Integration of Genomics Database and Bioinformatics to Identify GenomeWide Variants for Myasthenia Gravis Across Multiple Continents Dwiki Fitri1 , Lalu Muhammad Irham1*, Nanik Sulistyani1 , Muhammad Ma'ruf1 , Anisa Nova Puspitaningrum2 , Wirawan Adikusuma3 , Maulida Mazaya4 . Rockie Chong5 1Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 2Departement of Pharmacy, Telogorejo Semarang College of Health Sciences, Semarang, Indonesia 3Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram, Indonesia 4Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong, Indonesia 5Department of Chemistry and Biochemistry, University of California, Los Angeles, USA *Corresponding author: lalu.irham@pharm.uad.ac.id ABSTRACT Autoimmune disease is an immune response that damages the body's tissues, thereby disrupting the body's physiological functions. Myasthenia gravis represents one such condition characterized by muscle weakness due to impaired neuromuscular transmission. While it can affect anyone, it tends to be more prevalent among women aged 20-30 and men over 50. This disease, deemed a genetic disorder, typically emerges in old age when antibodies target receptors in the muscles. In this study, we sought to identify the genes that can affect myasthenia gravis by leveraging several databases, including the GWAS Catalog, HaploReg Version 4.2, GTEX portal, and ensemble. Specifically, our focus was on exploring genomic variants and the expression of the LTA and CTLA4 genes. Our findings reveal that two variants (rs2071591 and rs231770) impact LTA expression in both muscle and brain tissue, while affecting CTLA4 expression in testicular cell tissue. Subsequently, we assessed the allele frequency of these variants across regional populations, namely African, American, East Asian, European, and Southeast Asian. This study demonstrates that the LTA and CTLA4 genes have a higher frequency in African, East Asian, and European populations compared to American and Southeast Asian populations. Consequently, our finding suggests that the latter two populations might have relatively higher susceptibility to the autoimmune disease myasthenia gravis. Therefore, variations in these genes not only offer insights into disease susceptibility, diagnostic and prognostic biomarkers but also open avenues for identifying candidate drug targets through genomic-driven drug repurposing. Keywords: Myasthenia Gravis, Autoimmune, Bioinformatics, Genetic Variation. 1. Introduction Myasthenia gravis (MG) is a neuromuscular disease characterized by voluntary muscle weakness (Sanders, Wolfe, & Narayanaswami, 2017) (Gilhus & Verschuuren, 2015). This disease has different symptoms that vary in other patients depending on the degree of involvement of the striated muscles. The most common symptoms in patients with myasthenia gravis are ocular symptoms, which present as ptosis and diplopia. These symptoms usually occur later in the day, Journal Pre-proof and following activities such as watching TV or driving are more common. Excessive fatigue has been reported due to frequent exertion in patients with this disease. Myasthenia gravis is an autoimmune disease that connects nerves to muscles (Murai, 2014), produced by different antibodies against synaptic membrane proteins (Benatar et al., 2016). It usually accounts for more than 85% of cases and is caused by a type of antibody to the skeletal muscle acetylcholine receptor (AChR-Ab) (Berrih-Aknin, Frenkian-Cuvelier, & Eymard, 2014). However, components other than AChR, such as muscle-specific receptor tyrosine kinase or lipoprotein-associated protein 4 (LRP4), can also be targeted for autoimmune attacks (Mehling et al., 2011). Based on the mechanism of autoimmune disease and antibodies, molecular skeletal muscle invasiveness, thymus status, genetic characteristics, disease phenotype, and response to treatment, myasthenia gravis is divided into early and late ocular subtypes, seronegative, thymoma, LRP4. Diagnosis of the MG subtype influences treatment decisions and disease

prognosis (Kerty, Elsais, Argov, Evoli, & Gilhus, 2014). Approximately 50% of patients with ocular MG develop generalized myasthenia gravis (GMG) over a 2-year, which affects other muscles and is manifested by visual weakness and symptoms (Khan & Wang, 2020). According to a systematic population-based study, Car et al. (Carr, Cardwell, McCarron, & McConville, 2010) estimated the incidence and prevalence of MG to be 54 per million and 77.7 per million, respectively. However, significant changes have been reported in various studies. The incidence of this disease has been shown to range between 1.77 and 21.3 per million people and a prevalence of 15 to 179 million people (Carr et al., 2010). Many epidemiological studies, especially in Western Europe and Asia, report significant differences in the incidence and prevalence of MG. The incidence of myasthenia gravis ranges from 1.7 to 30 per million per year(Breiner et al., 2016). This disease has two age peaks: 40-40 years, mainly affecting women, and the other 80-60 years, which occurs equally in men and women (Benatar et al., 2016). In summary, MG is a neuromuscular disorder characterized by muscle weakness, with its manifestations varying among patients. Ocular symptoms, notably ptosis and diplopia, are common, often exacerbated by activities like watching TV or driving. MG results from autoimmune processes triggered by antibodies against synaptic membrane proteins, predominantly the acetylcholine receptor (AChR-Ab). The disease is categorized into subtypes, impacting treatment choices and prognosis. Epidemiologically, MG's incidence and prevalence vary across regions and age groups. Genetic variations play a significant role, with GWAS cataloging identifying relevant SNPs. This study aims to investigate the variants associated with MG through an approach based on bioinformatics, offering insights into disease susceptibility and progression. In addition, gene expression profile patterns and allele frequencies of genetic variant populations were assessed using various databases. The results will enable future studies to determine whether these variants may be associated with multiple risks of MG infection, as well as MG progression and disease susceptibility. 2. Method This study employed a bioinformatics-based approach to prioritize pathogenic variants potentially linked to MG. A detailed outline of the study design is visually represented in Figure Journal Pre-proof 1. To collect data on MG-associated variants, we leveraged the GWAS Catalog from the National Human Genome Research Institute (NHGRI) Database (http:// www.ebi.ac.uk/gwas), accessed on May 22, 2023. Employing the keyword “Myasthenia Gravis” (MG), we extracted information on variants associated with MG. We found a total 36 SNPs number of MG-associated variants. Next, subsequent analysis of MG-related SNPs was conducted using HaploReg (version 4.2) HaploReg is a tool designed to analyze non-coding genome annotations from published GWAS or new variants. It aids in understanding the functional outcomes of GWAS results, predicting potential causal variants, identifying involved cell types, and predicting candidate target genes (Ward & Kellis, 2016) (Ward & Kellis, 2012). In this study, the GWAS catalog inclusion criteria for SNPs were those with a p-value <10-8 (Lee et al., 2007). The investigation further entailed assessing associations between various genetic variants and gene expression profiles utilizing expression quantitative trait loci (QTL) available on the GTEx Portal database (http://www.gtexportal.org/home/) (Blauwendraat, 2022) accessed on May 22, 2023. Then, we confirmed the variant using the Ensembl Genome Browser (https://www.ensembl.org/index.html) (Ozaki et al., 2004) accessed on May 22, 2023. In addition, allele frequencies of the MG-associated variants were evaluated across diverse populations, encompassing African, American, East Asian, European, and Southeast Asian people. Journal Pre-proof Figure 1. Bioinformatics workflow for the identification of genetic variations associated with Myasthenia Gravis (MG) 3. Results This study utilized the GWAS

database to identify SNPs associated with MG. From the GWAS database, 36 SNPs associated with MG were identifieden with 26 unique SNPs associated with MG further identified after removing all SNP duplications (Table 1.) Based on the number of SNPs obtained, the SNPs were forwarded using HaploReg version 4.2, with a p-value <10-8 . Table 1. GWAS catalog result the duplicates removed are obtained from 26 SNPs with significance (p-value <10-8 ) SNP P-value rs3093958 4x10-42 rs9271375 2x10-19 rs4369774 6x10-19 Journal Pre-proof rs111945767 3x10-17 rs76815088 6x10-16 rs4409785 2x10-07 rs4574025 7x10-14 rs35274388 1x10-12 rs2476601 2x10-12 rs2071591 4x10-12 rs150881176 1x10-11 rs231770 9x10-11 rs4574025 4x10-07 rs9963862 4x10-07 rs12653117 5x10-07 rs6914704 2x10-06 rs4128527 4x10-06 rs4518467 4x10-06 rs2476601 7x10-06 rs9266277 7x10-10 rs6998967 9x10-10 rs35274388 1x10-09 rs4263037 1x10-08 rs73007767 4x10-08 rs2245569 6x10-08 rs9270986 6x10-08 Based on the data presented in Table 2, this study focused on two genomic variants of the same gene that qualified as biological risk SNPs in this MG study. Table 2. Varian risk allele which codes two genes SNP GENCODE P-value Allele Location rs2071591 LTA 4x10-12 Missense rs231770 CTLA4 9x10-11 Missense Journal Pre-proof Using an integrative bioinformatics approach, two variants with missense mutations rs2071591 and rs231770 encoding the LTA and CTLA4 genes were prioritized as MG biological risk SNPs. MG disease is characterized by muscle and tissue weakness that occurs when the immune system is impaired and produces antibodies that attack the tissues in the body (Benatar et al., 2016). It was also reported that Lymphotoxin α (LTA) is a cytokine secreted by lymphocytes and is a member of the Tumor Necrosis Factor (TNF) family. LTA gene variations can contribute to threshold brain excitability, the spread of neural hyperexcitability (Aurora & Welch, 2000). In Renton's (2015), the gene for cytotoxic T lymphocyte-associated protein (CTLA4) was previously suggested as a cause of myasthenia gravis susceptibility. The CTLA4 gene also multiplies when symptoms are present regardless of age, indicating that it is responsible for aberrant autoimmune responses that lead to neuromuscular junction dysfunction. CTLA4 45-kD immunoglobulin is expressed by activated T cells and has a significant sequence identity with CD28 (Renton et al., 2015). LTA gene expression in muscle and brain tissue. In atherosclerotic plaques, intimal cells, some spindle-shaped or have globular, vacuolated cytoplasm, show immunoreactivity for LTA and galectin-2. Binds to adjacent portions of antitender muscle (SMC) cells. Galectin-2 and LTA are expressed in human smooth muscle cells and macrophages affected by atherosclerotic lesions (Vergoossen, Keo, Mahfouz, & Huijbers, 2021). In Feroni's (2022), states that LTA genes encode cytokines that can modulate many inflammatory, immunological, and antiviral responses. It has been postulated that the inflammatory process modulated by LTA may contribute to the propagation of neural hyperexcitability by acting as an initiation and maintenance factor during migraine attacks (Ma'ruf et al., 2023). These results agree with a study on 439 Korean migraine patients genotyped for several LTA gene polymorphisms (Lee et al., 2007). Migraines occur because the blood vessels in the brain experience dilation or expansion, the main form of headache, characterized by debilitating headache attacks and symptoms of autonomic nervous system dysfunction (Olesen, 2018). Journal Pre-proof Figure 2. LTA gene expression associated with myasthenia gravis across several human tissues from the GTEx Portal. Expression of the CTLA4 gene in cell tissue in the testis. Our new analysis does not show the CTLA4 location in the previous GWAS (rs231770). After tracing using HaploReg4.2, it was found that neighboring missenses may show variations in alleles that impact the risk of myasthenia gravis in various populations. Although these loci still make biological sense, more extensive studies are needed to prove that they are related to each other (Blauwendraat, 2022). Another study in Vergoosen

2021 said that almost all MG-related genes were found in the testes and ectocervix. This study did not say a specific CTLA4 gene existed in testicular tissue. Still, the genes in question were RAPSN and CHRNA1 expression mostly limited to skeletal muscle, with some additional words in the tibial nerve and the testicular and pituitary glands, respectively (Yasumizu et al., 2022). Overall, the expression of MG-related genes is prominent in skeletal muscle and brain, but individual genes are also expressed in other tissues of the human body (Vergoossen et al., 2021) Journal Pre-proof . Figure 3. CTLA4 gene expression associated with myasthenia gravis across several human tissues from the GTEx Portal. Relationship between LTA and CTLA4 genes with eQTL from GTEx portal database. The GTEx Portal database aims to evaluate MG gene expression in various human tissues, including muscle tissue. Using the GWAS catalog database, we located 37 SNPs and identified genomic variation in LTA and CTLA4 gene expression. We found 26 SNPs with the highest p values from this analysis. After this investigation, two statistically significant SNPs were found and prioritized. Based on an extended SNP count analysis using HaploReg version 4.2, we prioritized the two SNPs at risk for MG because the functional annotations of the SNPs were 10-8 . According to (Yasumizu et al., 2022), the increase in CHRNA1 expression tend to be lower when compared with other neuromuscular antigens such as GABRA5 and RYR3. Although a moderate rise in CHRNA1 expression seems sufficient to cause severe symptoms in MG, it is related to the availability of acetylcholine receptor antibodies and their crucial biological role. These findings provide valuable clues for understanding the pathogenesis of various autoimmune neurological diseases. The following Table 3 shows the results of genetic variation from MG. Table 3. e-QTL's result for the Myasthenia gravis from GTEX portal database SNP GENCODE ID (ENSG00000-) Gene symbol P-value NES Tissue Actions rs2071591 226979.8 LTA 0.000065 0.29 Brain - Cerebellum GG>GA>AA Journal Pre-proof 226979.8 LTA 0.00024 0.14 Muscle - Skeletal GG>GA>AA rs231770 163599.14 CTLA4 0.0000016 -0.28 Testis CC>CT>TT Myasthenia gravis (MG) candidate variant allele frequencies across continents After identifying candidate variants related to LTA and CTLA4 gene expression, allele frequencies have been determined across populations of all continents as shown in (Table 4). Allele frequencies for the four variants were evaluated in different people, including populations of 1195 individuals (Africa), 567 individuals (America), 1112 individuals (East Asia), 693 individuals (Europe), and 554 individuals (South East Asia). Using the Ensemble Genome Browser, we obtained allele frequencies in Africa, America, East Asia, Europe, and Southeast Asia (http://www.ensembl.org). Allele frequencies across populations differ for each LTA and CTLA4 gene variant. Table 4 and Figure 4. shows the gene expression level at a higher frequency of the related allele (T) rs231770 than the corresponding allele (A) rs2071591. In the associated alleles (T) rs231770 and (A) rs2071591, the African and East Asian populations are much higher than the American, Southeast Asian, and European Asian people. Figure 4. Summary of allele frequency analysis of LTA and CTLA4 gene expression in Africa, America, East Asia, Europe, and Southeast Asia. Journal Pre-proof Table 4. Allele frequencies for SNPs examined in this study SNP Position (hg38) Gene Pvalue Location Allele Allele Frequency (N) Ref Eff* AFR AMR EAS EUR SAS rs2071591 Chr6 31548022 LTA 4x10-12 Missense G A 0.512 (677) 0.336 (233) 0.469 (473) 0.310 (312) 0.261 (255) rs231770 Chr2 203864430 CTLA 4 9x10-11 Missense C T 0.392 (518) 0.481 (334) 0.634 (639) 0.379 (381) 0.306 (299) The higher frequency of the related allele is (T) rs231770 compared to the other related allele, namely (A) rs2071591. In the associated alleles (T) rs231770 and (A) rs2071591, the African and East Asian populations are much higher than the American, Southeast Asian, and European Asian people. The

rs2071591 allele indicated a differential variant prevalence contribution for LTA gene expression, while the rs231770 allele indicated a differential variant prevalence contribution for CTLA4 gene expression. Allele frequencies in all populations differ for each SNP, as shown in Figure 4. It is generally known that the A and T allele frequencies for rs2071591 and rs231770 also appear to have a higher frequency in Europe with alleles rs2071591 (47%), rs231770 (63%), East Asia with alleles rs2071591 (47%), rs231770 (63%), compared to America with alleles rs2071591 (34%), rs231770 (48%), Africa with alleles rs2071591 (51%), rs231770 (39%), Asia southeast with alleles rs2071591 (26%), rs231770 (31%). Identification of gene variations that are unique and pathogenic for a disease is very interesting for clinical research and validation. Identification of these variants not only provide clues to disease susceptibility or as a diagnostic and prognostic biomarker (Irham, Adikusuma, & Perwitasari, 2022) but also can be used to discover candidate drug targets or what is known as drug repurposing (genomic driven drug repurposing) (Irham et al., 2020) (Afief et al., 2022). The authors hope that the discovery of the candidate gene variations found can be validated in clinical settings and can become a diagnostic and prognostic biomarker for myasthenia gravis disease. In conclusion, a bioinformatics-based approach revealed pathogenic variants potentially associated with MG. We propose that this variant may be used for further studies to identify MG and prognosis diagnostic biomarkers. However, we acknowledge that there are limitations to the bioinformatics-based approaches used to investigate genetic variants associated with MG. Notably, not all variants necessarily correspond to the identifiable gene (i.e., non-coding variants). ven when genes or genetic variants are present, they might not be suitable drug targets. Therefore, we recommend further clinical validation to corroborate our findings and gain deeper insight into the etiology and functional implications of MG disease. 4. Conclusion In this study, we utilized a state-of-the-art bioinformatics approach to analyze genomic databases revealing distinct gene expressions of the LTA and CTLA4 genes across muscle, brain, and testicular cell tissues for MG disease. Prominent gene variants include rs2071591, expressed in muscle and brain tissue, and rs231770, prevalent in testicular cell tissue. Notably, these variants Journal Pre-proof exhibit an overall higher frequency in Europe with alleles rs2071591 (47%), rs231770 (63%), East Asia with alleles rs2071591 (47%), rs231770 (63%), in comparison to America with alleles rs2071591 (34%), rs231770 (48%), Africa with alleles rs2071591 (51%), rs231770 (39%), and Southeast Asia with alleles rs2071591 (26%), rs231770 (31%). We recommend that the discovered candidate gene variations be validated clinically, which could serve as diagnostic or prognostic biomarkers for MG.