

Research Article

Identification of Candidate Genes Associated with Charcot-Marie-Tooth Disease by Network and Pathway Analysis

Min Zhong,¹ Qing Luo,¹ Ting Ye,¹ XiDan Zhu,¹ Xiu Chen ,² and JinBo Liu ¹

¹Department of Laboratory Medicine, The Affiliated Hospital of Southwest Medical University, 25 Taiping Street, Luzhou, 646000 Sichuan, China

²Department of Neurology, The Affiliated Hospital of Southwest Medical University, 25 Taiping Street, Luzhou, 646000 Sichuan, China

Correspondence should be addressed to JinBo Liu; liulab2019@163.com

Received 2 May 2020; Revised 21 July 2020; Accepted 12 August 2020; Published 24 September 2020

Academic Editor: Marco Fichera

Copyright © 2020 Min Zhong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Charcot-Marie-Tooth Disease (CMT) is the most common clinical genetic disease of the peripheral nervous system. Although many studies have focused on elucidating the pathogenesis of CMT, few focuses on achieving a systematic analysis of biology to decode the underlying pathological molecular mechanisms and the mechanism of its disease remains to be elucidated. So our study may provide further useful insights into the molecular mechanisms of CMT based on a systematic bioinformatics analysis. In the current study, by reviewing the literatures deposited in PUBMED, we identified 100 genes genetically related to CMT. Then, the functional features of the CMT-related genes were examined by R software and KOBAS, and the selected biological process crosstalk was visualized with the software Cytoscape. Moreover, CMT specific molecular network analysis was conducted by the Molecular Complex Detection (MCODE) Algorithm. The biological function enrichment analysis suggested that myelin sheath, axon, peripheral nervous system, mitochondrial function, various metabolic processes, and autophagy played important roles in CMT development. Aminoacyl-tRNA biosynthesis, metabolic pathways, and vasopressin-regulated water reabsorption were significantly enriched in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway network, suggesting that these pathways may play key roles in CMT occurrence and development. According to the crosstalk, the biological processes could be roughly divided into a correlative module and two separate modules. MCODE clusters showed that in top 3 clusters, 13 of CMT-related genes were included in the network and 30 candidate genes were discovered which might be potentially related to CMT. The study may help to update the new understanding of the pathogenesis of CMT and expand the potential genes of CMT for further exploration.

1. Introduction

Charcot-Marie-Tooth Disease (CMT), also known as hereditary motor sensory neuropathy (HMSN), was first reported by French neurologists Charcot and Marie and British neurologist Tooth in 1886 [1, 2]. It is the most common clinical single-gene genetic disease of the peripheral nervous system with high clinical heterogeneity and genetic heterogeneity, with a prevalence of about 1/2500 [3, 4]. Although the disease progresses slowly in most patients, mild to moderate functional impairment does not affect life expectancy. However, there are currently no treatments to reverse the course of

CMT, which still limits the ability of patients to move independently, reduces the quality of life, and increases the risk of disability.

Like many other degenerative disorders, hereditary peripheral neuropathies have been difficult to treat. There are currently no effectual pharmacologic treatments for CMT, limiting historic treatment to supportive care. In nearly a decade of research, ascorbic acid, progesterone antagonists, and subcutaneous neurotrophin-3 (NT3) injections have shown initial success in animal models of CMT 1A (the most common subtype of CMT) but have failed to translate any effect in humans [5–8]. Considering the

difficulties of treatment, it is imperative to find out the molecular pathogenesis of CMT for the purpose of strategising potential future therapies.

The advent of next-generation sequencing (NGS) has expanded and accelerated the analysis of various diseases at the level of genome, especially in heterogeneous disorder groups such as CMT [9, 10]. Currently, due to the development of NGS, >100 genetic mutations have been found to cause or promote the clinical manifestations of CMT [11, 12]. They have been related to a variety of molecular pathological mechanisms, including either protein synthesis and posttranslational processing (dysfunction of mRNA processing, abnormal endosomal sorting and signaling, aberrant proteasome/protein aggregation, and myelin assembly abnormalities), dysfunction of ion channels (channelopathies), intracellular transportation (axonal transport/cytoskeletal abnormalities), or mitochondrial dysfunction [13, 14]. Based on the understanding of the genetic basis of hereditary neuropathy, different cell and animal models have been established to decode the molecular mechanism of CMT and provide treatment strategies. However, there is no causal treatment for hereditary neuropathy so far. In view of the increasing complexity of the genetics of neuropathies, this paper is aimed at combining bioinformatics with new directions [15].

Although many studies have focused on elucidating the pathogenesis of CMT, few focuses on achieving a systematic analysis of biology to decode the underlying pathological molecular mechanisms. In this study, we firstly made a comprehensive selection of genes genetically related to CMT. We then performed a functional enrichment analysis to identify important biological topics within these genetic factors. In order to further explore the pathogenesis of CMT, we constructed a biological process network to explore possible crosstalk among the significant biological processes. Furthermore, we made a PPI network of these CMT-related genes using MCODE [16, 17]. This study may provide further useful insights into the molecular mechanisms of CMT based on a systematic bioinformatics analysis and find causative mutations in the heterogeneous group of disorders (hereditary neuropathies) in the era of NGS.

2. Materials and Methods

2.1. Identification of CMT-Related Genes. Candidate genes associated with CMT were collected by retrieving the human genetic association studies deposited in PUBMED (<http://www.ncbi.nlm.nih.gov/pubmed/>). Referring to published studies [17–19], we searched for reports related to CMT with the term (Charcot-Marie-Tooth Disease [MeSH]) and (polymorphism [MeSH] or genotype [MeSH] or alleles [MeSH]) not (neoplasms [MeSH]). Up to 10 November 2019, a total of 674 publications were retrieved for the disorder. Then, we reviewed the abstracts of all above publications, collecting the genetic association studies of CMT. We narrowed our selection by concentrating on the selected publications reporting a considerable association of one or more genes with CMT. Instead, for the purpose of reducing the number of false-positive finding, we excluded studies reporting nega-

tive or insignificant associations, although some genes analyzed in these studies may be associated with CMT. For the selected publications, we reviewed the full texts to assure that the conclusion was consistent with its contents. Thus, we screened out the genes reported to be significantly related to CMT. No ethical approval was required as it did not involve human or animals.

2.2. Functional Enrichment Analysis of CMT-Related Genes.

The functional features of the CMT-related genes were examined by R software [20, 21] and KOBAS 3.0 [22, 23]. To understand the CMT-related genes underlying biological processes, Gene Ontology analyses (GO; <http://geneontology.org>) were conducted using R software (version 3.6.2). GO enrichment results were visualized using the R clusterProfiler package (version 3.14.3; 10.18129/B9.bioc.clusterProfiler). GO terms were selected with a false discovery rate (FDR) < 0.05. Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg>) pathway analysis of the candidate genes was performed using the KOBAS online analysis database (available online: <http://kobas.cbi.pku.edu.cn/>). On the whole, the genes with symbols and/or corresponding NCBI Entrez Gene IDs were uploaded to the server and compared with the genes contained in each canonical pathway according to the KEGG pathway database. All the pathways with one or more genes overlapping with the candidate genes were extracted, and a *P* value was assigned to each pathway through Fisher's exact test to indicate the significance of the overlap between the pathway and the input genes. Then, the pathways with FDR value less than 0.05 were considered to be significantly enriched.

2.3. Biological Process Crosstalk Analysis. We further performed biological process crosstalk analysis to explore the interactions among significantly enriched biological processes. To evaluate the overlap between any chosen pairs of biological processes, two measurements were introduced, that is,

$$\text{Jaccard coefficient (JC)} = \frac{|A \cap B|}{|A \cup B|}, \quad (1)$$

$$\text{Overlap coefficient (OC)} = \frac{|A \cap B|}{\min(|A|, |B|)},$$

where *A* and *B* represent the number of candidate genes in the two biological processes, respectively. To construct the biological process crosstalk, we further implemented the following rules:

- (1) Select a set of biological processes for crosstalk analysis. Only the biological processes with adjusted *P* values less than 0.05 were used. In the meantime, the biological processes containing less than five candidate genes were removed because biological processes with too few genes may have inadequate biological information

- (2) Count the number of shared candidate genes between pairs of biological processes, and remove the pair with less than three overlapped genes
- (3) Calculate the average score of the JC [24] and OC [25] values and rank them
- (4) Visualize the selected biological process crosstalk with the software Cytoscape (version 3.7.1) [26].

2.4. Construction of CMT-Related PPI Network via MCODE.

In the current study, we first identified the protein-protein interaction (PPI) network relationships of the CMT-related genes from human Integrated Interaction Database (IID) (available online: <http://iid.ophid.utoronto.ca>), which is one of the major PPI databases that integrate PPIs from multiple PPI databases, i.e., BioGRID, IntAct, I2D, MINT, InnateDB, DIP, HPRD, BIND, BCI. IID [27, 28] uses PPIs that are computationally predicted by state-of-the-art computational methods [29, 30]. Subsequently, we visualized the PPI network by means of Cytoscape software (version 3.7.1). Furthermore, the Molecular Complex Detection (MCODE) [17, 31, 32] in Cytoscape software was then performed to select appropriate modules of the PPI network to identify the most important MCODE clusters according to clustering scores. The detailed options set as degree cutoff = 2, K-core = 2, node score cutoff = 0.2, and MAX depth = 100. The study design schema is presented in Figure 1.

3. Results and Discussion

3.1. Identification of Genes Reported to Be Associated with CMT. By searching PUBMED, there were more than 600 studies related to Charcot-Marie-Tooth Disease that were collected. In these publications, it was reported that 100 genes were significantly related to CMT and formed a gene set (CMTgset) for subsequent analysis (Table S1). Among them were seven tRNA synthases: AARS1, GARS1, HARS1, KARS1, MARS1, WARS, and YARS1; four cytochrome c oxidases: COA7, COX10, COX6A1, and SCO2; four heat shock protein family members: DNAJB2, HSPB1, HSPB3, and HSPB8; two dynactins: DCTN1 and DCTN2; and two kinesin family of proteins members: KIF1B and KIF5A. Several genes were those involving the functions associated with fat metabolism (e.g., ABHD12, DGAT2, and MORC2), glucose metabolism (e.g., HK1, NAGLU, and PDK3), mitochondrial-related functional genes (e.g., AIFM1, ATP1A1, DHTKD1, HADHB, MFN2, MPV17, POLG, REEP1, SLC25A46, and SOD1), actin-related genes (e.g., FGD4, FIG4, INF2, MICAL1, and PFN2), and peripheral myelin-related genes (e.g., MPZ, PMP22, and PRX). These data indicated that the genes significantly related to CMT were sophisticated.

3.2. Biological Functions Enriched in CMTgset. Functional enrichment analysis can show a more specific function of these genes. GO enrichment analysis was performed to investigate the biological function of 100 genes in CMTgset (Table S2). In total, 200 GO terms were significantly enriched in the genes analyzed. Figure 2 shows the top 10

functions of the three functional assemblies of GO (biological process, cellular component, and molecular function). Among these clusters, some biological processes can be seen: GO terms related to myelin sheath (e.g., response to myelination, response to myelin assembly, and regulation of myelination) and axon (e.g., response to axon ensheathment and response to axonal transport) were enriched in genes in CMTgset. These results were consistent with the clinical classification of CMT, which is generally divided into demyelinating and axonal [33, 34]. Terms directly related to peripheral nervous system (e.g., peripheral nervous system development, myelination in peripheral nervous system, and peripheral nervous system axon ensheathment), Schwann cell (e.g., Schwann cell differentiation and Schwann cell development), mitochondrial function (e.g., mitochondrial fission, mitochondrion organization, and mitochondrial respiratory chain complex IV assembly), and various metabolic processes (e.g., tRNA metabolic process, cellular amino acid metabolic process, nucleoside monophosphate metabolic process, and sphingosine metabolic process) were included. Also, GO terms related to autophagy (e.g., positive regulation of autophagy, process utilizing autophagic mechanism, and regulation of autophagy) were also enriched in these genes. These results revealed that the candidate genes collected were relatively dependable for subsequent bioinformatics analysis.

3.3. Pathway Enrichment Analysis in CMTgset. Identifying biochemical pathways for the enrichment of candidate genes may provide valuable hints for us to understand the molecular mechanism of CMT. We searched for enriched pathways in the CMTgset using KOBAS 3.0 and found 14 significant enrichment pathways for CMT (Table 1). The signaling pathways of CMTgset were mainly enriched in the aminoacyl-tRNA biosynthesis, metabolic pathways, and vasopressin-regulated water reabsorption. Then use the Cytoscape to calculate the topological characteristics of the network and determine each node. Genes and pathway nodes are represented by semiellipses. The results are shown in Figure 3, in which we can see all the pathways except has00970 (aminoacyl-tRNA biosynthesis) were not independent; instead, they were connected via some genes.

3.4. Crosstalk among Significantly Enriched Biological Processes. Since CMT might involve many genes and biological processes, taking a further step to understand how significantly enriched biological processes interact with each other, we performed a biological process crosstalk analysis among the 135 significantly enriched biological processes. The method was based on the assumption that two biological processes were considered to crosstalk if they shared a proportion of CMTgset [35]. There were 83 biological processes containing five or more members in CMTgset, of which 81 biological processes met the criterion for crosstalk analysis; i.e., each biological process shared at least three genes with one or more other biological processes. All the biological process pairs (edges) formed by these biological processes were applied to construct the biological process crosstalk, and

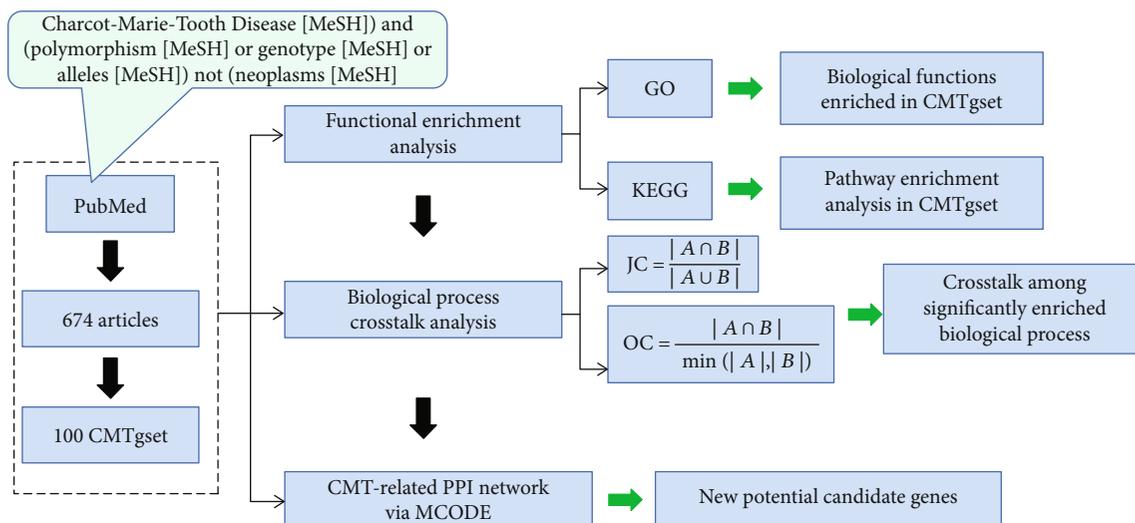


FIGURE 1: Study design and procedures. The green arrow to the right is the results.

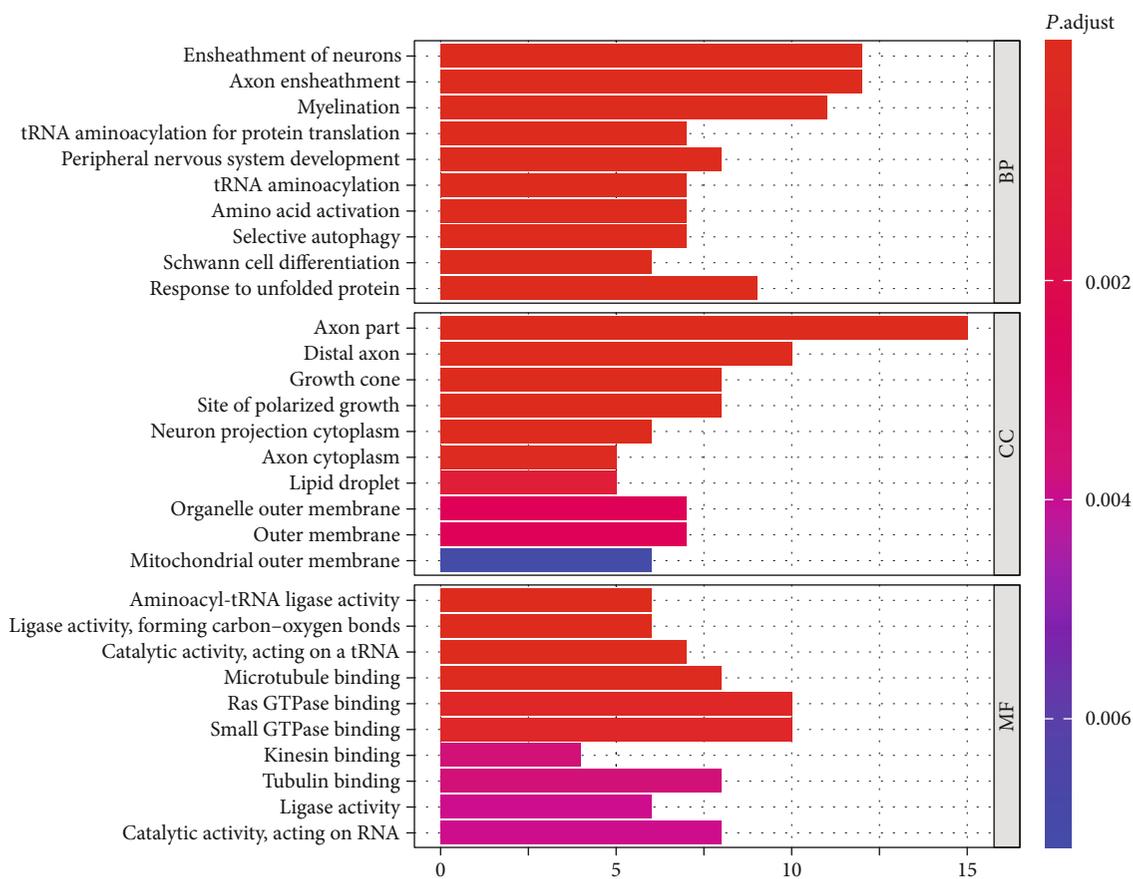


FIGURE 2: Top 10 significant GO terms and hub gene counts. For each term, the number of enriched genes is indicated by the bar size; while the level of significance is represented by the color. Blue indicates low significance while red represents high significance (FDR < 0.05). GO: Gene Ontology; BP: biological process; CC: cellular component; MF: molecular function; FDR: false discovery rate.

the overlapping level between two biological processes was measured according to the average scores of coefficients JC and OC. According to their crosstalk, the biological processes could be roughly divided into a correlative module and two

separate modules, with each module including biological processes that shared more interactions compared with other biological processes and may likely participate in the same or similar biological processes (Figure 4).

TABLE 1: Pathways enriched in CMTgset.

Pathways	ID	P value ^a	P_{BH} value ^b	Genes included in the pathway ^c
Aminoacyl-tRNA biosynthesis	hsa00970	$7.19E-10$	$2.09E-08$	GARS1, WARS, AARS1, HARS1, MARS1, YARS1, KARS1
Metabolic pathways	hsa01100	$3.16E-06$	$4.25E-05$	NAGLU, GAMT, SGPL1, COX10, COX6A1, MTMR2, PRPS1, DNMT1, SPTLC2, HADHB, HK1, DGAT2, SPTLC1, POLG
Vasopressin-regulated water reabsorption	hsa04962	$2.29E-04$	$1.63E-03$	DCTN1, DCTN2, DYNC1H1
Sphingolipid metabolism	hsa00600	$2.76E-04$	$1.89E-03$	SGPL1, SPTLC2, SPTLC1
Amyotrophic lateral sclerosis (ALS)	hsa05014	$3.46E-04$	$2.29E-03$	SOD1, NEFL, NEFH
Salmonella infection	hsa05132	$1.49E-03$	$7.65E-03$	PFN2, RAB7A, DYNC1H1
Huntington's disease	hsa05016	$1.57E-03$	$7.98E-03$	DCTN1, COX6A1, SOD1, DCTN2
Sphingolipid signaling pathway	hsa04071	$3.81E-03$	$1.61E-02$	SGPL1, SPTLC2, SPTLC1
Apoptosis	hsa04210	$5.67E-03$	$2.21E-02$	LMNA, TUBA8, AIFM1
Carbohydrate digestion and absorption	hsa04973	$6.52E-03$	$2.48E-02$	HK1, ATP1A1
Endocrine and other factor-regulated calcium reabsorption	hsa04961	$6.79E-03$	$2.55E-02$	ATP1A1, DNM2
Phagosome	hsa04145	$7.46E-03$	$2.75E-02$	TUBA8, RAB7A, DYNC1H1
Mineral absorption	hsa04978	$8.19E-03$	$2.95E-02$	ATP1A1, ATP7A
Central carbon metabolism in cancer	hsa05230	$1.31E-02$	$4.30E-02$	HK1, SCO2

CMTgset: Charcot-Marie-Tooth Disease-related gene set. ^a P values were calculated by Fisher's exact test. ^b P_{BH} values were adjusted by the Benjamini and Hochberg (BH) method. ^cOne hundred CMT-related genes included in the pathway.

The correlative module also could be grouped into five modules, which could be defined as nervous system-related biological processes (e.g., peripheral nervous system development, ensheathment of neurons, axon ensheathment, and myelination), antigen presentation processes (e.g., antigen processing and presentation of exogenous peptide antigen, antigen processing and presentation of exogenous peptide antigen via MHC class II, and antigen processing and presentation of peptide antigen via MHC class II), transport processes (e.g., transport along microtubule, cytoskeleton-dependent intracellular transport, and axonal transport), oxidation processes (e.g., response to reactive oxygen species, response to oxidative stress, and cellular response to toxic substance), and mitochondrial related biological processes (e.g., mitochondrion organization, purine ribonucleotide metabolic process, and ATP metabolic process). At the same time, these five modules were not independent, but connected by the interaction of several biological processes. The other two separate modules, one related to autophagy (e.g., positive regulation of autophagy and macroautophagy) and one related to tRNA (e.g., tRNA aminoacylation for protein translation, tRNA aminoacylation, and tRNA metabolic process).

3.5. Construction of CMT-Specific Protein Network via MCODE. Through the online PPI analysis of IID, 88 of the 100 candidate genes can be mapped to the human interactome network, and 1698 predicted genes were screened out. Pathway-based MCODE cluster analysis was used to analyze

the topological characteristics of PPI to help understand the potential biological mechanisms associated with the network. The higher the clustering scores, the more important the biological function of this clustering in the development of CMT. Then, we analyzed in detail the three clusters with the highest clustering scores. The results are shown in Figure 5. In Figure 5(a), the PPI cluster (MCODE cluster score = 4) had 40 genes in total, including 12 (30.00%) gene members in the CMTgset, while 28 (70.00%) genes were not included in the list. In Figure 5(b), the PPI cluster (MCODE cluster score = 3.905) had 43 genes in total, including 13 (30.23%) gene members in the CMTgset, but 30 (69.77%) genes were not in the list. In Figure 5(c), the PPI cluster (MCODE cluster score = 3.769) had 27 genes in total, in which 10 (37.04%) gene members were CMTgset but 17 (62.96%) genes have not been reported before. In these 3 PPI clusters, 13 of the CMTgset were included in the human interactome network, among which 30 predicted genes are likely to be highly associated with CMT based on MCODE clusters. These 30 genes are listed in Table 2, which provided a list of new potential candidates for CMT.

4. Discussion

In the past few decades, much has been learnt about the molecular mechanisms underlying Charcot-Marie-Tooth Disease from studies on cell models, animals, or human subjects [36–38]. With the high-throughput technology, developing more and more genes/proteins has been recognized

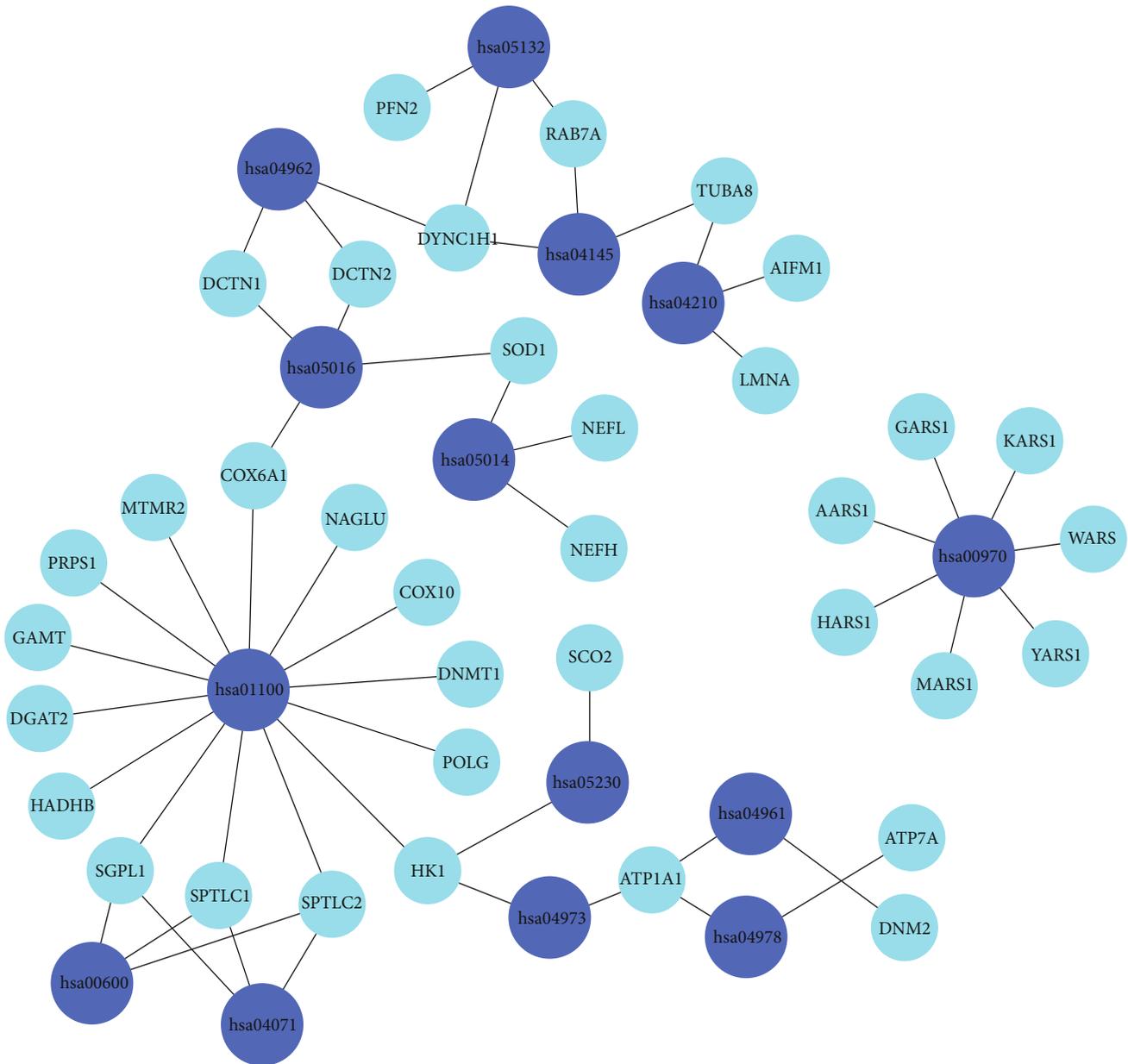


FIGURE 3: Significant pathway enrichment of CMTgset. Dark blue represents signaling pathway, and light blue represents candidate genes.

to be related to this disorder [15, 39], but it is still far from enough to fully understand the biological processes related to the pathogenesis of CMT at the molecular level. Therefore, the potential pathogenesis of CMT needs to be decoded at the system biology level. In this study, by collecting genes genetically related to CMT and using functional enrichment and network analysis to systematically explore the interaction of these genes, we provided a comprehensive and systematic framework to describe the relevant biochemical processes.

Although genetic association and biochemical studies based on candidate genes have provided us with the knowledge of factors involved in CMT, the systematic approach depicted in this work has clear advantages. Above all, we have comprehensively collected genes potentially genetically related to CMT in our study, which provided a valuable

resource for further analysis. From the perspective of molecular network level, it is very important to explore the biological characteristics of genes related to CMT. Moreover, functional enrichment analysis considering the biological relevance of genes can more robustly deal with possible false positives caused by different genes in various studies, and combining with network analysis could provide a more comprehensive view of the molecular mechanism of CMT.

Biological function enrichment analysis revealed the specific biological processes involved by CMTgset. GO enrichment analysis showed that these genes for CMT participated in myelin sheath and axon-related processes, peripheral nervous system, Schwann cell, mitochondrial function, various metabolic processes, and autophagy. In addition, terms such as ensheathment of neurons, axon

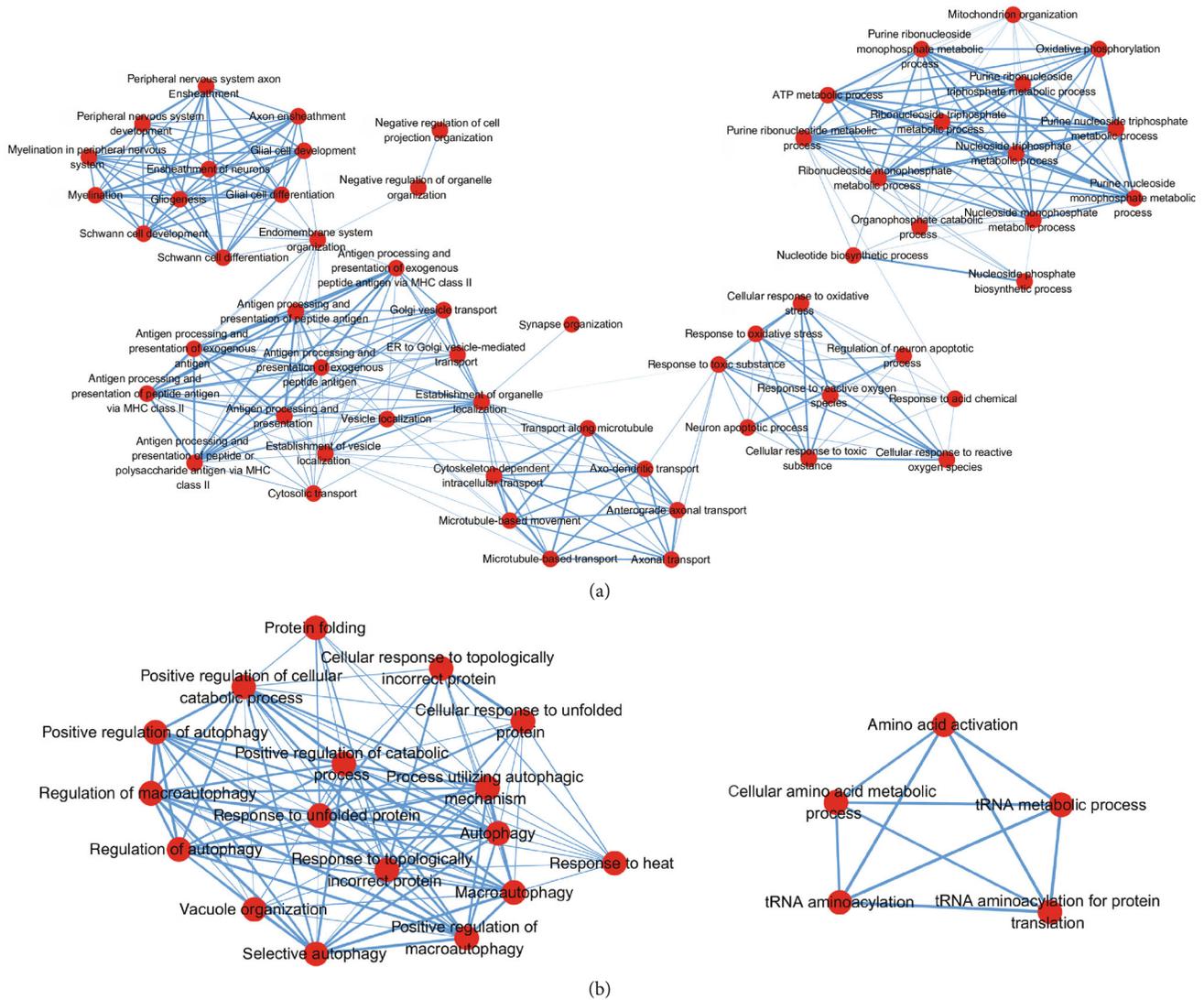


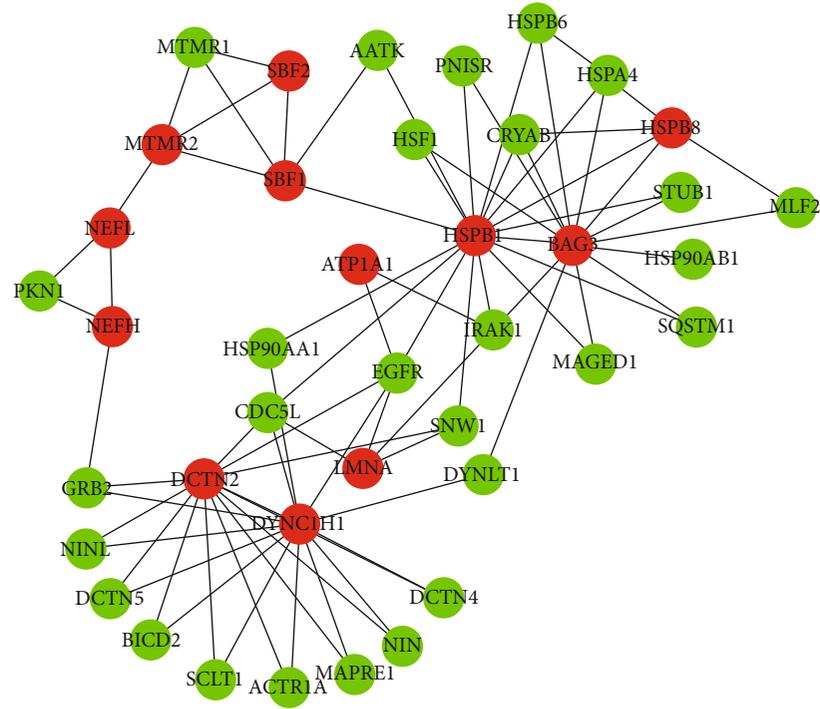
FIGURE 4: Biological process crosstalk among CMTgset-enriched biological processes. Nodes represent biological processes, and edges represent crosstalk between biological processes. Edge-width corresponds to the score of specific biological process pair. Larger edge-width indicates higher score. (a) Represents one correlative module. (b) Represents two separate modules.

ensheathment, myelination, tRNA aminoacylation for protein translation, peripheral nervous system development, tRNA aminoacylation, amino acid activation, selective autophagy, Schwann cell differentiation, and response to unfolded protein were in the top ten enriched GO terms, indicating the important roles of these activities in the pathologic processes of CMT. The above findings are basically consistent with previous reports [13, 33, 40, 41].

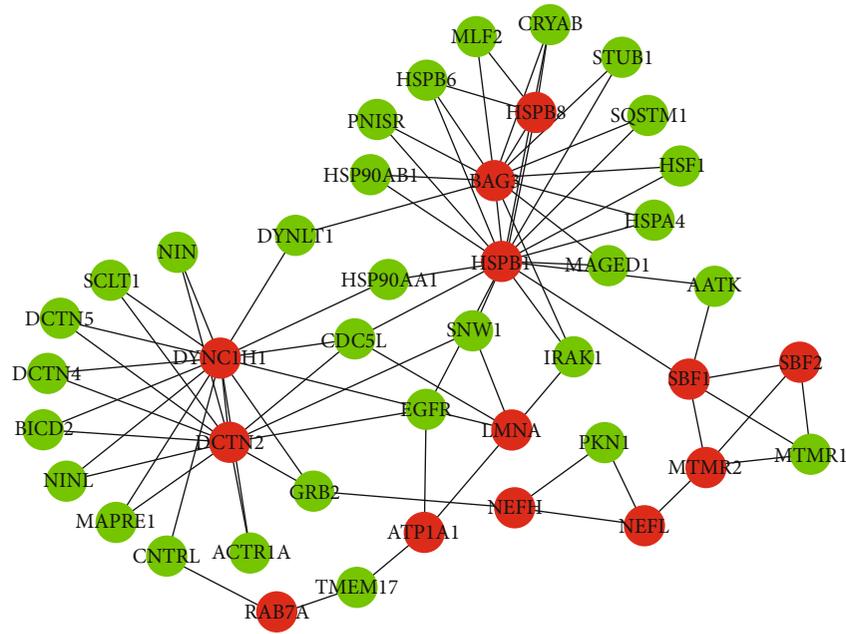
At the same time, the pathway analysis showed that 14 pathways were enriched and the first of which is aminoacyl-tRNA biosynthesis, which is considered to play an important role in CMT. Aminoacyl-tRNA synthetases (ARSs) are universally expressed enzymes accountable for charging tRNAs with their cognate amino acids, so it is crucial for the first step of protein synthesis [42, 43]. Just as illustrated by a large number of mutations in cytosolic and bifunctional tRNA synthetases causing CMT, the peripheral nerves are often affected [44]. The main mechanisms pro-

posed consist of reduction of aminoacylation activity, alteration of dimerization or localization, interaction of functionally acquired pathogens, and loss of noncanonical functions [45].

Of significance, in biological process crosstalk analysis, we identified a correlative module and two separate modules. The correlative module could be grouped into five modules: the first module was mainly dominated by the biological processes related to the activity of the nervous system. Among these biological processes, axon ensheathment, myelination, peripheral nervous system development, and Schwann cell development have been well studied, involving the axons, myelin sheaths, or peripheral nervous system as well as the progress of Charcot-Marie-Tooth Disease [46–49]. The second module and the third module can be defined together as a transmission function-related biological process. For example, many studies have determined that many major neurodevelopmental and neurodegenerative diseases are



(a)



(b)

FIGURE 5: Continued.

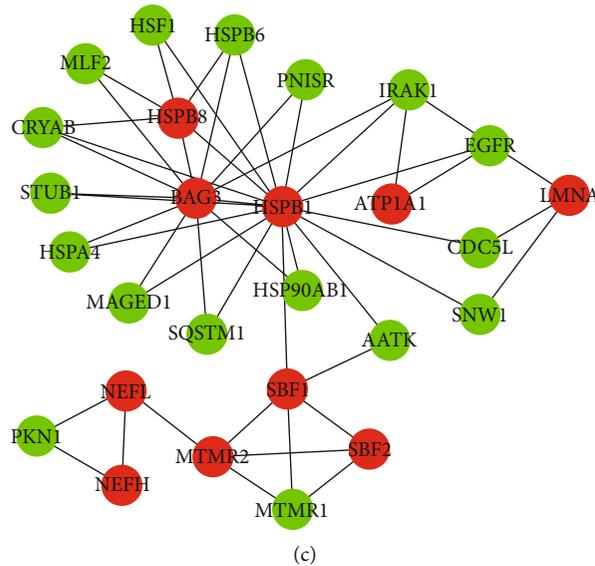


FIGURE 5: The red nodes are genes of CMTset, and the green nodes are nonoriginal/extended genes.

associated with mutations in microtubule-associated proteins. The microtubules play the role of tracks for organelle transport, and in neurons, microtubules give assistance to transporting membrane-bound organelles, extending neurites during development, providing scaffolding for neurites, and maintaining intracellular compartments [13, 50, 51]. Apart from this, insufficient axon transport is a common theme in many neurodegenerative diseases [52, 53], and some researches have found that NEFL [54] and HSPB1 [55] (included in the GO term of axon transport) are closely related to axonal transport dysfunction of CMT. The fourth and fifth modules are both primarily associated with energy-related biological processes, such as oxidative stress response (e.g., response to oxidative stress and oxidative phosphorylation) and mitochondrial related processes (e.g., mitochondrion organization and ATP metabolic process). For the purpose of maintaining membrane excitability and processing neurotransmission and plasticity, neurons rely on mitochondrial function [56]. Moreover, multiple studies have demonstrated that the pathology of CMT has been related to mitochondrial dynamics [57–60]. In addition, this correlative module is connected through several edges, indicating that these biological processes or modules may play a synergistic role in the pathogenesis of CMT, rather than acting alone.

The first three important PPI networks of the CMT-related genes were extracted from the human reference interactome network by MCODE clustering, which extends to its neighboring nodes with the seed node as the center. Then, the nodes that may interact with the seed nodes to build a complex in the PPI network were selected [61, 62]. It is worth noting that 30 extended genes appeared in the first three PPI networks, but were not included in CMT-related genes, which has not been previously reported to be related to CMT. For example, BICD2 (BICD cargo adaptor 2) has been implicated in dynein-mediated, minus end-directed motility along microtubules. Since whole-exome sequencing (WES)

has proven to be an efficient tool for mutation screening CMT [10], a novel variant c.1079C>T (p.A360V) in the gene BICD2 is identified in an individual with demyelinating CMT but requires further evidence of pathogenicity [63]. HSP90AA1, HSP90AB1, HSPA4, and HSPB6 are members of the heat shock protein family, which involves many physiological and pathological stresses, such as endoplasmic reticulum stress, protein misfolding and aggregation, and oxidative stress. Most importantly, many members of the heat shock protein family have been shown to resist many degenerative diseases, including CMT [64, 65]. EGFR (epidermal growth factor receptor) is a family member of the ErbB receptor kinase, which can regulate important signaling pathways (including cell proliferation and differentiation, cell cycle, and migration) and stimulate the growth and differentiation of neurons [66, 67]. There is increasing evidence that ErbB receptor-mediated signaling plays an important role in controlling Schwann cell axon communication and myelination in the peripheral nervous system [68]. Genes play a vital biological role in the form of proteins. So, the above events or functions reflected by the PPI network may be closer to the actual biological processes of CMT-related genes. As demonstrated by the results elaborated above, our methods extend the findings of previous studies and provide more information about the pathogenesis of CMT. At the same time, our MCODE cluster analysis not only provided a meaningful inferred network of CMT-related genes but also has the potential to identify potential candidate genes.

So far, the detection and analysis approaches of CMT are diverse, including linkage analysis, functional studies, and studies in the model organisms. However, the research of this disorder's molecular mechanism have not yet broken through, which may be due to the inability of conventional single-gene analyses to explain complicated psychiatric phenotype. This paper takes a comprehensive analysis of potential causal genes within a pathway [35, 69] and/or a network [70, 71] framework which might provide more important

TABLE 2: Genes included in CMT top three specific PPI networks but not in the CMTgset.

Gene symbol	Gene name	Cluster
MTMR1	Myotubularin-related protein 1	Clusters A, B, and C
AATK	Apoptosis-associated tyrosine kinase	Clusters A, B, and C
HSF1	Heat shock transcription factor 1	Clusters A, B, and C
PNISR	PNN interacting serine and arginine-rich protein	Clusters A, B, and C
HSPB6	Heat shock protein family B (small) member 6	Clusters A, B, and C
CRYAB	Crystallin alpha B	Clusters A, B, and C
HSPA4	Heat shock protein family A (Hsp70) member 4	Clusters A, B, and C
STUB1	STIP1 homology and U-box containing protein 1	Clusters A, B, and C
MLF2	Myeloid leukemia factor 2	Clusters A, B, and C
HSP90AB1	Heat shock protein 90 alpha family class B member 1	Clusters A, B, and C
SQSTM1	Sequestosome 1	Clusters A, B, and C
MAGED1	MAGE family member D1	Clusters A, B, and C
IRAK1	Interleukin 1 receptor-associated kinase 1	Clusters A, B, and C
EGFR	Epidermal growth factor receptor	Clusters A, B, and C
SNW1	SNW domain containing 1	Clusters A, B, and C
CDC5L	Cell division cycle 5 like	Clusters A, B, and C
PKN1	Protein kinase N1	Clusters A, B, and C
DYNLT1	Dynein light chain Tctex-type 1	Clusters A and B
GRB2	Growth factor receptor bound protein 2	Clusters A and B
NINL	Ninein like	Clusters A and B
DCTN5	Dynactin subunit 5	Clusters A and B
BICD2	BICD cargo adaptor 2	Clusters A and B
SCLT1	Sodium channel and clathrin linker 1	Clusters A and B
ACTR1A	Actin-related protein 1A	Clusters A and B
MAPRE1	Microtubule-associated protein RP/EB family member 1	Clusters A and B
NIN	Ninein	Clusters A and B
DCTN4	Dynactin subunit 4	Clusters A and B
HSP90AA1	Heat shock protein 90 alpha family class A member 1	Clusters A and B
CNTRL	Centriolin	Cluster B
TMEM17	Transmembrane protein 17	Cluster B

CMT = Charcot-Marie-Tooth Disease; PPI = protein-protein interaction.

insights. In addition, the data set CMTgset obtained by the approach proposed in this paper included the most recent researches. In CMTgset, we used a unified gene symbol standard to collect genes that are significantly related to CMT reported by the original authors. Since most of the genes came from association researches on individual genes, some of the genes may only have a moderate *P* value, but they worked synergistically with other genes to show a significant association with CMT, which made CMTgset a more comprehensive data set for CMT exploration.

Of course, this study still has a few limitations. First of all, our functional enrichment analysis and PPI network analysis results are entirely dependent on the genes reported in the retrieved literature that are related to CMT, but more evidence needs to be used to further verify specific new genes. In other words, the results obtained by/in the bioinformatics approach should be further verified in the DNA samples obtained from CMT-affected patients. Secondly, our research accepts the results of the original authors of each retrieved

study. Due to the imbalance and incompleteness of these existing studies, they will certainly bias our results. Third, to reduce the false positive rate of genes, we excluded publications that were negative or irrelevant. However, some of the genes in these studies may be related to the pathogenesis of CMT and were only excluded because of the small sample size or heterogeneity or other factors. In addition, although the quantity and quality of PPI databases have been greatly improved recently, the human interaction group is still incomplete and has many false positives.

5. Conclusion

In this study, we analyzed related genes collected from selected literature deposited in PUBMED, using the system biology framework to conduct a comprehensive, systematic biological function and network-based analysis of CMT. By integrating the information from GO, pathways, and biological process crosstalk analysis, the conclusions of this study

may help to update the new understanding of the pathogenesis of CMT and expand the potential genes of CMT for further exploration.

Data Availability

The statistical data of the article used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by grants from (1) the diagnostic value of circRNAs in breast cancer and the ceRNA regulatory network mechanism (2019YFS0332), (2) the key technologies of development and application of based on biochemical pipeline intelligent checks and interpretation system (2019YFH0010), and (3) construction of information management and shared service platform centered on clinical sample resources (2019YFS0038), all from the Science & Technology Department of Sichuan Province.

Supplementary Materials

Supplemental Table 1: list of genes associated with Charcot-Marie-Tooth Disease. Supplemental Table 2: Gene Ontology and biological process terms enriched in CMT-related genes. (*Supplementary Materials*)

References

- [1] J. M. Charcot and P. Marie, "Sur une form particulie're d'atrophie musculaire progressive, souvent familiale, debutant par les pieds et les jambes, et atteignant plus tard les mains," *Revolutions Medicines*, vol. 6, pp. 97–138, 1886.
- [2] H. H. Tooth, *The Peroneal Type of Progressive Muscular Atrophy*, H.K. Lewis, London, 1886.
- [3] H. Skre, "Genetic and clinical aspects of Charcot-Marie-Tooth's disease," *Clinical Genetics*, vol. 6, no. 2, pp. 98–118, 1974.
- [4] C. N. Martyn and R. A. Hughes, "Epidemiology of peripheral neuropathy," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 62, no. 4, pp. 310–318, 1997.
- [5] F. Désarnaud, A. N. Do Thi, A. M. Brown et al., "Progesterone stimulates the activity of the promoters of peripheral myelin protein-22 and protein zero genes in Schwann cells," *Journal of Neurochemistry*, vol. 71, no. 4, pp. 1765–1768, 1998.
- [6] M. W. Sereda, G. Meyer zu Hörste, U. Suter, N. Uzma, and K. A. Nave, "Therapeutic administration of progesterone antagonist in a model of Charcot-Marie-Tooth disease (CMT-1A)," *Nature Medicine*, vol. 9, no. 12, pp. 1533–1537, 2003.
- [7] E. Passage, J. C. Norreel, P. Noack-Fraissignes et al., "Ascorbic acid treatment corrects the phenotype of a mouse model of Charcot-Marie-Tooth disease," *Nature Medicine*, vol. 10, no. 4, pp. 396–401, 2004.
- [8] Z. Sahenk, H. N. Nagaraja, B. S. McCracken et al., "NT-3 promotes nerve regeneration and sensory improvement in CMT1A mouse models and in patients," *Neurology*, vol. 65, no. 5, pp. 681–689, 2005.
- [9] C. Gonzaga-Jauregui, T. Harel, T. Gambin et al., "Exome sequence analysis suggests that genetic burden contributes to phenotypic variability and complex neuropathy," *Cell Reports*, vol. 12, no. 7, pp. 1169–1183, 2015.
- [10] B.-O. Choi, S. K. Koo, M.-H. Park et al., "Exome sequencing is an efficient tool for genetic screening of Charcot-Marie-Tooth disease," *Human Mutation*, vol. 33, no. 11, pp. 1610–1615, 2012.
- [11] V. Fridman, B. Bundy, M. M. Reilly et al., "CMT subtypes and disease burden in patients enrolled in the Inherited Neuropathies Consortium Natural History study: a cross-sectional analysis," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 86, no. 8, pp. 873–878, 2015.
- [12] A. M. Rossor, P. J. Tomaselli, and M. M. Reilly, "Recent advances in the genetic neuropathies," *Current Opinion in Neurology*, vol. 29, no. 5, pp. 537–548, 2016.
- [13] N. U. Jerath and M. E. Shy, "Hereditary motor and sensory neuropathies: understanding molecular pathogenesis could lead to future treatment strategies," *Biochimica et Biophysica Acta*, vol. 1852, no. 4, pp. 667–678, 2015.
- [14] A. M. Rossor, J. M. Polke, H. Houlden, and M. M. Reilly, "Clinical implications of genetic advances in charcot-marie-tooth disease," *Nature Reviews. Neurology*, vol. 9, no. 10, pp. 562–571, 2013.
- [15] M. Pipis, A. M. Rossor, M. Laura, and M. M. Reilly, "Next-generation sequencing in Charcot-Marie-Tooth disease: opportunities and challenges," *Nature Reviews Neurology*, vol. 15, no. 11, pp. 644–656, 2019.
- [16] O. A. Alli, O. D. Ogbolu, and O. O. Alaka, "Direct molecular detection of Mycobacterium tuberculosis complex from clinical samples—an adjunct to cultural method of laboratory diagnosis of tuberculosis," *North American Journal of Medical Sciences*, vol. 3, no. 6, pp. 281–288, 2011.
- [17] R. Zhang, H. Guo, X. Yang et al., "Pathway-based network analyses and candidate genes associated with Kashin-Beck disease," *Medicine*, vol. 98, no. 18, article e15498, 2019.
- [18] H. Gu, Z. Huang, G. Chen et al., "Network and pathway-based analyses of genes associated with osteoporosis," *Medicine*, vol. 99, no. 8, article e19120, 2020.
- [19] Y. Hu, Z. Pan, Y. Hu, L. Zhang, and J. Wang, "Network and pathway-based analyses of genes associated with Parkinson's disease," *Molecular Neurobiology*, vol. 54, no. 6, pp. 4452–4465, 2017.
- [20] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *Omics: A Journal of Integrative Biology*, vol. 16, no. 5, pp. 284–287, 2012.
- [21] D. Zhou, Y. Sun, Y. Jia et al., "Bioinformatics and functional analyses of key genes in smoking-associated lung adenocarcinoma," *Oncology Letters*, vol. 18, no. 4, pp. 3613–3622, 2019.
- [22] C. Xie, X. Mao, J. Huang et al., "KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases," *Nucleic Acids Research*, vol. 39, Supplement 2, pp. W316–W322, 2011.
- [23] X. Yang, S. M. Zhu, L. Li et al., "Identification of differentially expressed genes and signaling pathways in ovarian cancer by integrated bioinformatics analysis," *Oncotargets and Therapy*, vol. 11, pp. 1457–1474, 2018.

- [24] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of jaccard coefficient for keywords similarity," in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 1, pp. 13–15, Hong Kong, 2013.
- [25] S. Mizuno, T. Yamaguchi, A. Fukushima, Y. Matsuyama, and Y. Ohashi, "Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations," *Clinical Trials*, vol. 2, no. 2, pp. 174–181, 2016.
- [26] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [27] M. Kotlyar, C. Pastrello, Z. Malik, and I. Jurisica, "IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species," *Nucleic Acids Research*, vol. 47, no. D1, pp. D581–D589, 2019.
- [28] H. Nakamura, K. Fujii, V. Gupta et al., "Identification of key modules and hub genes for small-cell lung carcinoma and large-cell neuroendocrine lung carcinoma by weighted gene co-expression network analysis of clinical tissue-proteomes," *PLoS One*, vol. 14, no. 6, article e0217105, 2019.
- [29] A. Elefsinioti, Ö. S. Saraç, A. Hegele et al., "Large-scale *de novo* prediction of physical protein-protein association.," *Molecular & Cellular Proteomics*, vol. 10, no. 11, 2011.
- [30] M. Kotlyar, C. Pastrello, F. Pivetta et al., "In silico prediction of physical protein interactions and characterization of interactome orphans," *Nature Methods*, vol. 12, no. 1, pp. 79–84, 2015.
- [31] G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC Bioinformatics*, vol. 4, no. 1, p. 2, 2003.
- [32] H. Tu, M. Wu, W. Huang, and L. Wang, "Screening of potential biomarkers and their predictive value in early stage non-small cell lung cancer: a bioinformatics analysis," *Translational Lung Cancer Research*, vol. 8, no. 6, pp. 797–807, 2019.
- [33] C. J. Klein, X. Duan, and M. E. Shy, "Inherited neuropathies: clinical overview and update," *Muscle & Nerve*, vol. 48, no. 4, pp. 604–622, 2013.
- [34] T. Stojkovic, "Hereditary neuropathies: an update," *Revue Neurologique*, vol. 172, no. 12, pp. 775–778, 2016.
- [35] J. J. Goeman and P. Buhlmann, "Analyzing gene expression data in terms of gene sets: methodological issues," *Bioinformatics (Oxford, England)*, vol. 23, no. 8, pp. 980–987, 2007.
- [36] M. Jouaud, S. Mathis, L. Richard, A.-S. Lia, L. Magy, and J.-M. Vallat, "Rodent models with expression of PMP22: relevance to dysmyelinating CMT and HNPP," *Journal of the Neurological Sciences*, vol. 398, pp. 79–90, 2019.
- [37] P. Sancho, L. Bartesaghi, O. Miossec et al., "Characterization of molecular mechanisms underlying the axonal Charcot-Marie-Tooth neuropathy caused by MORC2 mutations," *Human Molecular Genetics*, vol. 28, no. 10, pp. 1629–1644, 2019.
- [38] V. L. del Amo, M. Palomino-Schätzlein, M. Seco-Cervera et al., "A Drosophila model of GDAP1 function reveals the involvement of insulin signalling in the mitochondria-dependent neuromuscular degeneration," *Biochimica et Biophysica Acta - Molecular Basis of Disease*, vol. 1863, no. 3, pp. 801–809, 2017.
- [39] C. Saghira, D. M. Bis, D. Stanek et al., "Variant pathogenicity evaluation in the community-driven Inherited Neuropathy Variant Browser," *Human Mutation*, vol. 39, no. 5, pp. 635–642, 2018.
- [40] T. Murakami and Y. Sunada, "Schwann cell and the pathogenesis of Charcot-Marie-Tooth disease," *Advances in Experimental Medicine and Biology*, vol. 1190, pp. 301–321, 2019.
- [41] M. Fontés, "Charcot Marie Tooth disease. A single disorder?," *International Journal of Molecular Sciences*, vol. 19, no. 12, p. 3807, 2018.
- [42] R. Meyer-Schuman and A. Antonellis, "Emerging mechanisms of aminoacyl-tRNA synthetase mutations in recessive and dominant human disease," *Human Molecular Genetics*, vol. 26, no. R2, pp. R114–R127, 2017.
- [43] A. Antonellis and E. D. Green, "The role of aminoacyl-tRNA synthetases in genetic diseases," *Annual Review of Genomics and Human Genetics*, vol. 9, no. 1, pp. 87–107, 2008.
- [44] J. Ognjenović and M. Simonović, "Human aminoacyl-tRNA synthetases in diseases of the nervous system," *RNA Biology*, vol. 15, no. 4-5, pp. 623–634, 2018.
- [45] V. Boczonadi, M. J. Jennings, and R. Horvath, "The role of tRNA synthetases in neurological and neuromuscular disorders," *FEBS Letters*, vol. 592, no. 5, pp. 703–717, 2018.
- [46] M. A. C. Saporta, B. R. Shy, A. Patzko et al., "MpzR98C arrests Schwann cell development in a mouse model of early-onset Charcot-Marie-Tooth disease type 1B," *Brain*, vol. 135, no. 7, pp. 2032–2047, 2012.
- [47] R. Fledrich, T. Abdelaal, L. Rasch et al., "Targeting myelin lipid metabolism as a potential therapeutic strategy in a model of CMT1A neuropathy," *Nature Communications*, vol. 9, no. 1, article 3025, 2018.
- [48] M. Yamaguchi and H. Takashima, "Drosophila Charcot-Marie-Tooth disease models," *Advances in Experimental Medicine and Biology*, vol. 1076, pp. 97–117, 2018.
- [49] H. Wang, M. Davison, K. Wang et al., "Transmembrane protease serine 5: a novel Schwann cell plasma marker for CMT1A," *Annals of Clinical and Translational Neurology*, vol. 7, no. 1, pp. 69–82, 2019.
- [50] L. C. Kapitein and C. C. Hoogenraad, "Building the neuronal microtubule cytoskeleton," *Neuron*, vol. 87, no. 3, pp. 492–506, 2015.
- [51] K. Barlan and V. I. Gelfand, "Microtubule-based transport and the distribution, tethering, and organization of organelles," *Cold Spring Harbor Perspectives in Biology*, vol. 9, no. 5, 2017.
- [52] S. Millecamps and J.-P. Julien, "Axonal transport deficits and neurodegenerative diseases," *Nature Reviews Neuroscience*, vol. 14, no. 3, pp. 161–176, 2013.
- [53] S. Roy, B. Zhang, V. M.-Y. Lee, and J. Q. Trojanowski, "Axonal transport defects: a common theme in neurodegenerative diseases," *Acta Neuropathologica*, vol. 109, no. 1, pp. 5–13, 2005.
- [54] R. Perez-Olle, M. A. Lopez-Toledano, D. Goryunov et al., "Mutations in the neurofilament light gene linked to Charcot-Marie-Tooth disease cause defects in transport," *Journal of Neurochemistry*, vol. 93, no. 4, pp. 861–874, 2005.
- [55] J. Zhai, H. Lin, J.-P. Julien, and W. W. Schlaepfer, "Disruption of neurofilament network with aggregation of light neurofilament protein: a common pathway leading to motor neuron degeneration due to Charcot-Marie-Tooth disease-linked mutations in NFL and HSPB1," *Human Molecular Genetics*, vol. 16, no. 24, pp. 3103–3116, 2007.
- [56] A. B. Knott, G. Perkins, R. Schwarzenbacher, and E. Bossy-Wetzel, "Mitochondrial fragmentation in neurodegeneration," *Nature Reviews Neuroscience*, vol. 9, no. 7, pp. 505–518, 2008.
- [57] P. González-Sánchez, J. Satrústegui, F. Palau, and A. del Arco, "Calcium deregulation and mitochondrial bioenergetics in GDAP1-related CMT disease," *International Journal of Molecular Sciences*, vol. 20, no. 2, p. 403, 2019.

- [58] C.-j. Luan, W. Guo, L. Chen et al., "CMT2Q-causing mutation in the Dhdkd1 gene lead to sensory defects, mitochondrial accumulation and altered metabolism in a knock-in mouse model," *Acta Neuropathologica Communications*, vol. 8, no. 1, p. 32, 2020.
- [59] K. Panchal and A. K. Tiwari, "Mitochondrial dynamics, a key executioner in neurodegenerative diseases," *Mitochondrion*, vol. 47, pp. 151–173, 2019.
- [60] A. Niemann, K. M. Wagner, M. Ruegg, and U. Suter, "GDAP1 mutations differ in their effects on mitochondrial dynamics and apoptosis depending on the mode of inheritance," *Neurobiology of Disease*, vol. 36, no. 3, pp. 509–520, 2009.
- [61] P. O'Driscoll, E. Merenyi, C. Karmonik, and R. Grossman, "SOM and MCODE methods of defining functional clusters in MRI of the brain," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 734–737, Chicago, IL, USA, August 2014.
- [62] N. Lin, J. Jiang, S. Guo, and M. Xiong, "Functional principal component analysis and randomized sparse clustering algorithm for medical image analysis," *PLoS One*, vol. 10, no. 7, article e0132945, 2015.
- [63] A. P. Drew, D. Zhu, A. Kidambi et al., "Improved inherited peripheral neuropathy genetic diagnosis by whole-exome sequencing," *Molecular Genetics & Genomic Medicine*, vol. 3, no. 2, pp. 143–154, 2015.
- [64] L. K. Muranova, A. S. Ryzhavskaia, M. V. Sudnitsyna, V. M. Shatov, and N. B. Gusev, "Small heat shock proteins and human neurodegenerative diseases," *Biochemistry Biokhimiia*, vol. 84, no. 11, pp. 1256–1267, 2019.
- [65] R. I. Morimoto, "Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging," *Genes & Development*, vol. 22, no. 11, pp. 1427–1438, 2008.
- [66] R. J. Webster, K. M. Giles, K. J. Price, P. M. Zhang, J. S. Matlack, and P. J. Leedman, "Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7," *The Journal of Biological Chemistry*, vol. 284, no. 9, pp. 5731–5741, 2009.
- [67] A. Currais, T. Hortobágyi, and S. Soriano, "The neuronal cell cycle as a mechanism of pathogenesis in Alzheimer's disease," *Aging*, vol. 1, no. 4, pp. 363–371, 2009.
- [68] S. M. Lee, L.-S. Chin, and L. Li, "Dysregulation of ErbB receptor trafficking and signaling in demyelinating Charcot-Marie-Tooth disease," *Molecular Neurobiology*, vol. 54, no. 1, pp. 87–100, 2017.
- [69] G. V. Glazko and F. Emmert-Streib, "Unite and conquer: univariate and multivariate approaches for finding differentially expressed gene sets," *Bioinformatics*, vol. 25, no. 18, pp. 2348–2354, 2009.
- [70] P. Jia, C. F. Kao, P. H. Kuo, and Z. Zhao, "A comprehensive network and pathway analysis of candidate genes in major depressive disorder," *BMC Systems Biology*, vol. 5, Supplement 3, p. S12, 2011.
- [71] P. Jia, S. Zheng, J. Long, W. Zheng, and Z. Zhao, "dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks," *Bioinformatics (Oxford, England)*, vol. 27, no. 1, pp. 95–102, 2011.