



Using Whole Genome Sequencing in an African Subphenotype of Myasthenia Gravis to Generate a Pathogenetic Hypothesis

Melissa Nel¹, Nicola Mulder², Tarin A. Europa¹ and Jeannine M. Heckmann^{1*}

¹ Neurology Research Group, Division of Neurology, Department of Medicine, University of Cape Town, Cape Town, South Africa, ² Computational Biology Division, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

OPEN ACCESS

Edited by:

Fulvio Cruciani,
Sapienza University of Rome, Italy

Reviewed by:

Henry Kaminski,
George Washington University,
United States
Linda L. Kusner,
George Washington University,
United States

*Correspondence:

Jeannine M. Heckmann
jeanine.heckmann@uct.ac.za

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 12 November 2018

Accepted: 11 February 2019

Published: 01 March 2019

Citation:

Nel M, Mulder N, Europa TA and Heckmann JM (2019) Using Whole Genome Sequencing in an African Subphenotype of Myasthenia Gravis to Generate a Pathogenetic Hypothesis. *Front. Genet.* 10:136. doi: 10.3389/fgene.2019.00136

Myasthenia gravis (MG) is a rare, treatable antibody-mediated disease which is characterized by muscle weakness. The pathogenic antibodies are most frequently directed at the acetylcholine receptors (AChRs) at the skeletal muscle endplate. An ophthalmoplegic subphenotype of MG (OP-MG), which is characterized by treatment resistant weakness of the extraocular muscles (EOMs), occurs in a proportion of myasthenics with juvenile symptom onset and African genetic ancestry. Since the pathogenetic mechanism(s) underlying OP-MG is unknown, the aim of this study was to use a hypothesis-generating genome-wide analysis to identify candidate OP-MG susceptibility genes and pathways. Whole genome sequencing (WGS) was performed on 25 AChR-antibody positive myasthenic individuals of African genetic ancestry sampled from the phenotypic extremes: 15 with OP-MG and 10 individuals with control MG (EOM treatment-responsive). Variants were called according to the Genome Analysis Toolkit (GATK) best practice guidelines using the hg38 reference genome. In addition to single variant association analysis, variants were mapped to genes (± 200 kb) using VEGAS2 to calculate gene-based test statistics and HLA allele group assignment was inferred through “best-match” alignment of reads against the IMGT/HLA database. While there were no single variant associations that reached genome-wide significance in this exploratory sample, several genes with significant gene-based test statistics and known to be expressed in skeletal muscle had biological functions which converge on muscle atrophy signaling and myosin II function. The closely linked *HLA-DPA1* and *HLA-DPB1* genes were associated with OP-MG subjects (gene-based $p < 0.05$) and the frequency of a functional A > G SNP (rs9277534) in the *HLA-DPB1* 3'UTR, which increases *HLA-DPB1* expression, differed between the two groups (G-allele 0.30 in OP-MG vs. 0.60 in control MG; $p = 0.04$). Furthermore, we show that rs9277534 is an *HLA-DPB1* expression quantitative trait locus in patient-derived myocytes ($p < 1 \times 10^{-3}$). The application of a SNP to gene to pathway approach to this exploratory WGS dataset of African myasthenic individuals, and comparing dichotomous subphenotypes, resulted

in the identification of candidate genes and pathways that may contribute to OP-MG susceptibility. Overall, the hypotheses generated by this work remain to be verified by interrogating candidate gene and pathway expression in patient-derived extraocular muscle.

Keywords: myasthenia gravis, African, whole genome sequencing, extraocular muscle, ophthalmoplegia, *HLA-DPB1*, extreme phenotype, association

INTRODUCTION

Myasthenia gravis (MG) is a rare, but treatable antibody-mediated disease which results in fatigable weakness of skeletal muscles, including extraocular (or eye) muscles. In most individuals this is a result of pathogenic antibodies targeting the acetylcholine receptors (AChR) at the neuromuscular junction, which cause activation of complement at the muscle endplate and consequent muscle damage (Engel et al., 1977).

Though the incidence of AChR-antibody positive MG in sub-Saharan Africa is similar to global figures (Mombaur et al., 2015), and the response to MG therapies overall is similar among populations (Heckmann et al., 2007), we have recognized a subphenotype of treatment-resistant ophthalmoplegia, or OP-MG, among a subset of MG subjects of African genetic ancestry (Heckmann et al., 2007; Heckmann and Nel, 2017). This OP-MG subphenotype is characterized by severe, persistent extraocular muscle (EOM) weakness and commonly affects subjects with juvenile onset, but otherwise characteristic AChR-antibody positive MG (i.e., generalized muscle weakness which responds to treatment). The pathogenesis of the OP-MG subphenotype remains unknown though we hypothesize that individuals who develop this subphenotype may harbor African susceptibility variants which impact on the MG disease process in the particular context of the EOMs.

A previous extended whole exome sequencing (WES) study of OP-MG subjects, including untranslated region (UTR) coverage, identified a number of putative regulatory variants (Nel et al., 2017). However, this study suffered from several limitations including false positive variant calls which could not be validated by Sanger sequencing (likely PCR related) and limited coverage of the non-coding genome (which is expected to harbor a greater burden of variants contributing to complex disease risk). Here we identified a number of OP-MG associated variants in the *HLA* class II region though it was not possible to verify them with Sanger sequencing due to the complexity of this region. This was interesting because the genetic basis of MG has been investigated for more than three decades in individuals of European genetic ancestry and the consistent finding has been the association of the class I and II *HLA* region with individuals by age at MG onset (Nel and Heckmann, 2018).

The focus of the present study was to perform PCR-free whole genome sequencing (WGS) in a well characterized cohort of OP-MG and control MG individuals, all AChR antibody-positive and differing only by the responsiveness of their EOMs to standard therapy. Although the sample is small ($n = 25$), this discovery cohort represented highly selected individuals from the phenotypic extremes and matched for ancestry to maximize

the power to detect association signals. Single nucleotide polymorphisms (SNPs) which were suggestive of association with OP-MG were validated in a larger cohort and a SNP to gene to pathway approach was used to prioritize genes based on skeletal muscle expression patterns.

MATERIALS AND METHODS

Patient Samples

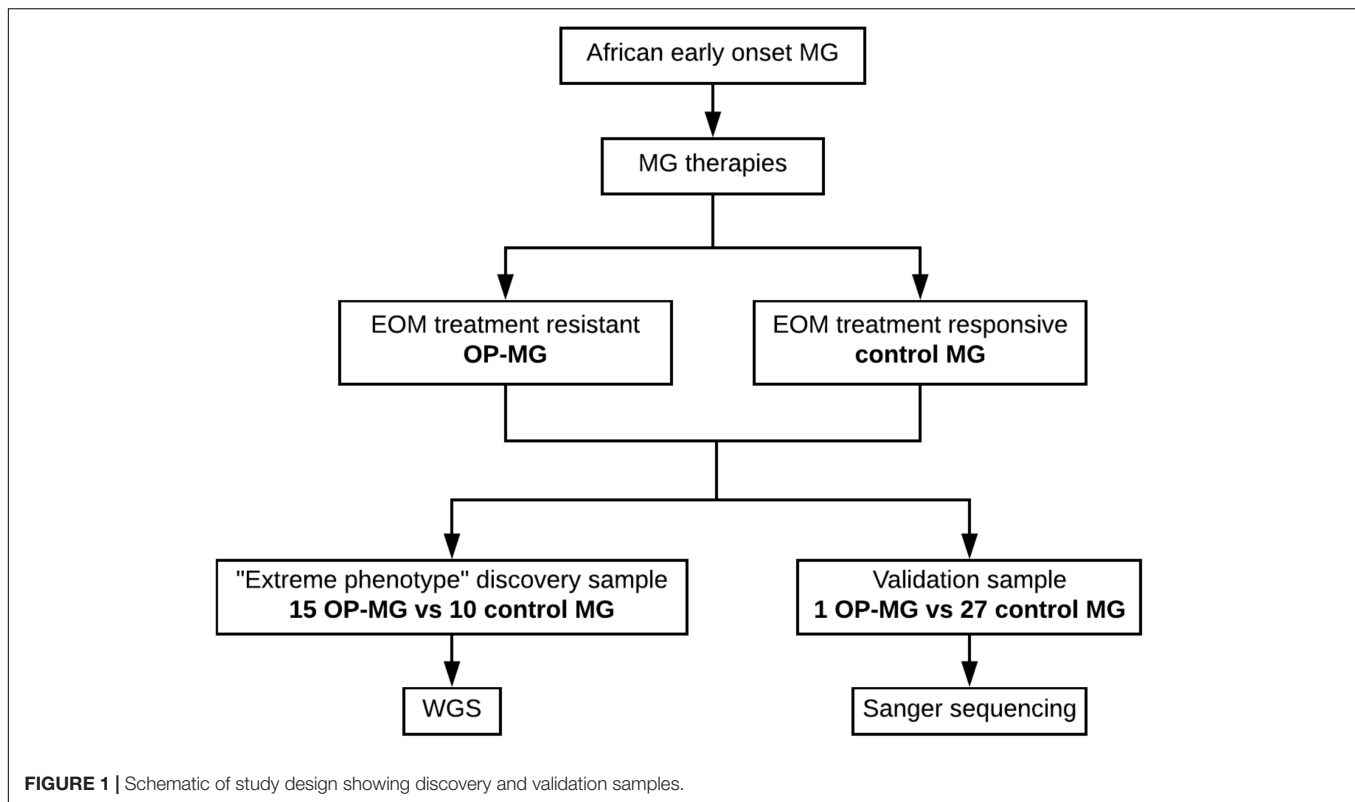
Patients with generalized myasthenia gravis (MG) of early-onset (<25 years) and African genetic ancestry (either black African or Cape mixed African ancestry) were recruited for WGS. This discovery sample represented the phenotypic extremes of treatment responsiveness to myasthenic-associated EOM weakness. The case group ($n = 15$) included individuals with OP-MG as previously described (Heckmann and Nel, 2017), defined as treatment resistant weakness of EOMs. The control group ($n = 10$) included individuals with no persistent EOM weakness, i.e., EOM weakness may have been present at disease presentation but responded appropriately to treatment. DNA samples from 28 African ancestry MG patients (1 OP-MG and 27 control MG) with early onset disease (<38 years) served as a validation sample to genotype selected variants. This study was approved by the UCT Faculty of Health Sciences Human Research Ethics Committee (HREC 591/2014) and all subjects gave written informed consent in accordance with the Declaration of Helsinki. The study design is outlined in **Figure 1**.

DNA Extraction and Whole Genome Sequencing

Genomic DNA was extracted from buffy coats of nucleated cells obtained from anticoagulated whole blood using the salting out method (Miller et al., 1988). Sequencing libraries (2×150 bp read length) were prepared from DNA samples using the TruSeq PCR-free library preparation kit (Illumina). Libraries were sequenced on Illumina HiSeq sequencing instruments ($30\times$ coverage) at the Kinghorn Centre for Clinical Genomics (Sydney, Australia) and the Centre for Genomic Regulation (Barcelona, Spain).

Read Alignment and Variant Calling

Paired end sequencing reads (FASTQ files) were aligned to the hg38 reference genome (including HLA contigs) using BWA MEM v0.7.15 (1000 Genomes Project Consortium et al., 2012) to generate BAM files. The Genome Analysis Toolkit best practice guidelines for germline SNPs and Indels were followed (GATK v3.7) (Van der Auwera et al., 2013) including



duplicate read removal and base quality score recalibration of BAM files followed by variant calling using Haplorecal (first on individual samples to generate GVCF files and then on the entire cohort to generate a final multisample VCF file) (Poplin et al., 2017). Variant quality score recalibration (VQSR) was performed separately for SNPs and Indels using a tranche sensitivity threshold of 99% to remove false positive calls. Variants were annotated using the Ensembl variant effect predictor (McLaren et al., 2016).

HLA Allele Determination

Reads aligning to the HLA region on chromosome 6 (33,064,568–33,080,777) and to the HLA contigs were extracted from the BAM files and realigned to reference sequences from the IMGT/HLA database (v3.29.0.1, 2017). HLA allele group assignment was inferred through “best-match” alignment of reads against the IMGT/HLA alleles using HLA Explore Software (Omixon).

Case Control Association Analysis

Autosomal, bi-allelic variants were extracted from the VCF file and PLINK v1.9 (Chang et al., 2015) was used to perform various quality control procedures prior to association testing. Variant level filtering (excluding variants with MAF <5%, call rate <95% and Hardy–Weinberg (HW) equilibrium p -value < 1×10^{-6} in controls) and sample level filtering (excluding individuals with outlying missing genotype or heterozygosity rates) was performed according to previously described guidelines (Anderson et al., 2010; Reed et al., 2015).

To exclude any large-scale differences in ancestry between the OP-MG and control MG groups (which could confound the case-control association analysis), principal component analysis of 50 357 variants was performed after LD based SNP pruning using PLINK v1.9 ($-indep-pairwise$ 1000 50 0.15). The allelic association of each marker with OP-MG was tested using Fisher’s exact test (considering the unpruned dataset) and the genomic inflation estimate (lambda) was calculated for the unadjusted model based on median χ^2 .

Gene and Pathway Based Analyses

As a complimentary approach to single variant association analysis, VEGAS2 was used to calculate gene (Mishra and Macgregor, 2015) and pathway (Mishra and MacGregor, 2017) based p -values. This software tool maps SNPs to genes based on their genomic location. We performed two analyses in parallel: stringent mapping (including SNPs within a gene plus any SNPs outside of the gene with $r^2 > 0.8$ with SNPs within the gene) and less stringent mapping (including SNPs within a gene plus any SNPs 200 kb upstream and downstream of the 5’UTR and 3’UTR boundaries). For each mapping approach, the p -values from each mapping SNP are aggregated accounting for the linkage disequilibrium (LD) between SNPs and correcting for the gene size (number of SNPs). To compute pathway-based test statistics, the gene-based test statistics for gene lists in curated pathways (multiple sources including BIOCARTA, REACTOME and KEGG databases) and custom pathways (mined from various sources of EOM gene expression data, **Supplementary Table S2**) are aggregated and corrected for pathway size bias.

Sanger Sequence Verification of Variants

Two variants were verified by Sanger sequencing in a validation sample of myasthenics with African genetic ancestry consisting of 1 OP-MG subject and 27 control MG subjects using the following primers: CCAGGCTGAGAGACAAAGCAGACC forward and CGTACTTATGTGCCACACAAGAC reverse for rs16834631 in *FAM92A1* and GATGGAGCTTCCGGAAGTCTTGG forward and CAAGGCAACTGCCTCTCTGCACC reverse for rs7816955 in *PEF1*.

Cell Cultures

Dermal fibroblasts from 10 OP-MG and 5 control MG individuals were obtained from skin punch biopsies using the explant method. These were transduced with an RGD fiber modified adenovirus containing a human MyoD transgene as previously described (Nel et al., 2019). Briefly, transduced fibroblasts were maintained in differentiation medium (DMEM + 5% horse serum + 1% P/S) for 48 h to induce myogenic transdifferentiation and generate myocytes. Myocytes stained positively for sarcomeric myosin and successful myogenic transdifferentiation was further confirmed by demonstrating muscle-specific gene expression in myocytes (*CHRNA1*, *MYOD1*, and *MYOG*). Importantly, based on muscle-specific gene expression levels, the degree of myogenic transdifferentiation was similar in both OP-MG and control MG myocytes. To mimic MG-induced gene expression changes *in vitro*, myocytes were stimulated with 5% homologous MG sera for 24 h before harvesting RNA. Sera samples were sourced from AChR antibody-positive, treatment-naïve MG patients with generalized myasthenia and severe extraocular muscle involvement.

Quantitative Polymerase Chain Reaction (qPCR)

RNA was extracted from myocytes using the HighPure RNA extraction kit (Roche) according to the kit protocol. RNA concentration and purity was determined using the Nanodrop® ND1000 spectrophotometer [Thermo Scientific and all ratios were within the recommended ranges ($A_{260}/_{280} = 1.8\text{--}2.0$; $A_{260}/_{230} > 1.7$)]. 400 ng total RNA was reverse transcribed to cDNA using the RT² First Strand Kit (Qiagen) according to the manufacturer's specifications. Quantitative PCR was performed

on the cDNA samples using proprietary Quantitect primer assays (Qiagen) (*RPLP0*, *HLA-DPB1*) and RT² SYBR Green Mastermix (Qiagen) on the 7900HT Fast Real-Time PCR System (Applied Biosystems). *RPLP0* was selected from a panel of 10 reference genes which were screened for their expression stability in myocytes (Nel et al., 2019). Individual data points were calculated as $2^{-\Delta Cq}$, where $\Delta Cq = \text{target gene Cq} - \text{reference gene Cq}$ (Schmittgen and Livak, 2008).

Data Visualization

Quantile–Quantile (Q–Q) and manhattan plots were created in R (version 3.5.1) using the qqman package (Turner, 2018). The heatmap of skeletal muscle tissue RNAseq expression data from the Genotype-Tissue Expression (GTEx) project was generated using the GTExPortal [1]. Graphs of qPCR expression data were created using Prism 7 (version 7.0c).

Computation

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: hpc.uct.ac.za and the Bioinformatics Unit at the Centre for Genomic Regulation (CRG), Barcelona.

RESULTS

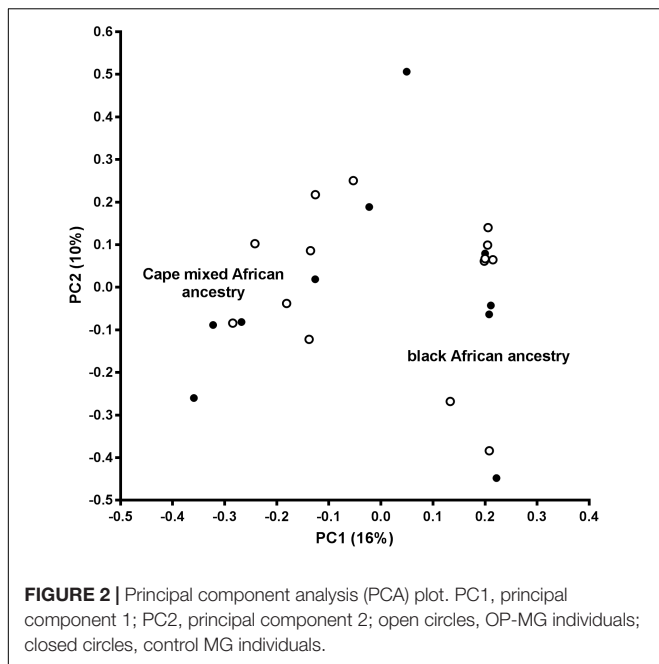
Clinical Characteristics of Study Participants

The clinical characteristics of the study participants are summarized in **Table 1**. For the WGS discovery sample, all subjects had early onset MG and there was no significant difference in the age of disease onset between OP-MG and control MG groups (14 years vs. 16 years, $p = 0.450$), or the sex ratios ($p = 0.13$). The WGS sample comprised 11 black African ancestry individuals (44%) and 14 Cape mixed African ancestry (M/A) individuals (56%) with similar ancestry proportions in OP-MG and control MG groups. While the sex and ancestry proportions were similar between the WGS and validation samples, the age at MG onset was significantly higher in the validation sample compared to the WGS sample ($p = 4 \times 10^{-8}$). This was primarily because we tried to

TABLE 1 | Demographic characteristics of the study participants.

	WGS sample		P-value	Validation sample		P-value
Clinical characteristics	OP-MG (<i>n</i> = 15)	Control MG (<i>n</i> = 10)	OP-MG vs. control MG	1 OP-MG + 27 control MG (<i>n</i> = 28)	WGS vs. validation sample	
Age at disease onset yrs (IQR)	14 (11–17)	16 (15–22)	0.450	26 (22–31)	4×10^{-8}	
Sex						
Female <i>n</i> (%)	9 (60)	9 (90)	0.131	23 (82)	0.404	
Male <i>n</i> (%)	6 (40)	1 (10)		5 (18)		
Ancestry						
Black African <i>n</i> (%)	8 (53)	3 (30)	0.288	6 (21)	0.091	
Cape mixed African <i>n</i> (%)	7 (47)	7 (70)		22 (79)		

IQR, interquartile range.



reduce confounders for the highly selected sample undergoing WGS by matching for age which was previously identified as a biological factor.

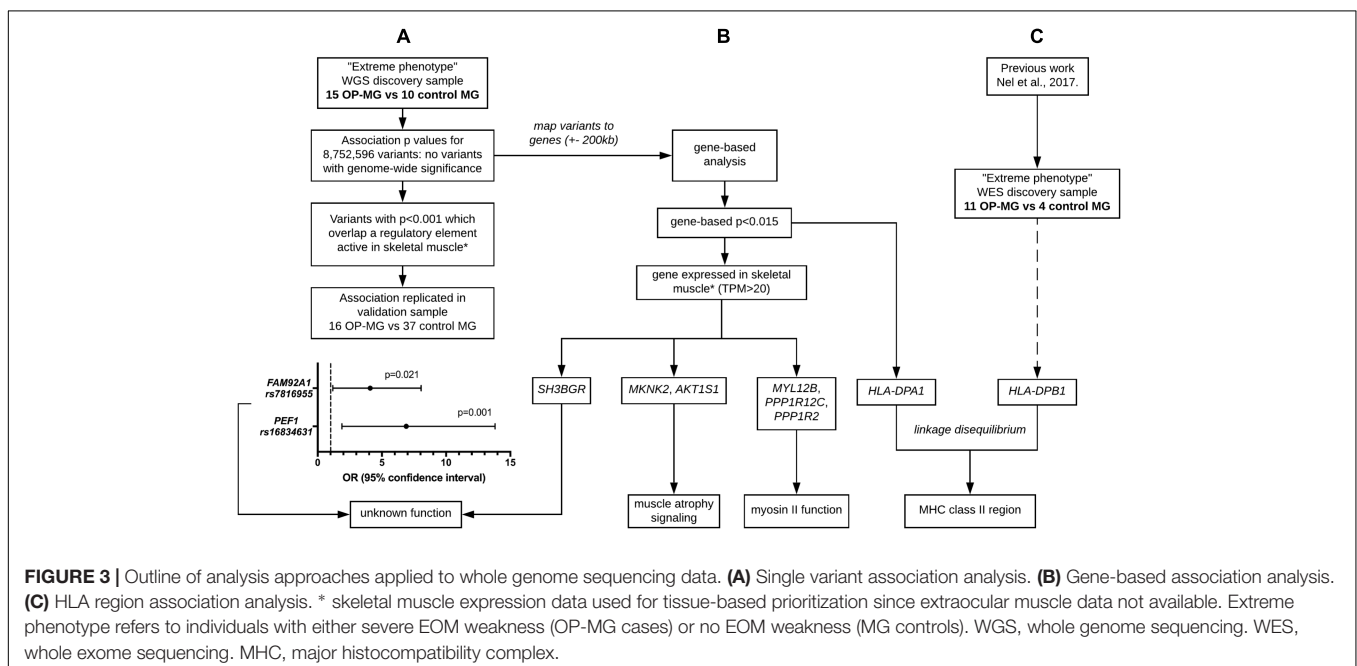
Description of Variants

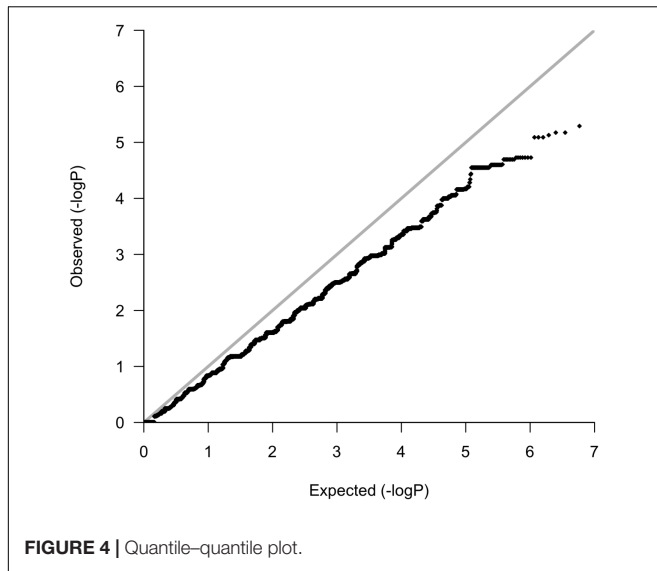
The final VQSR filtered callset contained ~18 million variants, including ~2 million (11%) novel variants, with an overall Ti/Tv ratio of 2.12 and a heterozygous/homozygous ratio of 2.06 which is in line with previously published genome-wide quality control metrics, particularly for African datasets

(DePristo et al., 2011; Guo et al., 2014). A high proportion of the detected variants were singletons 29% (~5 million). Overall there were ~5 million variants per genome which is consistent with previously published data for African populations (Auton et al., 2015).

Population Structure

An assessment of the population structure within the dataset was investigated using principal component analysis (PCA) after LD based SNP pruning. Combined, principal components 1 and 2 explain 26% of the total variance within the dataset; these are visualized on the PCA plot shown in **Figure 2**. The samples segregate into two clusters reflecting the black African and Cape M/A groups. Present day South Africans include a major ethnolinguistic group of black African South-Eastern Nguni-language (isiXhosa and isiZulu) speakers. The Cape M/A ancestry population (predominantly Khoisan and Nguni-speaking African ancestry as well as smaller genetic contributions from Europeans and Southeast Asians) (De Wit et al., 2010; Quintana-Murci et al., 2010) comprise the most prevalent sub-population in the Western Cape region where this study was conducted. Despite their shared African ancestry with the black African ancestry individuals, the Cape M/A ancestry individuals form a dispersed but distinct cluster reflecting the admixed nature of this population which has considerable ancient African hunter-gatherer (Khoisan) and lesser non-African genetic contributions (Choudhury et al., 2017). Two out of the three outlier samples in the black African ancestry cluster represent individuals from other African countries (Zimbabwe and Burundi). Importantly, OP-MG and control MG individuals are equally represented in both ancestry groups which indicates that the case control association analysis will not be confounded by differences in population structure.





Analysis of WGS Data to Identify Association Signals

Various approaches were used in parallel (outlined in **Figure 3**) to identify OP-MG associated variants and genes. The results of these analyses were interpreted in conjunction with our previous work, involving WES of OP-MG and control MG subjects (Nel et al., 2017), to collectively generate hypotheses regarding OP-MG susceptibility pathways.

Single Variant Association Analysis

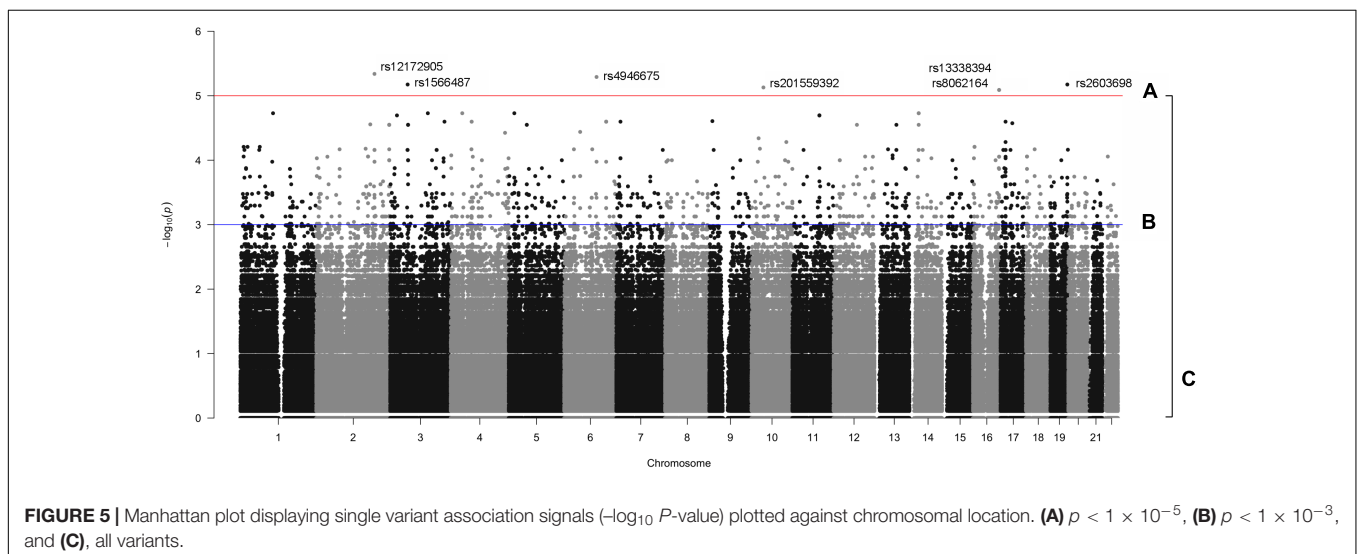
Following variant and sample level filtering, the frequency of 8,752,596 variants were compared between case and control groups (i.e., OP-MG vs. control MG) using Fisher's exact test (**Figure 3A**). The black points in the quantile-quantile (Q-Q) plot in **Figure 4** show the observed p values (sorted from largest to smallest) plotted against the expected p -values from a

theoretical χ^2 -distribution (Ehret, 2010). The gray straight line in the Q-Q plot indicates the distribution of SNPs under the null hypothesis. The black points form a straight line which is “deflated” relative to the gray line suggesting that the analyses were underpowered due to the small sample size in this study. Consequently, there were no variant associations which reached genome-wide significance ($p < 5 \times 10^{-8}$).

The manhattan plot shows 7 variants with suggestive association with either the OP-MG or control MG phenotype ($p < 1 \times 10^{-5}$, **Figure 5A**) which are summarized in **Supplementary Table S1**. Five out of 7 variants had a lower frequency in OP-MG compared to control MG and all variants are common in African populations (1000 genomes data). None of these top associated variants had any predicted functional consequences. While this may be true of many top GWAS hits in studies using chip data, where the top SNPs may not themselves be pathogenic but may “tag” other functional variants in LD, this is an unlikely scenario in our study since we have genome-wide variant coverage.

Therefore, in order to further prioritize variants with sub-genome wide significance thresholds, we screened the VEP “impact” annotations of 1,751 variants with $p < 0.001$ (**Figure 5B**). Seven variants were classified as “low impact” (splice region, intron and synonymous variants) and two variants were classified as “moderate impact” (missense variants) but were predicted to be benign by various prediction tools. The remainder of the variants were classified as “modifiers” and included intergenic variants and variants in up- and downstream gene regions, some of which overlapped regulatory features.

Since non-coding genetic variation was hypothesized to contribute to OP-MG susceptibility, we applied a tissue-specific prioritization approach to identify which modifier variants overlapped a regulatory feature active in muscle [human skeletal muscle myoblast and myotube (HSMM and HSMMtube) samples from the ENCODE project and psoas muscle samples from the Roadmap Epigenomics Project]. Muscle samples were chosen since there is no publically available expression data for human



EOM. This analysis identified 13 variants which were more common in OP-MG compared to control MG. Two upstream gene variants, rs7816955 in *FAM92A1* ($p = 9.2 \times 10^{-4}$) and rs16834631 in *PEF1* ($p = 1.5 \times 10^{-4}$), overlapped Ensembl regulatory features classified as active promoters based on epigenome activity in relevant muscle cell lines. While the *FAM92A1* variant did not overlap any Ensembl motif features, the *PEF1* variant overlapped 24 putative transcription factor binding sites based on binding matrices, one of which was a high information position with predicted decreased binding of the RFX3::FIGLA transcription factor pair.

Both variants had a reported frequency ≤ 0.30 among African controls and ≤ 0.10 among European controls (1000 Genomes Project) and were validated by Sanger sequencing in the WGS sample. Their frequency was also determined in a validation sample ($n = 28$) which confirmed the association of these variants with OP-MG: *PEF1* rs16834631 0.57 in OP-MG vs. 0.16 in control MG ($p = 0.001$) and *FAM92A1* rs7816955 0.47 in OP-MG vs. 0.18 in control MG ($p = 0.021$; **Figure 3A**).

Gene-Based Association Analysis

A single variant association testing approach, while unbiased, is limited by stringent genome-wide significance thresholds which are difficult to reach after correcting for multiple testing (particularly relevant with our small sample size). Searching for association signals in single variants assumes that all affected individuals (i.e., OP-MG cases) have the same pathogenic variant(s) which does not fit with our current understanding of the genetic architecture of complex disease, which may be attributed to the joint effect of many causal loci with small effect sizes (Fu et al., 2013). To interrogate the collective biological meaning of the sub-threshold single variant associations, all variants (**Figure 5C**) were mapped to genes and their modest association signals were aggregated using VEGAS2 to derive gene based p -values (**Figure 3B**). A mapping threshold of 200 kb upstream and downstream of gene boundaries was chosen since this distance has been shown to increase the number of significant phenotype-pathway associations, particularly for autoimmune diseases (Brodie et al., 2016).

While no genes had significant p -values after correcting for multiple testing of 23,361 genes, 38 genes had a p -value ≤ 0.015 . These were prioritized by determining their tissue expression using RNAseq expression data from the Genotype-Tissue Expression (GTEx) project (Aguet et al., 2017). Since there is no available expression data for the specific allotype of EOM, we prioritized genes based on their expression level in skeletal muscle tissues. Eleven genes had a medium expression level in skeletal muscle defined as a transcripts per million (TPM) value of 11–1,000 (shown in blue boxes in **Figure 6**). The functions of proteins encoded by genes with TPM > 20 are summarized in **Table 2**.

Pathway-Based Association Analysis

The sample size was not sufficient to produce meaningful pathway-based test statistics from the VEGAS2 pathway analysis which interrogated both curated and custom pathways.

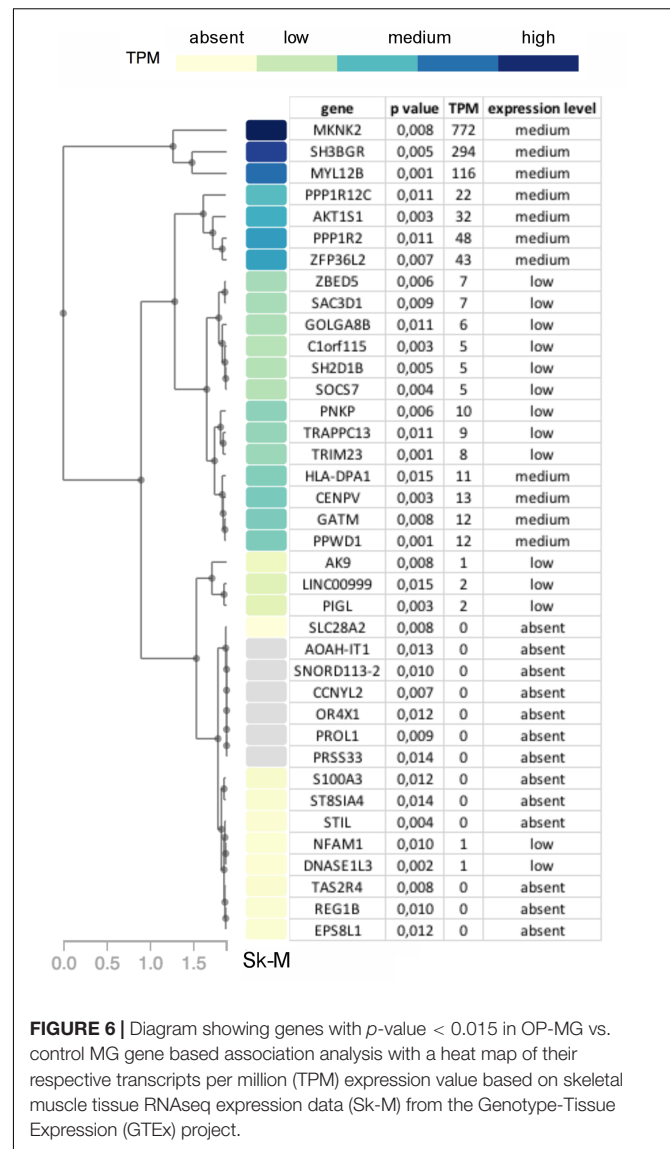


FIGURE 6 | Diagram showing genes with p -value < 0.015 in OP-MG vs. control MG gene based association analysis with a heat map of their respective transcripts per million (TPM) expression value based on skeletal muscle tissue RNAseq expression data (Sk-M) from the Genotype-Tissue Expression (GTEx) project.

HLA Region Associations

In our previous work we identified a unique “HLA signature” spanning the class II region of the MHC in OP-MG subjects (Nel et al., 2017) (**Figure 3C**) and the gene-based analysis in the present study also identified association signals in this region (*HLA-DPA1* $p = 0.015$ and *HLA-DPB1* $p = 0.033$). We therefore performed HLA typing (see section “HLA Allele Determination”) to interrogate differences in *HLA-DPA1* and *HLA-DPB1* allele frequencies between OP-MG and control groups. In our sample, *HLA-DPB1* allele diversity (12 alleles plus ambiguous alleles for 6 individuals) was higher than *HLA-DPA1* allele diversity (5 alleles) which is similar to studies in European populations (Hollenbach et al., 2012). We found differences in the frequency of 3 *HLA-DP* alleles between OP-MG and control MG (**Table 3**). Interestingly, for the *HLA-DPB1* locus, where alleles can be divided into two groups based on their associated *HLA-DPB1* expression levels, we found that the proportion of “low expression” and “high

TABLE 2 | Top 7 genes (based on GTEx skeletal muscle expression data, TPM > 20) with the muscle-specific function of their encoded proteins.

Gene	Function of encoded protein in skeletal muscle
<i>MKNK2</i>	Increased expression in muscle atrophy ¹ Downstream inhibitor of IGF1/Akt/mTOR hypertrophy signaling ²
<i>AKT1S1</i>	Subunit of mTORC1 multi-protein complex
<i>SH3BGR</i>	Critical for sarcomere formation in striated muscle tissues ³ Higher expression in mouse EOM compared to limb fibroblasts ⁴
<i>MYL12B</i>	Myosin regulatory light chain which maintains the integrity of myosin I and II ⁵
<i>PPP1R12C</i>	Encodes a subunit of myosin phosphatase which dephosphorylates myosin II ⁶
<i>PPP1R2</i>	Binds to the catalytic subunit of protein phosphatase 1, a subunit of myosin phosphatase
<i>ZFP36L2</i>	RNA-binding protein which is downregulated during satellite cell activation ⁷

¹Hu et al., 2012; ²Egerman and Glass, 2014; ³Jang et al., 2015; ⁴Kusner et al., 2010; ⁵Park et al., 2011; ⁶Ito et al., 2004; ⁷Galloway et al., 2016.

expression” alleles differed between the OP-MG and control MG groups ($p = 0.021$). The *HLA-DPA1*105:01* allele, the most common “low expression” allele in our sample and only observed in OP-MG individuals, appears to be common in African populations. The expression level of *HLA-DPB1* alleles was shown to be correlated with the genotype at rs9277534, a functional *A > G* SNP located in the 3'UTR of *HLA-DPB1* (Thomas et al., 2012). The *G*-allele of this SNP increases *HLA-DPB1* expression levels by altering the binding affinity of various microRNAs (Shieh et al., 2018). In keeping with the observed *HLA-DPB1* frequency differences (Schöne et al., 2018), we found a higher frequency of the rs9277534 *G*-allele in the control MG group.

HLA-DPB1 rs9277534 genotype-expression correlations have been demonstrated in blood (Yamazaki et al., 2018) but there is no data on this expression quantitative trait locus (eQTL) in skeletal muscle tissue. We therefore analyzed *HLA-DP* expression grouped by rs9277534 genotype in myocytes derived from transdifferentiated dermal fibroblasts from OP-MG and control MG subjects (Nel et al., 2019) and found that the *G*-allele increased *HLA-DPB1* expression levels (**Figure 7**, $p < 1 \times 10^{-3}$).

DISCUSSION

In this study we have used various strategies to mine WGS data in an attempt to generate hypotheses regarding the pathogenetic basis of a subphenotype of a rare autoimmune disease, myasthenia gravis. The subphenotype is characterized by treatment resistance of the eye muscles, or EOMs, whereas the non-ocular muscles respond to standard MG therapies (Heckmann and Nel, 2017). EOM is a specific allotype of muscle tissue because it differs from limb muscles in many respects (Porter et al., 2001). Since only a proportion of MG subjects develop the OP-MG subphenotype, the pool of affected individuals available for genetic studies is small. Nonetheless, we employed a focused strategy using extreme subphenotype sampling of OP-MG cases vs. MG disease controls to perform a genome wide analysis. Putative OP-MG susceptibility variants, genes and pathways were identified following prioritization based on known tissue-specific expression patterns in skeletal muscle since gene expression data for EOM is not available.

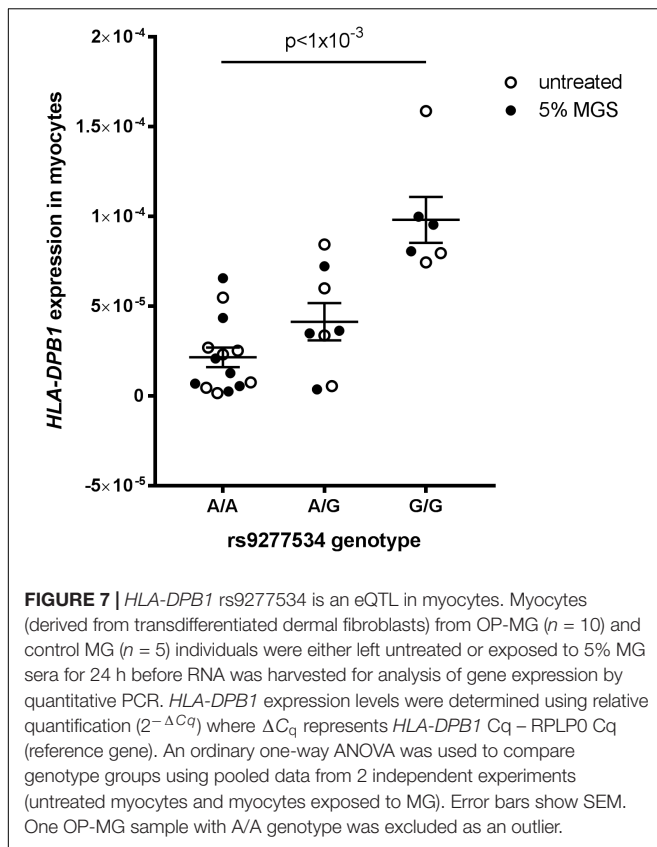
We have identified three main candidate pathogenic themes which we postulate are involved in developing OP-MG, and preliminary functional studies show at least some support for these hypotheses. Briefly, we summarize evidence gleaned from other areas, using the principle of triangulation, to lend support to the generated hypotheses.

The first two themes relate to muscle atrophy and muscle recovery/remodeling. The EOMs may be more susceptible to complement-mediated muscle endplate injury during MG (in Soltys et al., 2008) due to their relatively lower expression levels of complement regulatory proteins, particularly decay accelerating factor (DAF) (Kaminski et al., 2004). We previously screened the *DAF* gene in OP-MG subjects and found a higher frequency of a functional *DAF* promoter polymorphism compared to controls which impaired transcriptional upregulation of *DAF* expression in patient-derived cell lines following a lipopolysaccharide immune stimulus (Heckmann et al., 2010). Also, clinically and at surgery, the EOMs in the most severe cases of OP-MG are thin/atrophic, not fibrotic and unable to generate muscle force (Heckmann and Nel, 2017). Although there is limited histological data on EOMs in MG, neurogenic atrophy is a common pathological observation in the muscle biopsies of MG cases, (Oosterhuis and Bethlem, 1973) and likely to be the result of “functional denervation,” or the disconnection between the

TABLE 3 | Frequency of *HLA-DPA1* and *HLA-DPB1* alleles and rs9277534 *A > G* in OP-MG and control MG subgroups and population controls.

<i>HLA-DP</i> alleles	OP-MG freq <i>n</i> = 15*	Control MG freq <i>n</i> = 10*	<i>p</i> -value OP-MG vs. control MG	African control freq ¹ <i>n</i> = 48	Caucasian control freq ¹ <i>n</i> = 75	<i>p</i> -value African vs. Caucasian control
<i>HLA-DPA1*01:03</i>	15 (0.50)	3 (0.15)	$p = 0.009$	30 (0.31)	115 (0.77)	$p < 1 \times 10^{-7}$
<i>HLA-DPA1*02:01</i>	3 (0.10)	9 (0.45)	$p = 0.007$	23 (0.24)	26 (0.17)	0.242
<i>HLA-DPB1*105:01</i> (low expression ²)	7 (0.39)	0	$p = 0.009$	25 (0.26)	2 (0.01)	$p < 1 \times 10^{-7}$
<i>HLA-DPB1</i> rs9277534 <i>A > G</i> (high expression ³)	9 (0.30)	12 (0.60)	$p = 0.04$	789 (0.60)	317 (0.32)	$p < 1 \times 10^{-7}$

¹Goldfein, 2017; ²Schöne et al., 2018; ³Thomas et al., 2012. *For *HLA-DPB1*, allele assignment was ambiguous for 6 OP-MG and 3 control MG individuals and these samples were excluded from the analysis.



nerve and muscle endplate secondary to MG-induced damage (Nakano and Engel, 1993). With that in mind, the gene-based analysis identified two genes (*MKNK2*, *AKT1S1*) involved in the IGF1/AKT/mTOR pathway, which is a key pathway in promoting muscle atrophy following denervation (Tang et al., 2014) (Table 2). In keeping with these unbiased findings, our previous gene expression profiling of OP-MG myocytes using a panel of genes relevant in several MG studies found expression of genes from this pathway (*IGF1*, *AKT1*, and *AKT2*) were strongly correlated in OP-MG myocytes but not in the myocytes from control MG cases (Nel et al., 2019). Interestingly, *IGF1* is highly expressed in EOMs where it regulates both the muscle mass and force generation of these muscles, and its signaling is dysregulated in paralyzed EOMs (Altick et al., 2012).

Subsequent to MG damage we would expect the EOMs to undergo “regeneration” or remodeling due to their high numbers of resident satellite cells (McLoon and Wirtschafter, 2003), and this process requires the synthesis of new structural muscle proteins. We were therefore interested to observe that 3 of the 7 genes (*MYL12B*, *PPP1R12C*, and *PPP1R2*) identified by the gene-based analysis (Table 2) are involved in the stability and regulation of myosin II which is a prominent isoform in EOMs expressed by fast type IIA and IIB muscle fibers, respectively (Park et al., 2011).

While unbiased, genome-wide association studies (GWAS) typically employ very large samples, the application of this approach to the study of susceptibility to MG, has not been very

informative in terms of identifying new disease loci. In two recent GWAS in MG, the strongest association signals identified were localized to the *HLA* region (Gregersen et al., 2012; Renton et al., 2015), which was already identified over 3 decades ago in a small case-control sample (Compston et al., 1980). The third theme we identified relates to the *HLA* region since we found an association signal with lower *HLA-DPB1* expression in OP-MG which results from a functional polymorphism in the 3'UTR of *HLA-DPB1*. Although the *HLA-DP* locus is not in LD with other *HLA* loci, the expression levels of *HLA-DPB1* are increasingly recognized to have clinical relevance (Fleischhauer, 2015). While the main MG susceptibility locus lies in the class I or II region depending on the age at symptom onset (Nel and Heckmann, 2018), *HLA-DPB1* alleles may influence the phenotypic manifestations of the MG disease process in different individuals.

We also identified association signals in *PEF1* and *FAM92A1* which were validated in an independent sample, although the functional relevance of these genes in EOM is unknown. This highlights the importance of validating the hypotheses generated by this work in patient-derived EOM tissue, preferably from OP-MG individuals. It is worth noting that candidate gene associations, such as those previously identified in the regulatory region of *DAF*, were not identified following the filtering criteria used in this study. This is likely due to the sample size constraints imposed by WGS which limits the ability to detect significant associations for low frequency variants such as *DAF-198 C > G*. This SNP had a frequency of 0.13 among the OP-MG subjects in this study ($p = 0.119$) which is comparable to the statistically significant association previously reported using a larger sample size (0.12, $p = 0.001$) (Heckmann et al., 2010), albeit with an overlap of two OP-MG samples between the two studies.

In conclusion, despite the limitations of using a small sample to mine whole genome data to generate pathogenic hypotheses in a structured yet unbiased approach, several lines of evidence suggest we have achieved our aims. The next step will be to analyze the functionality of these genes and pathways in patient-derived extraocular muscle tissue.

DATA AVAILABILITY

The whole genome sequencing data on which the findings of this manuscript are based, have been deposited in the European Genome-Phenome Archive (EGA): <https://www.ebi.ac.uk/ega/home> and can be found under the following accession ID: EGAS00001003462. Access to the dataset is governed by a data access committee.

AUTHOR CONTRIBUTIONS

MN designed the myocyte model, performed the functional studies in myocytes, performed the genomic data analysis, and wrote the manuscript. NM provided computing resources, bioinformatics support, and editorial input. TE performed the qPCR experiments on ocular fibroblasts. JH conceived and

designed the study, collected DNA samples and clinical data, and provided editorial input and funding support.

FUNDING

MN was funded by a Novartis-Africa Mobility award and UCT Faculty of Health Sciences postdoctoral research fellowship. JH received funding from the French Muscular Dystrophy Association (AFM-Téléthon) (20049), National Research Foundation (NRF) of South Africa (113416), and the Rare Disease Foundation microgrant program and the BC Children's Hospital Foundation (2016).

REFERENCES

- 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632
- Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213. doi: 10.1038/nature24277
- Altick, A. L., Feng, C. Y., Schlauch, K., Alan Johnson, L., and von Bartheld, C. S. (2012). Differences in gene expression between strabismic and normal human extraocular muscles. *Investig. Ophthalmol. Vis. Sci.* 53, 5168–5177. doi: 10.1167/iovs.12-9785
- Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., and Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nat. Protoc.* 5, 1564–1573. doi: 10.1038/nprot.2010.116
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- Brodie, A., Azaria, J. R., and Ofran, Y. (2016). How far from the SNP may the causative genes be? *Nucleic Acids Res.* 44, 6046–6054. doi: 10.1093/nar/gkw500
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8
- Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardiën, S., Botha, G., et al. (2017). Whole-genome sequencing for an enhanced understanding of genetic variation among south africans. *Nat. Commun.* 8, 1–12. doi: 10.1038/s41467-017-00663-9
- Compston, D. A., Vincent, A., Newsom-Davis, J., and Batchelor, J. R. (1980). Clinical, pathological, HLA antigen and immunological evidence for disease heterogeneity in myasthenia gravis. *Brain A. J. Neurol.* 103, 579–601. doi: 10.1093/brain/103.3.579
- De Wit, E., Delpont, W., Rugamika, C. E., Meintjes, A., Möller, M., van Helden, P. D., et al. (2010). Genome-wide analysis of the structure of the south african coloured population in the western cape. *Hum. Genet.* 128, 145–153. doi: 10.1007/s00439-010-0836-1
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Egerman, M. A., and Glass, D. J. (2014). Signaling pathways controlling skeletal muscle mass. *Crit. Rev. Biochem. Mol. Biol.* 49, 59–68. doi: 10.3109/10409238.2013.857291
- Ehret, G. B. (2010). Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. *Curr. Hypertens. Rep.* 12, 17–25. doi: 10.1007/s11906-009-0086-6
- Engel, A. G., Lambert, E. H., and Howard, F. M. (1977). Immune complexes (IgG and C3) at the motor end-plate in myasthenia gravis: ultrastructural and light microscopic localization and electrophysiologic correlations. *Mayo Clin. Proc.* 52, 267–280.

ACKNOWLEDGMENTS

MN wishes to thank Dr. Julia Ponomarenko (head of the Bioinformatics Unit at the CRG) and her team for the provision of bioinformatics training.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00136/full#supplementary-material>

- Fleischhauer, K. (2015). Immunogenetics of HLA-DP — A new view of permissible mismatches. *N. Engl. J. Med.* 373, 669–672. doi: 10.1056/NEJMe1505539
- Fu, W., O'Connor, T. D., and Akey, J. M. (2013). Genetic architecture of quantitative traits and complex diseases. *Curr. Opin. Genet. Dev.* 23, 678–683. doi: 10.1016/j.gde.2013.10.008
- Galloway, A., Saveliev, A., Ukasiak, S., Hodson, D. J., Bolland, D., Balmanno, K., et al. (2016). RNA-binding proteins ZFP36L1 and ZFP36L2 promote cell quiescence. *Science* 352, 453–459. doi: 10.1126/science.aad5978
- Goldfein, H. (2017). *The Association of HLA Class II Genetic and Expression Level Variation with Response to the Hepatitis B Vaccine in South African Laboratory Workers*. dissertation thesis. Johannesburg: University of the Witwatersrand.
- Gregersen, P. K., Kosoy, R., Lee, A. T., Lamb, J., Sussman, J., McKee, D., et al. (2012). Risk for myasthenia gravis maps to a 151Pro→Ala change in TNIP1 and to human leukocyte antigen-B*08. *Ann. Neurol.* 72, 927–935. doi: 10.1002/ana.23691
- Guo, Y., Zhao, S., Sheng, Q., Ye, F., Li, J., Lehmann, B., et al. (2014). Multi-perspective quality control of illumina exome sequencing data using QC3. *Genomics* 103, 323–328. doi: 10.1016/j.ygeno.2014.03.006
- Heckmann, J. M., and Nel, M. (2017). A unique subphenotype of myasthenia gravis. *Ann. N. Y. Acad. Sci.* 1412, 14–20. doi: 10.1111/nyas.13471
- Heckmann, J. M., Owen, E. P., and Little, F. (2007). Myasthenia gravis in south africans: racial differences in clinical manifestations. *Neuromuscul. Disord.* 17, 929–934. doi: 10.1016/j.nmd.2007.07.002
- Heckmann, J. M., Uwimpuhwe, H., Ballo, R., Kaur, M., Bajic, V. B., and Prince, S. (2010). A functional SNP in the regulatory region of the decay-accelerating factor gene associates with extraocular muscle pareses in myasthenia gravis. *Genes Immun.* 11, 1–10. doi: 10.1038/gene.2009.61
- Hollenbach, J. A., Madbouly, A., Gragert, L., Vierra-Green, C., Flesch, S., Spellman, S., et al. (2012). A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer. *Immunogenetics* 64, 559–569. doi: 10.1007/s00251-012-0615-3
- Hu, S.-I., Katz, M., Chin, S., Qi, X., Cruz, J., Ibejunjo, C., et al. (2012). MNK2 inhibits eIF4G activation through a pathway involving serine-arginine-rich protein kinase in skeletal muscle. *Sci. Signal.* 5:ra14–ra14. doi: 10.1126/scisignal.2002466
- Ito, M., Nakano, T., Erdödi, F., and Hartshorne, D. J. (2004). Myosin phosphatase: Structure, regulation and function. *Mol. Cell. Biochem.* 259, 197–209. doi: 10.1023/B:MCBI.0000021373.14288.00
- Jang, D. G., Sim, H. J., Song, E. K., Medina-Ruiz, S., Seo, J. K., and Park, T. J. (2015). A thioredoxin fold protein Sh3bgr regulates enah and is necessary for proper sarcomere formation. *Dev. Biol.* 405, 1–9. doi: 10.1016/j.ydbio.2015.06.005
- Kaminski, H. J., Li, Z., Richmonds, C., Lin, F., and Medof, M. E. (2004). Complement regulators in extraocular muscle and experimental autoimmune myasthenia gravis. *Exp. Neurol.* 189, 333–342. doi: 10.1016/j.expneurol.2004.06.005
- Kusner, L. L., Young, A., Tjoe, S., Leahy, P., and Kaminski, H. J. (2010). Perimysial fibroblasts of extraocular muscle, as unique as the muscle fibers. *Invest. Ophthalmol. Vis. Sci.* 51, 192–200. doi: 10.1167/iovs.08-2857

- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4
- McLoon, L. K., and Wirtschafter, J. (2003). Activated satellite cells in extraocular muscles of normal adult monkeys and humans. *Invest. Ophthalmol. Vis. Sci.* 44, 1927–1932. doi: 10.1167/iov.02-0673
- Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16:1215. doi: 10.1093/nar/16.3.1215
- Mishra, A., and Macgregor, S. (2015). VEGAS2: Software for more flexible gene-based testing. *Twin Res. Hum. Genet.* 18, 86–91. doi: 10.1017/thg.2014.79
- Mishra, A., and MacGregor, S. (2017). A novel approach for pathway analysis of GWAS data highlights role of bmp signaling and muscle cell differentiation in colorectal cancer susceptibility. *Twin Res. Hum. Genet.* 20, 1–9. doi: 10.1017/thg.2016.100
- Mombaur, B., Lesosky, M. R., Liebenberg, L., Vreede, H., and Heckmann, J. M. (2015). Incidence of acetylcholine receptor-antibody-positive myasthenia gravis in south africa. *Muscle Nerve* 51, 533–537. doi: 10.1002/mus.24348
- Nakano, S., and Engel, A. G. (1993). Myasthenia gravis: quantitative immunocytochemical analysis of inflammatory cells and detection of complement membrane attack complex at the end-plate in 30 patients. *Neurology* 43, 1167–1172. doi: 10.1212/WNL.43.6.1167
- Nel, M., and Heckmann, J. M. (2018). “Epidemiology and Genetics of Myasthenia Gravis,” in *Myasthenia Gravis and Related Disorders*, ed. H. J. Kaminski (Cham: Springer International Publishing), 71–84. doi: 10.1007/978-3-319-73585-6_5
- Nel, M., Jalali Sefid Dashti, M., Gamielien, J., and Heckmann, J. M. (2017). Exome sequencing identifies targets in the treatment-resistant ophthalmoplegic subphenotype of myasthenia gravis. *Neuromuscul. Disord.* 27, 816–825. doi: 10.1016/j.nmd.2017.06.009
- Nel, M., Prince, S., and Heckmann, J. M. (2019). Profiling of patient-specific myocytes identifies altered gene expression in the ophthalmoplegic subphenotype of myasthenia gravis. *Orphanet J. Rare Dis.* 14:24. doi: 10.1186/s13023-019-1003-y
- Oosterhuis, H., and Bethlem, J. (1973). Neurogenic muscle involvement in myasthenia gravis. *J. Neurol. Neurosurg. Psychiatry* 36, 244–254. doi: 10.1136/jnnp.36.2.244
- Park, I., Han, C., Jin, S., Lee, B., Choi, H., Kwon, J. T., et al. (2011). Myosin regulatory light chains are required to maintain the stability of myosin II and cellular integrity. *Biochem. J.* 434, 171–180. doi: 10.1042/BJ20101473
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* [Preprint]. doi: 10.1101/201178
- Porter, J. D., Khanna, S., Kaminski, H. J., Rao, J. S., Merriam, A. P., Richmonds, C. R., et al. (2001). Extraocular muscle is defined by a fundamentally distinct gene expression profile. *PNAS* 98, 12062–12067. doi: 10.1073/pnas.211257298
- Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., et al. (2010). Strong Maternal khoisan contribution to the south african coloured population: a case of gender-biased admixture. *Am. J. Hum. Genet.* 86, 611–620. doi: 10.1016/j.ajhg.2010.02.014
- Reed, E., Nunez, S., Kulp, D., Qian, J., Reilly, M. P., and Foulkes, A. S. (2015). A guide to genome-wide association analysis and post-analytic interrogation. *Stat. Med.* 34, 3769–3792. doi: 10.1002/sim.6605
- Renton, A. E., Pliner, H. A., Provenzano, C., Evoli, A., Ricciardi, R., Nalls, M. A., et al. (2015). A genome-wide association study of myasthenia gravis. *JAMA Neurol.* 72:396. doi: 10.1001/jamaneurol.2014.4103
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Schöne, B., Bergmann, S., Lang, K., Wagner, I., Schmidt, A. H., Petersdorf, E. W., et al. (2018). Predicting an HLA-DPB1 expression marker based on standard DPB1 genotyping: linkage analysis of over 32,000 samples. *Hum. Immunol.* 79, 20–27. doi: 10.1016/j.humimm.2017.11.001
- Shieh, M., Chitnis, N., Clark, P., Johnson, F. B., Kamoun, M., and Monos, D. (2018). Computational assessment of miRNA binding to low and high expression HLA-DPB1 allelic sequences. *Hum. Immunol.* 80, 53–61. doi: 10.1016/j.humimm.2018.09.002
- Soltys, J., Gong, B., Kaminski, H. J., Zhou, Y., and Kusner, L. L. (2008). Extraocular muscle susceptibility to myasthenia gravis: unique immunological environment? *Ann. N. Y. Acad. Sci.* 1132, 220–224. doi: 10.1196/annals.1405.037
- Tang, H., Inoki, K., Lee, M., Wright, E., Khuong, A., Khuong, A., et al. (2014). mTORC1 promotes denervation-induced muscle atrophy through a mechanism involving the activation of FoxO and E3 ubiquitin ligases. *Sci. Signal.* 7:ra18. doi: 10.1126/scisignal.2004809
- Thomas, R., Thio, C. L., Apps, R., Qi, Y., Gao, X., Marti, D., et al. (2012). A novel variant marking hla-dp expression levels predicts recovery from hepatitis b virus infection. *J. Virol.* 86, 6979–6985. doi: 10.1128/JVI.00406-12
- Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* 3:731. doi: 10.21105/joss.00731
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From fastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* 43, 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43
- Yamazaki, T., Umemura, T., Joshita, S., Yoshizawa, K., Tanaka, E., and Ota, M. (2018). A cis-eQTL of HLA-DPB1 affects susceptibility to type 1 autoimmune hepatitis. *Sci. Rep.* 8:11924. doi: 10.1038/s41598-018-30406-9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Nel, Mulder, Europa and Heckmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.