Integration of genomic variants and bioinformatic-based approach to drive drug repurposing for multiple sclerosis

Arief Rahman Afief a,1, Lalu Muhammad Irham a,3, Wirawan Adikusuma b,*, Dyah Aryani Perwitasari a, Ageng Brahmadhi c, Rockie Chong d

a Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
b Department of Pharmacy, University of Muhammadiyah Mataram, Mataram, Indonesia
c Faculty of Medicine, Universitas Muhammadiyah Purwokerto, Purwokerto, Central Java, Indonesia
d Department of Chemistry and Biochemistry, University of California, Los Angeles, USA

ARTICLE INFO

Keywords:
Autoimmune disease
Bioinformatics
Drug repurposing
Genomic variants
Multiple sclerosis

ABSTRACT

Multiple sclerosis (MS) is a chronic autoimmune disease in the central nervous system (CNS) marked by inflammation, demyelination, and axonal loss. Currently available MS medication is limited, thereby calling for a strategy to accelerate new drug discovery. One of the strategies to discover new drugs is to utilize old drugs for new indications, an approach known as drug repurposing. Herein, we first identified 421 MS-associated SNPs from the Genome-Wide Association Study (GWAS) catalog (p-value < 5 × 10−8), and a total of 427 risk genes associated with MS using HaploReg version 4.1 under the criterion $r^2 > 0.8$. MS risk genes were then prioritized using bioinformatics analysis to identify biological MS risk genes. The prioritization was performed based on six defined categories of functional annotations, namely missense mutation, cis-expression quantitative trait locus (cis-eQTL), molecular pathway analysis, protein-protein interaction (PPI), genes overlap with knockout mouse phenotype, and primary immunodeficiency (PID). A total of 144 biological MS risk genes were found and mapped into 194 genes within an expanded PPI network. According to the DrugBank and the Therapeutic Target Database, 27 genes within the list targeted by 68 new candidate drugs were identified. Importantly, the power of our approach is confirmed with the identification of a known approved drug (dimethyl fumarate) for MS. Based on additional data from ClinicalTrials.gov, eight drugs targeting eight distinct genes are prioritized with clinical evidence for MS disease treatment. Notably, CD80 and CD86 pathways are promising targets for MS drug repurposing. Using in silico drug repurposing, we identified belatacept as a promising MS drug candidate. Overall, this study emphasized the integration of functional genomic variants and bioinformatic-based approach that reveal important biological insights for MS and drive drug repurposing efforts for the treatment of this devastating disease.

1. Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease in the central nervous system (CNS) marked by inflammation, demyelination, and axonal loss. The onset of MS usually occurs between 20 and 40 years of age and more predominantly in women [1]. MS also causes a series of other heterogeneous symptoms due to varying involvements of the motor, sensor, visual, and autonomous systems. It is characterized by optic neuritis (optic nerve inflammation), Uhthoff’s phenomenon (temporary fluctuation and worsened MS symptoms with increased body temperature), and Lhermitte’s sign (abnormal electrical-shock-like sensation over the spinal cord or body parts during neck flexion) [2], and tends to develop in genetically susceptible individuals who are exposed to a diversity of triggering environmental
factors (e.g., Epstein-Barr virus, tobacco use, and vitamin D deficiency) [3]. The genes involved in MS have long been sought after. A number of approaches to this problem have been applied with varying degrees of success. The candidate gene approach has been in use over several decades, where potentially MS-associated genes are selected based on autoimmune MS prognosis, involving class I and II immune-response-gene-controlling human leukocyte antigen (HLA) [4].

Treatments for MS have been divided into three categories: 1) acute relapse management; 2) disease-modifying therapies; and 3) symptomatic treatments [2]. One MS treatment available and approved is dimethyl fumarate (Tecfidera) [2,5]. So far, these medications can help people with MS that have fewer and less severe relapses. However, the problem is still arising from those medications, including resistance and toxicity [6]. Under such circumstances, drug repurposing emerges as one of the solutions to identify new candidate drugs for MS disease. In addition, further investigations such as clinical validation and in vivo experiments are needed to accelerate new discoveries for the treatment of MS disease, which aim to maximize the likelihood of success during pre-clinical development and validation [7].

The concept of drug repurposing is to find new indications for existing drugs that are already available on the market [8]. The drug repurposing approach has several advantages compared to the traditional ones, such as time and cost-effectiveness [9], safety profile (drugs have previously passed clinical trials), dosage, and the toxicity of existing drugs has been vetted [10]. Genome-wide association study (GWAS) can potentially be leveraged for precision drug repurposing by applying functional annotation [11]. Several studies were applied to the risk variants from GWAS, and were able to prioritize the biological risk genes based on the functional annotations to drive drug repurposing for various diseases, including chronic hepatitis B [12], atopic dermatitis [13], asthma [14], colorectal cancer [15] and rheumatoid arthritis [16]. Of interest to this study, GWAS has revolutionized MS genetic analyses, including the MS variants. These variants have consistently implicated genes associated with immunological processes, mostly lie in regulatory rather than coding areas, and are often associated with other autoimmune diseases [17]. This study aimed to implement bioinformatics strategies to identify biological MS candidate genes through an integrated gene network. Six functional annotations (missense mutation, cis-expression quantitative trait locus (cis-eQTL), molecular pathway analysis, protein-protein interaction (PPI), overlap knockout mouse phenotype, and primary immunodeficiency (PIDD)) were used to find biological MS risk genes. Finally, we overlapped the biological MS risk genes with the drug databases and prioritized the candidate drugs to be repurposed for MS disease.

2. Methods

A detailed description of the study design of drug repurposing utilizing the genomic information for MS is provided in Fig. 1. MS-associated single nucleotide polymorphisms (SNPs) were obtained from the GWAS catalog under the criterion p-value $< 5 \times 10^{-8}$ and expanded using HaploReg (v4.1) based on the criterion of $r^2 \geq 0.8$ in Asian (ASN) populations. The data was retrieved from the 1000 Genome Project Phase I data [18,19]. Genes matching MS-associated SNPs are denoted as “MS-associated genes”. Then, genomic data were prioritized based on six functional annotation criteria. Every functional annotation is assigned a score of 1, and genes with a score $\geq 2$ are defined as “biological MS risk genes”. Biological MS risk genes were used in advanced analysis using the STRING database to extend the list of candidate genes as drug-target genes. This study mapped an approved expanded list of drug-target genes in the DrugBank and the Therapeutic Target Database. The drug-target genes were checked with ClinicalTrials.gov to determine the clinical status.

2.1. Functional annotations of MS risk genes

Functional annotation describes a gene’s biological identity by compiling the relevant biological information for a particular gene. Herein, six categories of functional annotations were used to build an assessment system representing the candidate genes most likely to be MS targets. The first category of annotation was missense or nonsense mutation according to HaploReg v4.1, which contains functional consequence annotations of the SNP database (db). HaploReg v4.1 also connected genetic variants to cis-expression quantitative trait loci (cis-eQTLs) [16]. If a gene has an MS risk SNP with cis-eQTL effect throughout the blood, the gene was assigned one point. Then, to gain an understanding of the relationship between a mutant gene and the phenotype, WebGestalt 2019 was used for functional enrichment analysis. The data source was the Mammalian Phenotype (MP) Ontology, which contains information on the mouse and other mammalian phenotypes [15]. The genes from human Ensembl ID were converted into mouse Ensembl ID using BioMart. Clusters of genes with FDR $<0.05$ in the enrichment analysis were considered significant. Specifically, the gene ontology (GO) biological process categories were analyzed for this stage. The result significance was set at FDR $<0.05$. Enrichment analysis was performed on molecular pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG). Genes enriched on the KEGG pathway (FDR $<0.05$) were assigned a score of 1. Primary immunodeficiency (PIDD) was the last annotation criterion. It refers to inborn immunity diseases that are genetic disorders associated with increased severity [13,15]. Data enrichment analysis was performed using the hypergeometric test; p $< 0.05$ was used in this stage as the significance criterion.

![Fig. 1. Scheme of drug repurposing using genomic database for multiple sclerosis (MS).](image-url)
It is important to note that each functional annotation is assigned a score of 1, and genes with a score ≥2 are defined as “biological MS risk genes”. Biological MS risk genes were used in advanced analysis from the STRING database.

2.2. STRING database

The use of the STRING database (http://string-db.org) aimed to integrate functional interactions related to protein expressions by inputting and regulating data associated with the predicted protein-protein interactions [20,21]. The majority of protein networks in various diseases can be the targets of the diseases [22]. The biological MS risk genes were expanded using the STRING database to gain more candidate drug targets. This step emphasized that the genomic information of MS has given insight into the biological risk genes for MS.

2.3. Validation and drug discovery

The drug target gene’s candidate was overlapped with drug databases such as the DrugBank (https://www.drugbank.ca/) and Therapeutic Target Database (http://db.idrblab.net/ttd/) to find the candidate drug to be repurposed for MS disease. DrugBank and Therapeutic Target Database are databases widely used to identify the drug target precisely. It also contributes to driving drug repurposing for various diseases [23]. Drug-target genes were used to analyze the database based on several criteria, such as drugs with pharmacological activity, effectiveness in humans, and approved annotations in clinical trials or drug experiments [24]. Furthermore, the identified drugs were reviewed with ClinicalTrials.gov (https://clinicaltrials.gov) to identify clinical examinations for MS or other diseases. The data visualization of drugs under clinical investigation for MS were built by using R (Chord diagram) with the circlize package (RStudio 4.0.3 program).

3. Results

The susceptibility of various MS genomic variants was retrieved from the genomic database. A variety of genomic databases can be used including GWAS databases. GWAS not only provides information on the susceptibility of diseases but also provides information on the biological insight of diseases. In this study, 420 MS-associated SNPs were obtained from the GWAS database (Table S1). The expansion was then performed using HaploReg v4.1 under the criterion $r^2 > 0.8$, resulting in 427 MS-associated risk genes (Table S2).

3.1. Functional annotations of MS risk genes

Six categories of biological functional annotations were applied to prioritize biological MS risk genes. One point was assigned to each category of functional annotation. The assessment of each of the 427 candidate genes under the six criteria was as follows: (1) genes with missense mutation MS risk variants (n = 29); (2) cis-eQTL genes (n = 134); (3) genes in knockout mouse phenotypes (n = 101); (4) genes with the KEGG pathways (n = 93); and (6) number of genes overlapped with the PID (n = 13) (Fig. 2 and Fig. 3) (Table S2). Biological scoring was conducted after data collection. There were 173 genes with a score of 0, 110 genes with a score of 1, 67 genes with a score of 2, 47 genes with a score of 3, 7 genes with a score of 5, and 2 genes with a score of 6. A total of 144 genes had a score ≥2 (Fig. 4). We found that interferon-gamma receptor 2 (IFNGR2) and interleukin 7 receptor (IL7R) were the top two biological MS risk genes, each with a score of 6.

3.2. Expansion of biological MS risk genes list

The STRING database was utilized to expand the investigation of biological MS risk genes. The expansion is based on the rationale that, the more biological MS risk genes we find, the more candidate drug targets for MS drug repurposing can be identified. We successfully obtained 194 genes on the list (Table S3). These genes were included on the list of new candidate drug genes for further analysis.
3.3. MS drug targets finding

Finally, the drug target genes were prioritized based on the network analysis and the drug databases. Herein, we obtained 2904 gene pairs from the PPI network (Table S4) and 27 genes targeted by 68 new candidate drugs based on DrugBank and Therapeutic Target Database (Table S5). We found one drug, dimethyl fumarate that has been clinically approved for MS treatment (Fig. 5). This drug is an effective medicinal option, administered twice a day in MS medication [25, 26]. This study emphasized that the biological functional annotation we applied can be validated through a known drug used in the clinic for MS disease. This study also found eight drug-target genes bound to 8 drugs approved for other diseases and under clinical investigation for MS, including human immunoglobulin G, antithymocyte immunoglobulin (rabbit), liothyronine, abatacept, topiramate, and phenytoin (Table S6). These drugs can potentially repurpose MS medication (Fig. 6). An example of a drug repurposed for MS is abatacept, which is approved for rheumatoid arthritis, targeting CD80 and CD86 gene pathways. This drug is currently under clinical investigation for MS in a phase II trial (NCT01116427) and has a considerable potential to be used for MS [27]. Thus, we would like to emphasize that integrating genomic variants and gene networking can potentially guide the drug repurposing for MS disease.

4. Discussion

The focus of our present work is to narrow down candidate drugs for a debilitating disease, multiple sclerosis, through leveraging large bioinformatics datasets together with human genetics data. More specifically, we use genomic data and genetic mapping to guide drug repurposing for MS. In particular, this study focused on using new candidate drugs for MS by prioritizing candidate genes derived and identified from the GWAS database. Six categories of functional annotations were used to build an assessment system, in order to prioritize the MS risk genes as leads for new candidate drugs. We hypothesized that the broad strategy of genetic variant prioritization, using the functional annotations described in this study, would enable us to translate the risk genes to meaningful, actionable insights on MS. According to our analyses, we ensure the sensitivity of our study results by setting the threshold of a biological score ≥ 2 to screen a much higher number of genes as biological MS risk genes, and thereby as candidates for MS drug targets.

In this research, 27 drug-target gene products were found to bind to 68 drugs. In addition, 8 drug-target gene products were found to bind to 8 drugs, of which some were under clinical testing for MS, namely human immunoglobulin G (NCT00220779), anti-thymocyte immunoglobulin (rabbit) (NCT03342638), liothyronine (NCT02760056), abatacept (NCT01116427), topiramate (NCT00217295), and phenytoin (NCT01451593). From the data collected, one available drug approved for MS is dimethyl fumarate (Fig. 5). Dimethyl fumarate, also known as BG-12, was licensed as first-line therapy and oral therapy for MS in 2013. It is also known as neuroprotective and immunomodulatory [28, 29]. The mechanism of action of dimethyl fumarate is to react with cysteine residues from KEAP1 (Kelch-like ECH-associated protein 1), which causes KEAP1 to be dissociated from the nuclear factor (erythroid-derived 2)-like 2 (Nrf2) pathway toward Nrf2 nuclear translocation. Nrf2 then binds antioxidant response element (ARE) and drives antioxidant target gene expression toward neuronal protection, reduces astrocyte activation, and prolongs cell life [30].

We identified eight promising targets overlapping with drugs that could potentially be repurposed to treat MS. These include C3, CD4, CD86, THRA, CSF2RB, CD80, IFNGR2, and CACNA1S. Among them, we highly proposed CD80 and CD86 as potential targets for MS, since these
targets are closely related to \textit{IL7R} as biological MS risk genes with high functional annotation scores (Fig. 6). The \textit{CD80}/\textit{CD86} pathway is essential for controlling T cell activation and preserving immunological tolerance to self-antigens. The findings of functional and genome-wide investigations demonstrate that genes encode molecules that fit in. This pathway may increase the likelihood of developing autoimmune illnesses and may be viewed as a potential MS candidate gene \cite{31–33}. In addition, we identified \textit{CD80}- and \textit{CD86}-targeting drugs, including anti-thymocyte immunoglobulin (rabbit), abatacept, and belatacept. Among these drugs, in fact, anti-thymocyte immunoglobulin (rabbit) (NCT03342638) and abatacept (NCT01116427) are currently under clinical investigation for MS. Therefore, from this perspective, targeting \textit{CD80} and \textit{CD86} might become novel therapeutic options for MS therapy. Further clinical evidence generation would be needed to validate these targets.

Combining results from GWAS and bioinformatic, gene-level annotation of human genetic variants is a powerful approach to identify candidate new drugs for MS. However, it is important to consider that this approach is not without limitations, such as not all the identified target genes can either be targeted and/or demonstrate the pharmacological activity with the desired profile for use in the clinic. The genes identified in this manner would therefore miss the theoretical druggability for the particular disease. Further investigation is thus required to determine the effect of the candidate drugs on the clinic. Hence, we suggest that the use of current findings with subsequent functional studies to ascertain the role of drug target genes discovered in this study. Preclinical and clinical validation is necessary and important to ensure whether our drug candidates produced the desired interaction (intended from the study), any undesired side effects, or suboptimal effects.

5. Conclusions

Our study utilizes MS functional genomic variants to open up additional avenues for the repurposing of existing drugs from other therapeutic classes and disease areas. Herein, we identify \textit{CD80} and \textit{CD86} as potential targets for MS treatment. The involvement of these genes with MS might be highly significant, thereby requiring further examination. Through targeting \textit{CD80} and \textit{CD86}, belatacept could be a promising therapy option for MS therapy. However, additional studies from animal models and clinical trials are needed to confirm the biological mechanisms of the drug for new diseases. Overall, we combined the drug repurposing approach with integrated bioinformatics methodology to identify drugs with new indications for MS. Finally, this study emphasizes the vast potential of utilizing functional genomic variants as a basis to drive drug repurposing for MS disease.

Fig. 6. Relationship between biological MS genes, and drugs approved for other indications and under clinical investigation for MS.
References

1. A. Vidal-Jordana, X. Montalban, Multiple sclerosis: epidemiological, clinical, and therapeutic aspects, neuroimaging clin, N. Am. 27 (2017) 195–204, https://doi.org/10.1016/j.neuci.2016.12.001.
2. A. Desai, J. Chataway, Multiple sclerosis, a treatable disease, Clin Med. 16 (2016) S3–S9, https://doi.org/10.7861/clinmed.16-6.s3.
3. A.E. Handel, G. Giovannoni, G.C. Ebers, S.V. Ramagopalan, Environmental factors and the risk of secondary progressive multiple sclerosis, PLoS One 10 (2015), https://doi.org/10.1371/journal.pone.0125071.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://doi.org/10.1016/j.bbrep.2021.101337.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

We used the databases in the current study which were retrieved from GWAS (https://www.ebi.ac.uk/gwas/); ClinicalTrials.gov (https://clinicaltrials.gov/); Drug bank (https://www.drugbank.ca/); HaploReg Version 4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php); STRING database (https://string-db.org/); Therapeutic Targets Database (http://dtd.db.idrblab.net/ttd/).

These authors declare no conflict of interest.

Data integration of gene networking and genomic information, front, Immunol. 12 (2022), https://doi.org/10.3390/immunol.12.2472.

11. A. Taniguchi, R. Yamada, M. Kubo, J.S. Liu, S.C. Bae, J. Worthington, L. Padyukov, L. Klotz, Dimethyl fumarate treatment alters circulating T helper cell subsets in multiple sclerosis, Front Immunol. 12 (2021), https://doi.org/10.3389/fimmu.2021.753929.

12. H. Ruffner, A. Bauer, T. Bouwmeester, Human protein-protein interaction networks and the value for drug discovery, Drug Discov Today. Off. 12 (2007) 709–716, https://doi.org/10.1016/j.drudis.2007.07.011.

13. Y. Masoudi-Sobhanzadeh, Y. Omidi, M. Amanlou, A. Masoudi-Nejad, Drug databases and their contributions to drug repurposing, Genomics 112 (2020) 1087–1095, https://doi.org/10.1016/j.ygeno.2019.06.021.

14. S.D. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayevera, N. Assempour, I. Jyikkaran, Y. Liu, A. Maclejewski, N. Gale, A. Wilson, L. Aon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 47 (2019) D607–D613, https://doi.org/10.1093/nar/gky1131.

15. M. Li, H. Chen, P. Yin, J. Song, F. Zhang, X. Tang, X. Fan, C. Xu, Y. Wang, Y. Xue, B. Han, H. Wang, G. Li, D. Zhong, Identification and clinical validation of key extracellular proteins as the potential biomarkers in relapsing-remitting multiple sclerosis, Front Immunol. 12 (2021), https://doi.org/10.3389/fimmu.2021.753929.

16. N. Santri, Y.V.A. Phiri, R. Chong, The use of genomic variants to drive drug repurposing, Nat. Rev. Genet. 22 (2021) 658–677, https://doi.org/10.1038/s41576-021-00387-z.

17. R.A. Linker, D.H. Lee, S. Ryan, A.M. Van Dam, R. Conrad, P. Bista, W. Zeng, R.A. Afief et al.

18. S.C. Gupta, B. Sung, B.B. Aggarwal, Cancer drug discovery by repurposing: teaching new tricks to old dogs, Trends Pharmacol. Sci. 34 (2013) 508–617, https://doi.org/10.1016/j.tips.2013.06.005.

19. M. Zhao, C.C. Yang, Drug repositioning to accelerate drug development using social media data: computational study on Parkinson disease, J. Med. Internet Res. 20 (2018), https://doi.org/10.2196/jmir.9646.

20. T.T. Ashburn, K.B. Thor, Drug repurposing: identifying and developing new uses for existing drugs, Nat. Rev. Drug Discov. 3 (2004) 673–683, https://doi.org/10.1038/nrd1460.

21. W.B. Reay, M.I. Caix, Advancing the use of genome-wide association studies for drug repurposing, Nat. Rev. Genet. 22 (2021) 658–671, https://doi.org/10.1038/s41576-021-00387-z.

22. L.M. Irham, W. Adikusuma, D.A. Perwitasari, H. Dania, R. Maliza, I.N. Faridah, I. N. Santri, Y.V.A. Phiri, R. Chong, The use of genomic variants to drive drug repurposing for chronic hepatitis B, Biochem. Biophys. Reports. 31 (2022) 101307, https://doi.org/10.1016/j.bbrep.2022.101307.

23. W. Adikusuma, L.M. Irham, W.H. Chou, H.S.C. Wong, E. Mugiyanto, J. Ting, D. A. Perwitasari, W.F. Chang, C.W. Chang, Drug repurposing for atopic dermatitis by
[32] M.T. Boylan, A.D. Crockard, G.V. McDonnell, M.A. Armstrong, S.A. Hawkins, CD80 (B7-1) and CD86 (B7-2) expression in multiple sclerosis patients: clinical subtype specific variation in peripheral monocytes and B cells and lack of modulation by high dose methylprednisolone, J. Neurol. Sci. 167 (1999) 79–89, https://doi.org/10.1016/S0022-510X(99)00152-X.

[33] M. Wagner, M. Sobczyński, L. Karabon, M. Bilinska, A. Pokryszko-Dragan, E. Pawlak-Adamska, M. Cyrl, P. Kuśnierscyk, M. Jasek, Polymorphisms in CD28, CTLA-4, CD80 and CD86 genes may influence the risk of multiple sclerosis and its age of onset, J. Neuroimmunol. 288 (2015) 79–86, https://doi.org/10.1016/j.jneuroim.2015.09.004.