 of CHM.
Figure 5
439x257mm (96 x 96 DPI)

Patient	Age at exam	Sex	Genotype	OD/OS	BCVA	ERG Scotopic b-wave, µV	ERG Photopic b-wave, µV	ERG 30-Hz flicker, µV	Visual Field	Fundus
5116	76	M	c.-98C>A	OD	6/19	0	0	0.2	< 5 degrees	Widespread chorioretinal atrophy
				OS	6/15	0	0	0.2	< 5 degrees	Widespread chorioretinal atrophy
5149	-	M	c.-98C>A	Affected; not examined						
5113	33	F	c.-98C/A heterozygous	OD	6/6	Not tested				Carrier signs
				OS	6/6	↓50%	Normal	↓50%	Not tested	Carrier signs
5147		M	c.-98C	Unaffected; not examined						
111	35	M	c.-98C>T	OD	6/5	Not tested				Focal areas of chorioretinal atrophy in periphery
				OS	6/5	Not tested				Focal areas of chorioretinal atrophy in periphery
151	-	M	c.-98C>T	Data unavailable						Chorioretinal atrophy
161	37	M	c.-98C>T	OD	Count fingers vision	Not tested			Moderately constricted	Small central areas of functional retina
				OS	6/36	Not tested			Moderately constricted	Small central areas of functional retina
190	-	M	c.-98C	Unaffected; not examined						
192	-	F	c.-98C/T heterozygous	Not examined						

Homo sapiens (human)	tcca--cccaa-gaactacaa-cacccagaatgcact-g-----tttt-ctt--ttc
Pan troglodytes (chimp)	tcca--cccaa-gaactacaa-cacccagaatgcact-g-----tttt-cct--ttc
Pongo abelii (orangutan)	ttccatcccaa-gaactacaa-cacccagaatgcatt-g-----tttttcct--ttc
Callithrix jacchus (marmoset)	tcca--cccaa-gaactacaa-cacccagaatgcatt-g-----tttt-cct--ttc
Mus musculus (mouse)	gcgt--ctcta-gaactacaa-cacccagaatgcact-g-----tttt-tcc--ttc
Rattus norvegicus (rat)	gcta--cccta-aaactacaa-cacccagaatgcact-g-----tttt-tcc--ttc
Equus caballus (horse)	gcca--cccaa-gaactacaa-taccccagaatgcact-g-----tttt-ccc--ttc
Canis familiaris (dog)	gcca--cccac-gaactacaa-tatccagaatgcatt-g-----tttt-tcc--ttc
Pteropus vampyrus (megabat)	gcca--cccaa-ggactacaa-taccccagaatgcact---------tt-tcc----c
Dasypus novemcinctus (armadillo)	tcaa--cca-atgaactacaa-taccccagaatgcact-a-----tttt-cct--ttc
Monodelphis domestica (opossum)	tcct--gccac-gaactacaa-atccccagaatgcatt-g-----cgct-ccc----g
Anolis carolinensis (lizard)	----------a-gaactacag-caaccacaatgcact-gcaatatggt-cac--atc
Gallus gallus (chicken)	==
Xenopus tropicalis (X. tropicalis)	==
Dani rerio (zebrafish)	

Base c.-98 and its corresponding location in other species is highlighted