國立臺灣大學醫學院分子醫學研究所

碩士論文

Department of Graduate Institute of Molecular Medicine
College of Medicine
National Taiwan University

Master Thesis

遺傳疾病次世代基因分析套組之建立

Development of a Next Generation Sequencing Panel for Inherited Diseases

林宜霖

Yi-Lin Lin

指導教授：簡穎秀 博士

Advisor: Yin-Hsiu Chien, M.D., Ph. D.

中華民國 106 年 7 月

July 2017

doi:10.6342/NTU201704070

國立臺灣大學（碩）博士學位論文
口試委員會審定書

論文中文題目

遺傳疾病次世代基因分析套組之建立

論文英文題目

Development of a Next Generation Sequencing Panel for Inherited Diseases

本論文係 林宜霖 君（學號 P04448013）在 國立臺灣大學 醫學院 分子醫學研究所 完成之碩士學位論文，於民國 106 年 07 月 17 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

（簽名）

（指導教授）

系主任、所長 （簽名）

doi:10.6342/NTU201704070

本研究榮獲財團法人罕見疾病基金會
「第十八屆罕見疾病博碩士論文獎助學金」獎助
謹於此特別致謝

摘要

背景

基因分析為遺傳疾病確認診斷的工具。毛細管電泳是基因序列分析的標準方法，雖然準確，但因為人力和時間的耗費，價格無法降低。最近由於基因醫學的快速進步，許多臨床上表現相似的遺傳疾病，已知是由多個基因引起。若使用毛細管電泳來分析，其價格依基因數目倍增，變得相當龐大。次世代基因分析用晶片的觀念，可同時分析許多的基因，甚至全基因 (whole genome) 分析都做得到。

目的

本研究擬建立遺傳疾病次世代基因分析套組，設計高達408個基因，包含肢帶肌肉失養症、Charcot Marie Tooth氏症 (CMT, 進行性神經性腓骨萎縮症)、Brugada 猝死症、努南症候群及溶小體儲積症等。可應用在多種疾病，符合臨床運用的目的。

方法

我們從受檢者血液中抽取DNA，標靶序列捕獲捕捉平台設計所含之基因的外顯子，聚焦在408個目標基因，再用生物資訊技術分析這些捕捉的序列。本研究先用四群疾病(肢帶肌肉失養症、CMT、神經肌肉疾病、成骨不全症／Ehlers-Danlos症候群)共40名病人來測試；並且同時監測各外顯子捕捉之效率。

結果

此NGS基因分析套組的平均覆蓋率為174.9x，95.7%的核苷酸讀深超過30x。在40人中，致病性變異點可在16個人身上被偵測到 (40%)。其中，又以肌肉失養症的診斷率最高，21個肌肉失養症病人可以確認出10名患者的致病基因變化(47.6%) ，其次為CMT 33% (2/6) 和OI／EDS 33% (3/9)。

結論

次世代定序基因套組可以快速有效的偵測患者的致病基因變化，提供患者與醫療人員得知正確的診斷以供後續照顧參考。

關鍵詞：次世代定序、基因分析套組、肢帶肌肉失養症、Charcot Marie Tooth 氏症、神經肌肉疾病

doi:10.6342/NTU201704070

ABSTRACT

Background

Genetic analysis is the conformational tool for the diagnosis of inherited diseases. With the rapidly progress of genetic medicine, many inherited diseases with similar clinical presentations are known to be caused by several of genes. Sanger sequencing is the gold standard for DNA sequencing. Though accurate, analyzing numerous genes one by one using is very costly. Next generation sequencing (NGS) is a concept of massive parallel sequencing by chips that can inspect many genes, even the whole genome, at the same time.

Purpose

In this project, we plan to set up an NGS panel of 408 genes to analyze inherited diseases, for instance, limb-girdle muscular dystrophy (LGMD), Charcot-Marie Tooth disease (CMT), neurometabolic disease (NMD), osteogenesis imperfecta (OI) / Ehlers-Danlos syndrome (EDS), Brugada syndrome, and lysosomal storage disease (LSD). Ultimately, we would like to use the technique not only for research but also apply to clinical service.

Methods

We analyzed 40 patients clinically diagnosed with LGMD, CMT, NMD, and OI/EDS. Captured-based exome sequencing was then applied, followed by bioinformatics analysis on the targeted genes. We also examined the coverage, accuracy, sensitivity, and specificity of the designed NGS panel.

doi:10.6342/NTU201704070

Result

The average coverage of this NGS panel was about 174.9x; 95.7% of the nucleotides have read depth >30x. The targeted NGS panel was able to identify pathogenic variants in 40% of the patients (16/40). Among the diseases, muscular dystrophy panel has the highest diagnostic rate 47.6% (10/21), followed by CMT panel 33.3% (2/6) and OI/EDS 33.3% (3/9).

Conclusion

NGS panel could indeed help to molecularly diagnosed and confirmed the diseases and their subclasses. After validation, we would transfer the targeted NGS panel for clinical service in hope to increase diagnosis rate and provide better medical services.

Keyword: Next-generation sequencing, Gene panel, Limb-girdle muscular dystrophy, Charcot-Marie Tooth, Neuromuscular disease

doi:10.6342/NTU201704070

TABLE OF CONTENT

doi:10.6342/NTU201704070

LIST OF TABLE

doi:10.6342/NTU201704070

LIST OF FIGURE

doi:10.6342/NTU201704070

CHAPTER 1. INTRODUCTION

Molecular Diagnosis of Genetic Diseases

Molecular diagnosis has become increasingly complicated. In the old days, genetic testing relied mainly on biochemical assays to determine the presence of gene products (Nowakowski 1992). With the advance of technology, tests could be performed at DNA level to help determine the molecular basis of diseases and provide better healthcare management, diseases prognosis, and family planning. Genetic tests have been applied for many purposes, such as new born screening, carrier testing, prenatal diagnosis, diagnostic testing, pre-symptomatic testing, and pharmacogenomics. Different types of mutation call for different testing methods: 1. Biochemical testing usually involves examining the function of the proteins. A change in any part of the protein could result in disruption of the function and cause diseases. Some of the measurement that could be done are protein activity, level of metabolites used or produced, and quantity of the protein (Nowakowski 1992). 2. Cytogenetics testing studies abnormalities in chromosome. Each individual chromosome could be identified after staining. This could be performed under microscope or use array comparative genomic hybridization (aCGH) to do a more detailed chromosomal study at higher resolution. 3. Direct DNA testing could be the most effective method to determine whether DNA has mutation, insertion, deletion, or duplication. Some of the techniques that have been developed to detect such variation are MLPA, Sanger sequencing, and next generation sequencing (Zhang 2014). With direct genetic testing, laboratory generally looks for the particular genetic variants that contribute to the symptoms. A list of methods applicable to perform genotyping and karyotyping is listed Table 1 (Katsanis 2015).

The Advantage of Next Generation Sequencing on Genetic Disease

Sanger sequencing has been regarded as the golden standard for diagnosis of genetic diseases in the past decades (Sanger 1977). It is an accurate and effective method to identify mutations, especially for single gene disorder. However, with the development of technology, more and more disorders are found to exhibit clinical and genetic heterogeneity. If Sanger sequencing were to be used to diagnose such disorders, the total cost would be high and it would often take a long time to go through all the probable causative genes, resulting in high total-cost and anxiety for both patients and their families

doi:10.6342/NTU201704070

(Craigen, Graham et al. 2013). Therefore, a high throughput large scale sequencing method is required.

Due to the fast development of variety of sequencing technologies, next generation sequencing (NGS) has made analysis of numerous genomic regions possible (Rogers 2005, Boycott 2013, Boycott 2014). Hundreds of genes, even the whole genome, could be surveyed at the same time. The ability to sequence multiple genes in multiple samples simultaneously allows NGS to be ideally suited for addressing the limitations of traditional sequencing technology. In fact, NGS has been applied successively to different research studies to identify new disease genes (Lin 2012, Nakagawa 2013, Jiang 2014, de Koning 2015, Nakagawa 2015). And it is the goal of many molecular laboratories to put this technique to clinical use.

As mentioned earlier, disorders with phenotypic and genetic heterogeneity are more complicated to diagnose clinically (de Koning 2015). Since pinpointing possible disease causing genes simply based on the clinical features is difficult, a more comprehensive and cost-effective molecular diagnostic method is required. For genes having more than 20 exons and diseases with genetic heterogeneity, mutation analysis with NGS panel is the most effective way. Furthermore, to make NGS more appealing to clinical practice, samples are pooled, which not only shortens waiting time, but also lowers the cost per experiment to an acceptable range.

Introduction of NGS into clinical practice has shown a broad range of usage in molecular diagnosis of various human genetic disorders, from single gene disorders to disorders caused by groups of related genes. NGS panels have been applied to multi-gene disorders, such as cardiomyopathy, deafness, mitochondrial disorders, movement disorders, neuromuscular disorders, and retinitis pigmentosa (Craigen, Graham et al. 2013, Danielsson 2014, Ankala 2015, Atik 2015, de Koning 2015, Mercimek-Mahmutoglu 2015)

Disadvantage of Next Generation Sequencing for Clinical Application

Yet, despite the benefit of high throughput NGS, there are still some disadvantages. Millions of reads are generated in one single experiment. They contain large amount of genetic information that requires extensive bioinformatics to extract out meaningful data,

doi:10.6342/NTU201704070

that is, disease-causing mutations (de Koning 2015, Zhang 2014). The efficiency of analysis would increase if only the genes of interest are captured or enriched. Moreover, one other disadvantage of NGS is the cost of the experiment. Fortunately, the cost of NGS per base pair has decreased by 10-fold over the years, with improved accuracy and speed. Hence, the price per test is lower than before. Even so, an adequate number of samples must be obtained to achieve full cost advantage (Yohe 2015). Since inherited diseases are rare already, obtaining sufficient test samples for the same disease is even more challenging. This problem could be solved by pooling samples from different diseases on a single platform. Pooling samples increases sample size, thus, reduces the cost. This greatly increases the possibility of utilizing NGS for clinical diagnosis.

The Multiplex Targeted Gene Panel

In this study, we multiplexed patient samples and different disease categories on a single platform, thus having adequate number of test samples per experiment. All related genes would be captured and sequenced but only those we are interested in would be analyzed. This strategy reduces costs and decreases turnaround time. Exome-captured libraries of more than 400 genes with various disease conditions were designed, focusing on limb-girdle muscular dystrophy (LGMD), Charcot-Marie Tooth disease (CMT), neurometabolic diseases (NMD), osteogenesis imperfecta (OI) and Ehlers-Danlos syndrome (EDS). We evaluated the coverage, accuracy, and clinical validity of NGS. Our goal is to develop this NGS panel that would be applicable to clinical utilization for helping clinical practitioners molecularly diagnose the diseases and care for patients.

Limb-Girdle Muscular Dystrophy

Limb-girdle muscular dystrophy (LGMD) is a genetically and clinically heterogeneous group of muscular dystrophies (Nigro 2014). Individuals with LGMD generally show weakness and wasting of the muscles in the arms and legs, and gradually worsen over time (Nigro et al. 2014). The most affected muscles are usually the proximal muscles (Pegoraro 2016). The age of onset, severity, and features vary among different subtypes of LGMD, sometimes even within the same family. The various types of LGMD are caused by mutations in different genes. The list of genes responsible for LGMD has grown in the last 10 years from 16 to 31 loci, which is far too much to screen each gene one-by-one (Nigro 2014). The NGS approach could save both time, cost, and labor.

doi:10.6342/NTU201704070

Charcot-Marie-Tooth Disease

Charcot-Marie-Tooth disease (CMT) is the most common inherited neurological disorder that affects both motor and sensory nerves (Swinney 2014). A typical feature includes weakness of lower leg muscles and foot; sometimes foot deformities may occur (Bird 1993). Progression of symptoms are gradual, but CMT is not a fatal disease and patients have a normal life expectancy. CMT is caused by mutations in many different genes. Until now, more than 75 genes have been identified to result in CMT phenotypes and new genes are still being identified, making molecular diagnosis of CMT even harder (Ø stern 2013, Tazir 2014, Ekins 2015). Therefore, a panel with multiple genes is necessary for clinical practice.

Neurometabolic Diseases

Neurometabolic diseases (NMDs) are a group of rare diseases with problems in both metabolism and brain function (Karimzadeh 2015). They are complex disorders with many causes, which are poorly understood. NMDs cover wide range of symptoms, including lysosomal storage disorders, leukodystrophies, neurotransmitter disorders, urea cycle disorders, peroxisomal disorders, aminoacidopathies, and other disorders causes mental retardation and epilepsy (Karimzadeh 2015). Patients with NMDs show neurological sign and symptoms, such as developmental delay, seizures, hypotension, and loss of consciousness. The diagnosis can be performed for some diseases using tandem mass spectrometry for metabolite analysis, plasma amino acid analysis, urine organic acid analysis, or other specific tests, like lysosomal enzyme assay (Szymańska 2014). For the rest of the disorders, the diagnosis is extremely difficult, especially in Taiwan. In the NGS panel for this study, we enclosed genes involved in basal ganglion calcification, leukodystrophy, and dystonia. Clinical implications of these genes are examined in hopes of helping diagnosing the patients.

Osteogenesis Imperfecta and Ehlers-Danlos Syndrome

Ehlers-Danlos syndrome (EDS) is a heritable connective tissue disorder characterized by joint laxity, cutaneous hyperextensibility, and abnormal wound healing. Other minor diagnostic criteria were established as well, for example, tissue extensibility and fragility, and complications of joint hypermobility. Various forms of EDS have been described and they are caused by mutations in different genes. The clinical presentations often overlap within subtypes. In fact, some of the phenotypes of EDS even overlap with another

doi:10.6342/NTU201704070

connective tissue disorder, osteogenesis imperfecta (OI), which is characterized by joint hypermobility, skin hyperexentibility, and easy bruising; other clinical manifestations of OI include blue sclera, short stature, and various degree of bone fragility. Like EDS, OI has several clinical subtypes based on the clinical, biochemical, and molecular nature of each type. Thus, a panel with multiple genes is beneficial in establishing diagnosis.

doi:10.6342/NTU201704070

CHAPTER 2. MATERIALS AND METHODS

Patient Selection and Evaluation

Patients were eligible to enroll in the study if they were clinically diagnosed with muscular dystrophy, CMT, NMDs, or OI/EDS phenotype, without age limitation; those who were selected to participate in the study were de-identified. We have 40 patients in total, collected in the period of January 2015 to December 2015, (patients A to AN): 21 with muscular dystrophy, 9 with OI/EDS, 2 with renal diseases, 6 with CMT, and 2 with NMDs. Among the muscular dystrophy patients, 13 were suspected of LGMD. Two muscular dystrophy/LGMD patients (patient I and P) had pathogenic mutations identified previously; muscular dystrophy patient G has a family history with known mutation. Patients were informed that participation would consist of donating approximately 3mL of blood for genetic testing. The study was approved by Institutional Review Board (IRB) and informed consent was obtained from all individuals (NTUH IRB No. 201505135RINA).

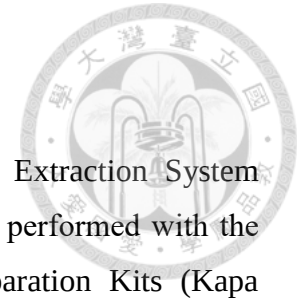
NGS Panel Gene List

Genes were selected through review of published literatures. All 195 genes known to be implicated in limb-girdle muscular dystrophy (LGMD) (39 genes), Charcot-Marie Tooth disease (CMT) (60 genes), neuromuscular diseases (NMDs) (43genes), osteogenesis imperfecta (OI) and Ehlers-Danlos syndrome (EDS) (54 genes) were targeted for enrichment. All genes were considered due to genetic heterogeneity in all diseases. Some genes are related with more than one disease, for example, *LMNA* gene is associated with LGMD and CMT. Genes enclosed in the 4 disease groups are listed in Table 2. The NGS gene panel itself contains an additional 213 genes that were indicated in other inherited diseases as well.

Library Design

For DNA capture, we designed a custom NimbleGen SeqCap EZ Choice Library to capture all exons and all exon-intron boundaries including at least 50 intronic nuecleotides. DNA bait selection follows the most stringent settings for probe design using NimbleDesign (http://www.nimblegen.com/products/nimbledesign/index.html). The uniqueness is tested with Sequence Search and Alignment using Hashing Algorithm (SSAHA).

doi:10.6342/NTU201704070



Sample Preparation

DNA was extracted from blood samples using Puregene DNA Extraction System according to manufacturer's manual (Qiagen). Genomic capture is performed with the libraries prepared using KAPA Hyper Prep DNA Library Preparation Kits (Kapa Biosystems) per the manufacturer's manual (NimbleGen). In short, a minimum of 1μg of genomic DNA is fragmented with Covaris S2 to obtain 3' and 5' overhangs that are about 200-300bp long, followed by end repair, and A-tailing. The libraries are indexed and amplified with PCR. Ten samples are pooled per reaction. The captured libraries are sequenced with Illumina MiSeq System (Illumina).

Data Output and Classification of Variants

The bioinformatics analysis pipeline is described in Figure 1. Resulting MiSeq paired-end reads were aligned to human reference genome hg19 (GrCh37) and variant calling was done using MiSeq Reporter (Illumina). Subsequent variant annotation was performed by wANNOVAR and Variant Studio. Variants were then filtered out with any of the following criteria met: i) quality score <30, ii) read depth <30, iii) not pass filter, iv) allele frequency >0.05 in population database, such as Exome Aggregation Consortium (ExAC), 1000 Genomes Project, ESP6500si, and dbSNP. Variants located in intron, 3' or 5' UTR, upstream or downstream, or non-coding regions, or those that would cause synonymous mutations and inframe indels were also not included, as majority of variations occurred in these regions have insufficient evidence. Remaining variants were prioritized according to the patient's clinical diagnosis and phenotypes. The probable mutations were evaluated in the order of the following consequences: initiation codon, stop loss/gain, splice site, frameshift, and missense. Candidate variants were classified into likely pathogenic (LP) / pathogenic (P), likely benign (LB) / benign (B), or variant of unknown significance (VUS) according to ACMG guideline (Richards 2015). We looked at databases like Human Gene Mutation Database (HGMD) and Online Mendelian Inheritance in Man (OMIM), the effect of each change on the protein and splice site, and different scores like nucleotide and amino acid conservation scores, SIFT, and PolyPhen v2. If no variant was found to be associated with the diseases, criteria ii) and iii) were alleviated. Assessment of the quality of sequencing and variants were done by Integrative Genomics Viewer (IGV). The coverage of exons in each gene is calculated. Those that did not have adequate coverage were highlighted.

doi:10.6342/NTU201704070

Verification by Sanger Sequencing

Sanger sequencing is carried out to verify the variants detected by NGS. Specific primers are designed using Primer3 v.0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/). They were purchased from Bio-Protech. Genomic DNA was amplified by touchdown PCR method using GeneAmp PCR System (Applied Biosystems). For PCR products with GC content exceeding 59%, 5% of DMSO was added to the mixture. Thermal cycles were performed with 1 cycle of 94°C for 5 min; 14 cycles of touchdown PCR with 94°C for 30 sec, touchdown annealing temperature for 30 sec, 72°C for 30 sec; 25 cycles of 94°C for 30 sec, 53°C for 30 se, 72°C for 30 sec; finally, 1 cycle at 72°C for 5 min. The annealing temperature of the touchdown cycles started at 60°C initially, and was decreased by 0.5°C each cycle, ending at 53°C. Agarose gel electrophoresis was used to confirm PCR products, and all primer sets produced a single band. The PCR products were purified with Gel/PCR DNA Fragments Extraction Kit (Geneaid), followed by direct sequencing using ABI Prism Big Dye Dideoxy Chain Terminator Cycle Sequencing Kit and ABI Prism 3100 genetic analyzer (Applied Biosystem).

doi:10.6342/NTU201704070

CHAPTER 3. RESULTS

Targeted NGS panel sequencing results

Targeted sequencing of 408 genes of various inherited diseases were performed on 40 unrelated individuals with different inherited diseases. On average, 81.4% of targeted regions were enriched, ranging from 76.8% to 85.6%. After aligning with hg19 reference genome, mean coverage of the overall region was 174.9X, from lowest 111.1X to highest 276.1X. More than 95% of the targeted regions were covered. At least 95.7% of nucleotide was covered at least 30X, while the lowest coverage was 91.3% and the highest coverage was 98%. The overall performance and coverage of the panel, as well as sequence capture performance per gene for LGMD, CMT, NMD, and OI/EDS, are shown in Table 3 and Table 4, respectively. Most exons with low coverage were similar between patients and coverage decreased with increasing GC content.

Variants identification

Variants were identified based on the bioinformatics analysis encompassing sequence mapping, variant calling, filtering, and variant classification. On average, 3057 variants were detected (Table 3). We found a mean of 2269 SNVs and 185 indels, and 180 were not found in dbSNP. After filtering of 3057 variants, a mean count of 20 variants affecting the splice sites or predicted to change the amino acid sequence were found in the 195 genes. These variants were then prioritized; those that fitted the diseases based on the clinical phenotypes and diagnosis were classified by ACMG guideline (Richards 2015). For patients with phenotypes that matched to several diseases, all genes fitting the different diseases were assessed. For example, patient D and F were diagnosed with muscular dystrophy and CMT; thus, all genes related to muscular dystrophy and CMT were studied.

Patient Testing Result

Results from de-identified patients received in the study in the period of January 2015 to December 2015 for testing using the designed NGS panel were collected. The patients enrolled are detailed in Table 5. Overall, we performed NGS tests on 40 patients. Six patients were suspected of CMT, 2 of which also have symptoms of muscular dystrophy. Of the 21 patients with muscular dystrophy, 2 already had a molecular diagnosis prior to

doi:10.6342/NTU201704070

entering the study. Both patient I and P carry mutations on *DYSF* gene, [c.937+1 G>A / c.3207G>A, p.Trp1069Ter] and [c.863A>T, p.Asp288Val / c.1667T>C, p.Leu556Pro], respectively. Patient G, on the other hand, have a family history, in which his maternal uncle and cousins have similar symptoms and they have molecular diagnosis of *DMD* gene, [c.2169-1G>T]. As both OI and EDS are of collagen defect and would have weaken joint and bone, the genes related to OI and EDS are placed in the same panel. Patient AB and AC were clinically diagnosed with EDS; patient AD, AE, AG, and AH have OI. Patient AF was suspected of OI as well; however, prior to joining the study, screening of collagen 1 and 2 gene were all negative. Two cases with Alpers syndrome and GM1 gangliosidosis, patient AM and AN, respectively, were categorized as NMD as they had some degree of neurological symptoms.

The disease causing mutations that would match the clinical diagnosis could be identified in 40% of our patients (16 out of 40). A detailed list of pathogenic mutations is summarized in Table 6 and the sequencing data of the mutations is depicted in Figure 2. Although the clinical information is important to define the subclass of diseases, we did not prioritize genes according to the subclasses of each disease based on more detailed clinical data, instead all genes associated with the diseases were classified. Patient B was suspected of CMT axonal type. Rather than just looking for the genes related to axonal type CMT, all genes associated with CMT were checked. In total, NGS identified 23 P/LP variants in 40 patients. Each variant was either a known pathogenic variant or its pathogenicity could be confirmed by segregation analysis or categorized by ACMG guideline. The patients in our cohort were affected by diverse diseases of different inheritance, autosomal recessive, autosomal dominant, and X-linked. About 50% of patients (8 out of 16) carried one single heterozygous mutation with autosomal dominant inheritance pattern. Six patients (37.5%) were identified to have diseases of autosomal recessive inheritance: 5 patients had compound heterozygous mutations and 1 patient had a homozygous mutation. About twelve percent of patients (2 out of 16) carried variants with X-linked inheritance in *DMD* gene. The identified variants were all in agreement with the clinical diagnosis of CMT, muscular dystrophy, LGMD, and OI/EDS.

Variant confirmation

Probable disease-causing mutations were identified in several patients. Sanger

doi:10.6342/NTU201704070

sequencing and segregation analysis were used to validate the presence of mutations in the original DNA, as well as confirm the association between genotypes and phenotypes. Segregation analysis and familial studies could only be performed with some patients, as DNA of the relatives of other patients were not available.

For CMT patients, one pathogenic variant was found in patient C and F in *MFN* gene, [c.839G>A, p.Arg280His] and [c.775C>T, p.Arg259Cys], respectively. Both variants were of autosomal dominant and were classified as P/LP variants. On the other hand, 14 variants were identified in 21 muscular dystrophy patients. Before participating in the study, patient I and P were molecularly diagnosed as LGMD 2B previously. Both patient I and P carry mutations on *DYSF* gene, [c.937+1 G>A] / [c.3207G>A, p.Trp1069Ter] and [c.863A>T, p.Asp288Val] / [c.1667T>C, p.Leu556Pro], respectively. Our NGS panel could identify the same mutations found in these two patients. *DYSF* [c.1667T>C, p.Leu556Pro] has unknown pathogenicity, while the others were P/LP variant. Patient Q was found to carry compound heterozygous mutations in *CAPN3* gene, [c.1309C>G, p.Arg437Gly] / [c.2092C>T, p.Arg698Cys]. Patient J had homozygous mutation in *TRAPPC11* gene, [c.2938G>A, p.Gly980Arg], which is associated with LGMD 2S. Sanger sequencing on J's parents revealed that both parents are carriers of the mutation, thus, validating the autosomal recessive inheritance pattern.

In addition, patient N, O, U, and V had one pathogenic variant in *MYH7*, *LMNA*, *HNRNPDL* (*HNRPDL*), and *CSF1R*, respectively. All variants were of autosomal dominant inheritance pattern. Patient N carried [c.1322C>T, p.Thr441Met] mutation in *MYH7* gene, which has been associated with distal and cardiomyopathy. Patient O with LGMD was classified as LGMD 1B by the genetic mutation in *LMNA* gene, [c.148C>G, p.Arg50Gly]. Family history was available for both patient U and V. Samples of several family members of patient U were available for segregation analysis. NGS revealed that patient U carried a mutation associated with LGMD 1G, *HNRNPDL* (*HNRPDL*) [c.1132 G>A, p.Asp378A]. Confirmation by Sanger sequencing on patient U, as well as on several family members showed that affected individuals all carried the mutation, while the asymptomatic individual did not have the mutation. For patient V, she carried *CSF1R* [c.2330G>A, p.Arg777Gln]. Sanger sequencing of parents of V revealed the same variant in *CSF1R* gene was only found in her father, whom was suspected to be affected as well.

doi:10.6342/NTU201704070

Three pathogenic variants in X-linked *DMD* gene were found in patient G and S, 1 in patient G (male) and 2 in patient S (female). Patient G was clinically diagnosed as Becker muscular dystrophy judging from his family history. His maternal uncle and cousins have similar symptoms, and was molecularly diagnosed with *DMD*, [c.2169-1G>T]. The NGS gene panel was able to identify the same *DMD* splice site variant in patient G, confirming that he, too, carried the mutation. Patient S was found to have 2 mutations in *DMD* gene, [c.2956C>T, p.Gln986Ter] and [c.2136G>C, p.Arg712Ser]. The former variant was classified as pathogenic using ACMG guideline; though latter has not been reported in the literature before, it was categorized as pathogenic by ACMG guideline.

As for patients diagnosed with OI/EDS, 3 P/LP variants and 1 VUS were detected. Patient AB with Ehlers-Danlos syndrome was found to carry a missense variant of uncertain significance, *COL5A1* [c.1345C>T, p.Arg449Trp]. Confirmation of the variant with Sanger sequencing revealed that AB's mother carried the same mutation in *COL5A1* gene without any symptoms. Patient AG, who was diagnosed with OI, had compound heterozygous variants for *WNT1* gene, [c.104+1G>A] / [c.105G>A p.Trp35Ter]. The result of Sanger sequencing for the parents revealed that each parent carried one pathogenic variant, thus, validating the autosomal recessive inheritance pattern. *WNT1* is classified as typeXV OI. Patient AH, also diagnosed with OI, had a pathogenic variant in *COL1A1* [c.661G>C, p.Gly221Arg], which is categorized as typeI OI. A pathogenic variant, [c.1781G>A, p.Arg594His] was found in *TRPV4* gene in patient AI with spondyloepiphyseal dysplasia. The result of Sanger sequencing showed that no parent of patient AM carried the same mutation in *TRPV4* gene. For patient AK, he was diagnosed with Bartter syndrome; and three P/LP variants in *SLC12A3* gene were found. The 3 variants identified in patient AI in *SLC12A3* gene was later confirmed with his parents using Sanger sequencing. Two variants [c.488C>T, p.Thr163Met] and [c.2612G>A, p.Arg871His] were from the father, and [c.734T>C, p.Leu245Pro] from the mother.

False positive

The presence of false positive variant, that is, probable disease-causing variant not present in the original DNA of patients, was examined as well. The probability of being false positive due to sequencing or mapping errors was about 1 out of the 24 variants tested, when the filters were shown strand bias (SB) and IGV revealed high SNPs around the variant. In the case of patient R, who was clinically diagnosed with LGMD, a probable

doi:10.6342/NTU201704070

disease causing variant was found in FLNC gene, [c.2389+2T>G]. The variant was proven false when confirmed by Sanger sequencing. No false positive was found when the filter PASS in QC was applied.

doi:10.6342/NTU201704070

CHAPTER 4. DISCUSSION

In this study, targeted massive parallel sequencing was performed in a cohort of patients with heterogeneous diseases. The NGS platform can detect several types of mutations including intronic and exonic changes. All mutations were obtained from sequencing and analysis of a single proband.

The overall coverage of our targeted NGS panel was about 174.9X with more than 95% of target region was covered. 95% of the targeted region was sequenced more than 30x. Compared with the previously published NGS panel regarding NMD, LGMD, OI/EDS, and CMT, which on average is about 471.5X (Høyer 2014, Sule 2013, Vesli 2012, Yu 2017), our coverage is about 2.5X lower. Low sequencing coverage could be due to insufficient production of reads during sequencing run. If more reads were generated per sequencing run, the coverage would increase. Alternatively, inability of reads mapping and alignment back to the target region could reduce coverage. Reads not aligning back to the target sequences on the reference could be due to several reasons: sequencing errors, the degree of mutations in the genome being sequenced, and the uniqueness of the target region. All of these would affect the accuracy of read alignment. Although our coverage is not as high as the others, our detection rate is comparable with the published NGS gene panels.

In our cohort of 40 patients, a genetic diagnosis was identified in 16 patients (40%) using our targeted NGS panel. When broken down to different diseases, detection rate differed for each disease. For our CMT portion of the panel, detection rate of is about 33.3% (2/6). Both patient C and F are clinically diagnosed as CMT2A. CMT is a clinically and genetically heterogeneous group of diseases with over 75 genes associated with the disease. Most mutations occurred in PMP22, GJB1, MFN2, and MPZ genes, which accounted for 92% of CMT patients (Høyer 2014); mutations in other genes are rare causes of CMT. Molecular diagnostic rate of CMT has been reported to be around 45.6% (37/81) (Høyer 2014). Yet, among the 37 CMT patients enrolled by Høyer (2014), 25 carried point mutation, while 12 carried duplication in *PMP22* and *MPZ*. If such deletion was in our group of patients, it might be missed by the NGS panel.

doi:10.6342/NTU201704070

For muscular panel, 47.6% (10/21) of muscular dystrophy patients could be molecularly diagnosed (patient G, I, J, N, O, P, Q, S, U, and V). Among the muscular dystrophy, 13 were clinically suspected of LGMD. Within this LGMD panel, diagnostic rate was about 61.5% (8/13), although the mutations in patient S was found in *DMD* gene. Patient S showed two missense changes in DMD gene, [c.2956C>T, p.Gln986Ter] and [c.2136G>C, p.Arg712Ser]. [c.2956C>T, p.Gln986Ter] is categorized as pathogenic in database and by ACMG guideline. [c.2136G>C, p.Arg712Ser], on the other hand, has not been reported yet. Even so, ACMG guideline classified the mutation as pathogenic. It's likely that both changes are either on the same allele or different alleles. Parental DNAs are required to investigate this further. Several literatures have evaluated the application of NGS on LGMD patients, and the average positive identification rate was about 30%, ranging from 16% to 68% (Ankala 2015, Ghaoui 2015, Seong 2015, Kitamura 2016, Kuhn 2016, Yu 2017). Our LGMD panel has a relatively better diagnostic rate than average.

Our OI/EDS panel has a detection rate of 33.3% (3/9), compared to the literature, which is about 33% (Weerakkody 2016). In addition to single variants, insertions, deletions, and duplication of collagen related genes, Weerakkody also identified rearrangements between *COL3A1* and *COL5A2*, as the cause of disease. Fifty percent (1/2) of patients with renal diseases would be molecularly diagnosed. No pathogenic variants could be could be identified in patients with NMD (0/2). However, in literature, the use of NMD panels could identify pathogenic mutations in about 30% of NMD patients on average (Dhamija 2015, Davis 2016), ranging from 16-60%.

One possibility that no variants could be found in the remaining 24 patients, could be due to the poor coverage of target regions. The coverage differs not only between different genes, but between exons within the same gene as well. High GC-content in the target region, poor alignment of the reads, and pseudogenes are some of the factors that would lead to reduction in coverage. Additionally, one of the limitation of NGS is the detection of large deletion, duplication, insertion, inversion or complex structural variation, and tissue-specific or low percentage of mosaicism. For instance, if one of the patients have large deletion in *PMP22* gene, NGS might not be able to detect the change. Another reason could be that the target regions were missed since no capture primers were designed. What's more, disease-causing mutations might be located in intron and non-

doi:10.6342/NTU201704070

coding protein regions which were not considered in the study. Alternatively, the disease-causing mutations might be in a novel gene not linked to the disease at this point.

According to our study, targeted NGS is more cost- and time- efficient than the conventional strategies, such as Sanger sequencing, in investigating suspected targeted genes one by one. Due to the overlapping of clinical and pathological phenotypes, it is quite challenging to choose the appropriate target gene, even for experienced specialists. Under such conditions, targeted NGS shows its advantages compared to Sanger sequencing. The NGS gene panel for inherited diseases could be applied in clinical use in hope to help confirm clinical diagnosis or establish diagnosis when clinical presentation is ambiguous. However, the study indicates the high demand to re-evaluate patients with unclassified diseases. One major challenge during NGS data analysis is the identification of the pathogenic change among the large list of variants. The analysis based on diseases and variant categorization allowed the identification of all known mutations. Detailed clinical and molecular data are necessary for confirmation and matching genetic data with phenotypes.

Though we planned to obtain all disease-related genes, new genes are still being identified. New genes should be regularly updated with the most current research to the capture library. On top of that, in this study, only exome was captured. Some of the studies suggest that mutations in intronic regions could cause diseases, too. The new version of NGS panel, thus, should contain these regions as well. For poorly coverage regions, other methods, whether it be add-on PCR or new capture design, should be applied to make the panel more reliable. To use NGS in the clinical diagnosis, the interpretation of the enormous amount of genomic data and the attempt to correlate genotype and phenotype is one of the pains that come with this developing technology. The assessment of each variant requires extensive research and careful validation. The development of targeted NGS gene panel, together with the combination with clinical evaluation, could help improve the molecular diagnosis success rate in inherited diseases, and decrease the cost of diagnostic tests. Furthermore, confirmation of genetic diagnosis could help improve the quality of the medical care for the patients.

doi:10.6342/NTU201704070

CHAPTER 5. REFERENCE

Ankala, A., da Silva, C., Gualandi, F., Ferlini, A., Bean, L.J., Collins, C., Tanner, A.K., & Hegde, M.R. (2015). "A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield." Annals of Neurology **77**(2): 206-214.

Atik, T., Bademci, G., Diaz-Horta, O., Blanton, S.H., & Tekin, M. (2015). "Whole-exome sequencing and its impact in hereditary hearing loss." Genetics Research **97**: e4.

Boycott, K. M., Dyment, D.A., Sawyer, S.L., Vanstone, M.R., & Beaulieu, C.L. (2014). "Identification of genes for childhood heritable diseases." Annual Review of Medicine **65**: 19-31.

Boycott, K. M., Vanstone, M.R., Bulman, D.E., MacKenzie, A.E. (2013). "Rare-disease genetics in the era of next-generation sequencing: discovery to translation." Nature Reviews Genetics **14**(10): 681-691.

Craigen, W. J., et al. (2013). "Exome sequencing of a patient with suspected mitochondrial disease reveals a likely multigenic etiology." BMC Med Genet **14**: 83.

Danielsson, K., Mun, L.J., Lordemann, A., Mao, J., & Lin, C.H. (2014). "Next-generation sequencing applied to rare diseases genomics." Expert Review of Molecular Diagnostics **14**(4): 469-487.

Davis, M., Allcock, R., and Laing, N. (2016). "Next generation sequencing for neuromuscular disease in a diagnostic setting - The Perth custom neuromuscular gene panel 3 year on." Neuromuscular Disorders 26(2016):S88

de Koning, T. J., Jongbloed, J.D., Sikkema-Raddatz, B., Sinke, R.J. (2015). "Targeted next-generation sequencing panels for monogenetic disorders in clinical diagnostics: the opportunities and challenges." Expert Review of Molecular Diagnostics **15**(1): 61-70.

Ekins, S., Litterman, N.K., Arnold, R.J., Burgess, R.W., Freundlich, J.S., Gray, S.J., Higgins, J.J., Langley, B., Willis, D.E., Notterpek, L., Pleasure, D., Sereda, M.W., & Moore, A. (2015). "A brief review of recent Charcot-Marie-Tooth research and priorities." F1000Research **4**: 53.

Høyer, H., Braathen, G.J., Busk, Ø.L., Holla, Ø.L., Svendsen M., Hilmarsen, H.T., Strand, L., Skjelbred, C.F., and Russell, M.B. (2014). " Genetic Diagnosis of Charcot-Marie-Tooth Disease in a Population by Next-Generation Sequencing." BioMed Research International 2014 (2014)

doi:10.6342/NTU201704070

Jiang, T., Tan, M.S., Tan, L., & Yu, J.T. (2014). "Application of next-generation sequencing technologies in Neurology." Annals of Translational Medicine **2**(12): 125.

Karimzadeh, P. (2015). "Approach to neurometabolic diseases from a pediatric neurological point of view." Iranian Journal of Child Neurology **9**(1): 1-16.

Katsanis, S.H. and Katsanis, N. (2015). "Molecular genetic testing and the future of clinical genomics." Nat Rev Genet 14(6):415

Lin, X., Tang, W., Ahmad, S., Lu, J., Colby, C.C., Zhu, J., & Yu, Q. (2012). "Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities." Hearing Research **288**(1-2): 67-76.

Mercimek-Mahmutoglu, S., Patel, J, Cordeiro, D., Hewson, S., Callen, D., Donner, E.J., Hahn, C.D., Kannu, P., Kobayashi, J., Minassian, B.A., Moharir, M., Siriwardena, K., Weiss, S.K., Weksberg, R., & Snead, O.C. (2015). "Diagnostic yield of genetic testing in epileptic encephalopathy in childhood." Epilepsia **56**(5): 707-716.

Nakagawa, H., Wardell, C.P., Furuta, M., Taniguchi, H., & Fujimoto, A. (2015). "Cancer whole-genome sequencing: present and future." Oncogene **34**(49): 5943-5950.

Nakagawa, H. S., T. (2013). "Comprehensive genome sequencing of the liver cancer genome." Cancer Letters **340(2):234-240.**(2): 234-240.

Nigro, V. S., M. (2014). "Genetic basis of limb-girdle muscular dystrophies: the 2014 update." Acta Myologica **33**(1): 1-12.

Nowakowski, R. and Thompson, J.H. (1992). "Laboratory diagnosis of genetic disorders" Optom Clin 2(1):87

Østern, R., Fagerheim, T., Hjellnes, H., Nygård, B., Mellgren, S.I., & Nilssen, Ø. (2013). "Diagnostic laboratory testing for Charcot Marie Tooth disease (CMT): the spectrum of gene defects in Norwegian patients with CMT and its implications for future genetic test strategies." BMC Medical Geneticsc **14**: 94.

Pegoraro, E. H., E. P. (2016). "Limb-Girdle Muscular Dystrophy Overview." GeneReviews(R).

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., and Rehm, H.L. (2013). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." Genetics in Medicine **17**:5: 405

doi:10.6342/NTU201704070

Rogers, Y. H., & Venter, J. C. (2005). "Genomics: massively parallel sequencing." Nature Reviews Genetics **437**(7057): 326-327.

Sanger, F., Nicklen, S., & Coulson, A.R. (1977). "DNA sequencing with chain-terminating inhibitors" Proceedings of the National Academy Sciences USA **74**(12): 5463-5467.

Sule, G., Campeau, P.M., Zhang, V.W., Nagamani, S.C.S., Dawson, B.C., Grover, M., Bacino, C.A., Sutton, V.R., Brunetti-Pierri, N., Lu, J.T., Lemire, E., Gibbs, R.A., Cohn, D.H., Cui, H., Wong, L.J.C., & Lee, B.H. (2013). "Next-generation sequencing for disorders of low and high bone mineral density. " Osteoporosis Internationa. **24**(8):2253

Swinney, D. C. X., S. (2014). "The discovery of medicines for rare diseases." Future Medicinal Chemistry **6**(9): 987-1002.

Szymańska, K., Szczałuba, K., Lugowska, A., Obersztyn, E., Radkowski, M., Nowakowska, B.A., Kuśmierska, K., Tryfon, J., & Demkow, U. (2014). "The analysis of genetic aberrations in children with inherited neurometabolic and neurodevelopmental disorders." BioMed Research International **2014**: 424796.

Tazir, M., Hamadouche, T., Nouioua, S., Mathis, S., Vallat, J.M. (2014). "Hereditary motor and sensory neuropathies or Charcot-Marie-Tooth diseases: an update." Journal of the Neurological Sciences **347**(1-2): 14-22.

Vasli,.N., Böhm, J., Le Gras, S., Muller, J., Pizot, C., Jost, B., Echaniz-Laguna, A., Laugel, V., Tranchant, C., Bernard, R., Plewniak, F., Vicaire, S., Levy, N., Chelly, J., Mandel, J.L., Biancalana, V., & Laporte, J. (2012). " Next generation sequencing for molecular diagnosis of neuromuscular diseases." Acta Neuropathol. **124**(2):273

Weerakkody, R.A., Vandrovcova, J., Kanonidou, C., Mueller, M., Gampawar, P., Ibrahim, Y., Norsworthy, P., Biggs, J., Abdullah, A., Ross, D., Black, H.A., Ferguson, D., Cheshire, N.J., Grahame, R., Ghali, N., Vandersteen, A., Pope, F.M., & Aitman, T.J. (2016). "Targeted next-generation sequencing makes new molecular diagnoses and expands genotype-phenotype relationship in Ehlers-Danlos syndrome." Genetics in Medicine. **18**(11):1119

Yohe, S., Hauge, A., Bunjer, K., Kemmer, T., Bower, M., Schomaker, M., Onsongo, G., Wilson, J., Erdmann, J., Zhou, Y., Deshpande, A., Spears. M.D., Beckman, K., Silverstein, K.A., & Thyagarajan, B. (2015). "Clinical validation of targeted next-generation sequencing for inherited disorders." Archives of Pathology & Laboratory Medicine **139**(2): 204-210.

Yu, M., Zheng, Y., Jin, S., Gang, Q., Wang, Q, Yu, P., Lv, H., Zhang, W., Yuan, Y., and Wang, Z. (2017). Mutational spectrum of Chinese LGMD patients by targeted next-

doi:10.6342/NTU201704070

generation sequencing. PLoS One. 2017; 12(4): e0175343

Zhang, W., Cui, H., Wong, L.J. (2014). "Application of next generation sequencing to molecular diagnosis of inherited diseases." Topics in Current Chemistry **336**: 19-45.

doi:10.6342/NTU201704070

Table 1. List of methodologies for molecular diagnosis

Method	Sensitivity	Specificity	Cost	Copy number variation	Repeat expansions	SNVs / indels	Translocation/ Inversion
FISH	High	High	Low	Yes			Yes
Array CGH	High	High	High	Yes			Yes
SNP microarray	Low	Low	Low	Yes			
Linkage analysis (STRs)	Low	Low	Low	Yes			
PCR	High	High	Low		Yes	Yes	
MLPA	High	High	Low	Yes	Yes		
Sanger sequencing	High	High	Low			Yes	
WES / WGS	High	High	High			Yes	

doi:10.6342/NTU201704070

Table 2. Genes included in the NGS panel for CMT, LGMD, NMD, and OI/EDS.

CMT						
AARS	AIFM1	ARHGEF10	BICD2	BSCL2	COX6A1	CTDP1
DCTN1	DHTKD1	DNAJB2	DNM2	DYNC1H1	EGR2	FBLN5
FGD4	FIG4	GAN	GARS	GDAP1	GJB1	GNB4
HARS	HINT1	HK1	HSPB1	HSPB3	HSPB8	IGHMBP2
KARS	KIF1B	KIF5A	LITAF	LMNA	LRSAM1	MARS
MED25	MFN2	MICAL1	MPZ	MTMR	NDRG1	NEFL
PDK3	PLEKHG5	PMP22	PRPS1	PRX	RAB7A	SBF1
SBF2	SH3TC2	SLC12A6	SPG11	SURF1	TDP1	TFG
TRIM2	TRPV4	TUBA8	YARS			

LGMD						
ANO5	CAPN3	CAV3	COL6A1	COL6A2	COL6A3	DAG1
DES	DNAJB6	DYSF	FKRP	FKTN	GAA	GMPPB
HNRNPDL	ISPD	ITGA7	LAMA2	LARGE	LIMS2	LMNA
MYOT	PLEC	PNPLA2	POMGNT1	POMK	POMT1	POMT2
SEPN1	SGCA	SGCB	SGCD	SGCG	SMCHD1	TCAP
TNPO3	TRAPPC11	TRIM32	TTN			

NMD						
ACTB	ALDH7A1	ANO3	ARSA	ASPA	ATM	ATP1A3
CIZ1	COL4A1	COL4A2	DBH	DDC	FUCA1	GALC
GCDH	GCH1	GJC2	GLB1	GNAL	HEXA	LAMP2
LAMP3	PAH	PLA2G6	PNKD	PNPO	PRKRA	PRRT2
PSAP	PTS	QDPR	SGCE	SLC17A5	SLC18A2	SLC2A1
SLC6A3	SPR	SUMF1	TAF1	TH	THAP1	TOR1A
TUBB4A						

OI; EDS; Osteopetrosis Gene						
ACTA2	ADAMTS2	B3GALT6	B4GALT7	BMP1	CBS	CHST14
CLCN7	COL1A1	COL1A2	COL2A1	COL3A1	COL5A1	COL5A2
CREB3L1	CRTAP	DSE	FBN1	FBN2	FKBP10	FKBP14
FLNA	FLNB	IFITM5	IKBKG	LEPRE1	LRP5	MBTPS2
OSTM1	P4HB	PLOD1	PLOD2	PLS3	PPIB	RIN2
SEC24D	SERPINF1	SERPINH1	SLC39A13	SMAD3	SP7	SPARC
TCIRG1	TGFBR1	TGFBR2	TMEM38B	TNFRSF11A	TNFSF11	TNXB
TRAPPC2	TRPV4	WNT1	XYLT2	ZNF469		

doi:10.6342/NTU201704070

Table 3. Overall sequencing performance and coverage.

		Average	Standard deviation	Min	Max
Sequencing	Enrichment	81.4	2.05	76.8	85.6
Coverage	All regions	174.9	27.23	111.1	276.1
	>30x coverage	95.7	1.83	91.3	98
Variants	All regions	3057	661	2379	4133
	SNV+indel	2454	477	2025	3358
	SNV+indel (without rs number)	180	45	102	273
	All regions after filtering	115	32	41	171
	Candidates (nonsynonymous missense mutation, stop gain, splice-site, frameshift, initation codon)	20	5	10	35

doi:10.6342/NTU201704070

Table 4. The average coverage of genes in LGMD, CMT, NMD, and OI/EDS.

LGMD Gene	Average Coverage	CMT Gene	Average Coverage	NMD Gene	Average Coverage	OI; EDS; Osteopetrosis Gene	Average Coverage
ANO5	190.69	AARS	192.91	ACTB	61.59	ACTA2	170.52
CAPN3	176.30	AARS2	145.24	ALDH7A1	214.68	ADAMTS2	87.79
CAV3	173.20	AIFM1	159.17	ANO3	203.71	B3GALT6	142.23
COL6A1	96.50	ARHGEF10	166.58	ARSA	93.34	B4GALT7	116.06
COL6A2	86.49	BICD2	113.58	ASPA	144.76	BMP1	128.60
COL6A3	194.88	BSCL2	156.57	ATM	188.82	CBS	97.59
DAG1	181.80	COX6A1	212.49	ATP1A3	117.18	CHST14	136.97
DES	130.76	CTDP1	134.72	CIZ1	121.86	CLCN7	97.11
DNAJB6	172.84	DCTN1	182.51	COL4A1	194.60	COL1A1	138.63
DYSF	145.07	DHTKD1	190.35	COL4A2	168.31	COL1A2	208.36
FKRP	161.64	DNAJB2	167.59	DBH	116.92	COL2A1	147.43
FKTN	185.44	DNM2	181.92	DDC	165.03	COL3A1	193.33
GAA	102.96	DYNC1H1	195.89	FUCA1	205.06	COL5A1	137.74
GMPPB	145.12	EGR2	159.29	GALC	192.01	COL5A2	188.09
HNRNPDL	264.51	FBLN5	195.00	GCDH	130.68	CREB3L1	124.45
ISPD	183.63	FGD4	181.03	GCH1	185.76	CRTAP	217.12
ITGA7	151.62	FIG4	184.24	GJC2	66.49	DSE	199.87
LAMA2	194.20	GAN	214.86	GLB1	170.50	FBN1	186.90
LARGE	186.59	GARS	189.69	GNAL	202.66	FBN2	187.17
LIMS2	98.79	GDAP1	190.19	HEXA	187.39	FKBP10	135.22
LMNA	121.57	GJB1	118.80	LAMP2	163.80	FKBP14	157.48
MYOT	205.76	GNB4	165.74	LAMP3	196.00	FLNA	101.77
PLEC	67.80	HARS	197.27	PAH	190.30	FLNB	191.96
PNPLA2	92.83	HINT1	192.45	PLA2G6	123.57	IFITM5	89.40
POMGNT1	174.27	HK1	173.75	PNKD	140.09	IKBKG	112.36
POMK	151.50	HSPB1	108.55	PNPO	179.41	LEPRE1	166.76
POMT1	166.89	HSPB3	231.68	PRKRA	197.81	LRP5	97.75
POMT2	169.50	HSPB8	171.79	PRRT2	110.62	MBTPS2	141.60
SEPN1	124.05	IGHMBP2	150.93	PSAP	176.84	OSTM1	172.95
SGCA	123.26	KARS	197.75	PTS	172.30	P4HB	162.14
SGCB	214.64	KIF1B	198.69	QDPR	161.02	PLOD1	116.35
SGCD	201.68	KIF5A	185.79	SGCE	170.96	PLOD2	201.61

doi:10.6342/NTU201704070

Table 4. (Cont'd) The average coverage of genes in LGMD, CMT, NMD, and OI/EDS.

LGMD Gene	Average Coverage	CMT Gene	Average Coverage	NMD Gene	Average Coverage	OI; EDS; Osteopetrosis Gene	Average Coverage
SGCG	175.69	LMNA	121.57	SLC17A5	187.30	PLS3	149.37
SMCHD1	191.42	LRSAM1	120.50	SLC18A2	165.41	PPIB	154.20
TCAP	127.87	MARS	211.51	SLC2A1	159.89	RIN2	179.57
TNPO3	214.71	MED25	116.12	SLC6A3	110.77		
TRAPPC11	186.40	MFN2	156.07	SPR	143.52		
TRIM32	124.90	MICAL1	141.40	SUMF1	168.05		
TTN	240.02	MPZ	165.22	TAF1	163.58		
		NDRG1	170.20	TH	69.05		
		NEFL	230.59	THAP1	233.88		
		PDK3	132.34	TOR1A	205.95		
		PLEKHG5	98.12	TUBB4A	87.84		
		PMP22	175.53				
		PRPS1	158.41				
		PRX	110.67				
		RAB7A	212.16				
		SBF1	85.89				
		SBF2	190.21				
		SH3TC2	177.38				
		SLC12A6	202.64				
		SPG11	197.08				
		SURF1	157.50				
		TDP1	170.76				
		TFG	169.78				

doi:10.6342/NTU201704070

Table 5. Forty patients participated in the study, listed according to disease category.

Patient ID	Clinical diagnosis	Sex	Age	Onset
A	Polyneuropathy / CMT	M	11	Progressive athletic function deterioration since age 11
B	Polyneuropathy / CMT axonol type	M	15	Symptoms appear since young age Maternal aunt's son has similar symptoms
C	Polyneuropathy / CMT	M	30	Gait abnormality for some years Leg atrophy since 29
D	CMT / Muscular dystrophy / Mitochondrial defect	M	53	Onset at 40 y/o Father similar symptoms
E	Polyneuropathy	M	61	Onset 49 y/o; weakness in legs
F	Muscular dystrophy / CMT / polyneuropathy	F	67	Onset 60 y/o; distal muscular dystrophy
G	Muscular dystrophy, Becker type	M	18	Weakness sine 12 y/o; progressive Maternal uncle and counsins found mutation in DMD gene
H	Muscular dystrophy	M	24	Since childhood
I	Muscular dystrophy / LGMD	M	20	First noted at 17 y/o; Diagnosed at Taichung China Medical University found [c.3207G>A p.Trp1069*]
J	Muscular dystrophy / LGMD	M	36	Onset 4-5 y/o; progressive
K	Muscular dystrophy / LGMD	F	44	

doi:10.6342/NTU201704070

Table 5. (Cont'd) Forty patients participated in the study, listed according to disease category.

Patient ID	Clinical diagnosis	Sex	Age	Onset
L	Muscular dystrophy / Mitochondrial defect	M	48	Weakness since 10 y/o
M	Muscular dystrophy / LGMD	M	40	Onset 27-28 y/o
N	Muscular dystrophy / LGMD	M	44	Onset 39 y/o
O	Muscular dystrophy / LGMD	F	23	Symptoms appear since childhood
P	Muscular dystrophy / LGMD	M	17	Muscle stiffness since 13 y/o Molecular diagnosis confirmed on *DYSF*
Q	Muscular dystrophy / LGMD	M	46	Diagnosed since 28 y/o
R	Muscular dystrophy / LGMD	F	50	Onset 10 y/o; wheel chair since 42 Younger brother and sister affected
S	Congenital myopathies / LGMD	F	54	Congenital Mother wheelchair bound
T	Muscular dystrophy / LGMD	F	49	Late onset; wheelchair bound
U	Congenital myopathies / LGMD	M	58	Onset 30 y/o Mother, brothers, and half-brothers affected
V	Encephalopathy / Muscular dystrophy	F	43	Since 32 y/o, psychological problem 34 y/o, brain abnormality by CT
W	Muscular dystrophy	F	15	Chronic muscle weakness since 3 years old
X	Congenital muscular dystrophy	M	1	
Y	Congenital muscular dystrophy	F	45	Onset 41 y/o

doi:10.6342/NTU201704070

Table 5. (Cont'd) Forty patients participated in the study, listed according to disease category.

Patient ID	Clinical diagnosis	Sex	Age	Onset
Z	Mitochondrial defect / Rhabdomyolysis	M	29	
AA	Mitochondrial defect / HyperCKnemia	F	32	
AB	Ehlers-Danlos syndrome	M	14	
AC	Ehlers-Danlos syndrome IV	F	12	
AD	Osteogenesis imperfecta	F	26	
AE	Osteogenesis imperfecta	F	19	
AF	Osteogenesis imperfecta / Ehlers-Danlos syndrome	F	4	
AG	Osteogenesis imperfecta	M	12	
AH	Osteogenesis imperfecta	F	49	
AI	Spondyloepiphyseal dysplasia	M	4	

doi:10.6342/NTU201704070

Table 5. (Cont'd) Forty patients participated in the study, listed according to disease category.

Patient ID	Clinical diagnosis	Sex	Age	Onset
AJ	Spondyloepiphyseal dysplasia	M	15	X-shaped lower extremities since 5 y/o
AK	Bartter syndrome	M	48	
AL	Hypokalemia / Hypercalcemia	F	43	Hypokalemia known in 39 y/o
AM	Alpers syndrome	M	1	
AN	GM1	F	15	Regression since 2.5 y/o

doi:10.6342/NTU201704070

Table 6. Pathogenic variants found in each patient.

Patient ID	Gene	Nucleotide change	Zygosity	Protein change	Molecular diagnosis	Inheritance
C	MFN2	c.839G>A	Het	p.Arg280His	CMT 2A	AD
F	MFN2	c.775C>T	Het	p.Arg259Cys	CMT 2A	AD
G	DMD	c.2169-1G>T	Hemi		Becker's muscular dystrophy	X-linked
I	DYSF	c.937+1 G>A	Het		LGMD 2B (Miyoshi)	AR
		c.3207G>A	Het	p.Trp1069*		
J	TRAPPC11	c.2938G>A	Hom	p.Gly980Arg	LGMD2S	AR
N	MYH7	c.1322C>T	Het	p.Thr441Met	Cardiomyopathy and distal myopathy	AD
O	LMNA	c.148C>G	Het	p.Arg50Gly	LGMD 1B	AD
P	DYSF	c.863A>T	Het	p.Asp288Val	LGMD 2B	AR
		c.1667T>C	Het	p.Leu556Pro		
Q	CAPN3	c.1309C>G	Het	p.Arg437Gly	LGMD 2A	AR
		c.2092C>T	Het	p.Arg698Cys		
S	DMD	c.2956C>T	Het	p.Gln986Ter	DMD	X-linked
		c.2136G>C	Het	p.Arg712Ser		
U	HNRNPDL /HNRPDL	c.1132 G>A	Het	p.Asp378Asn	LGMD 1G	AD

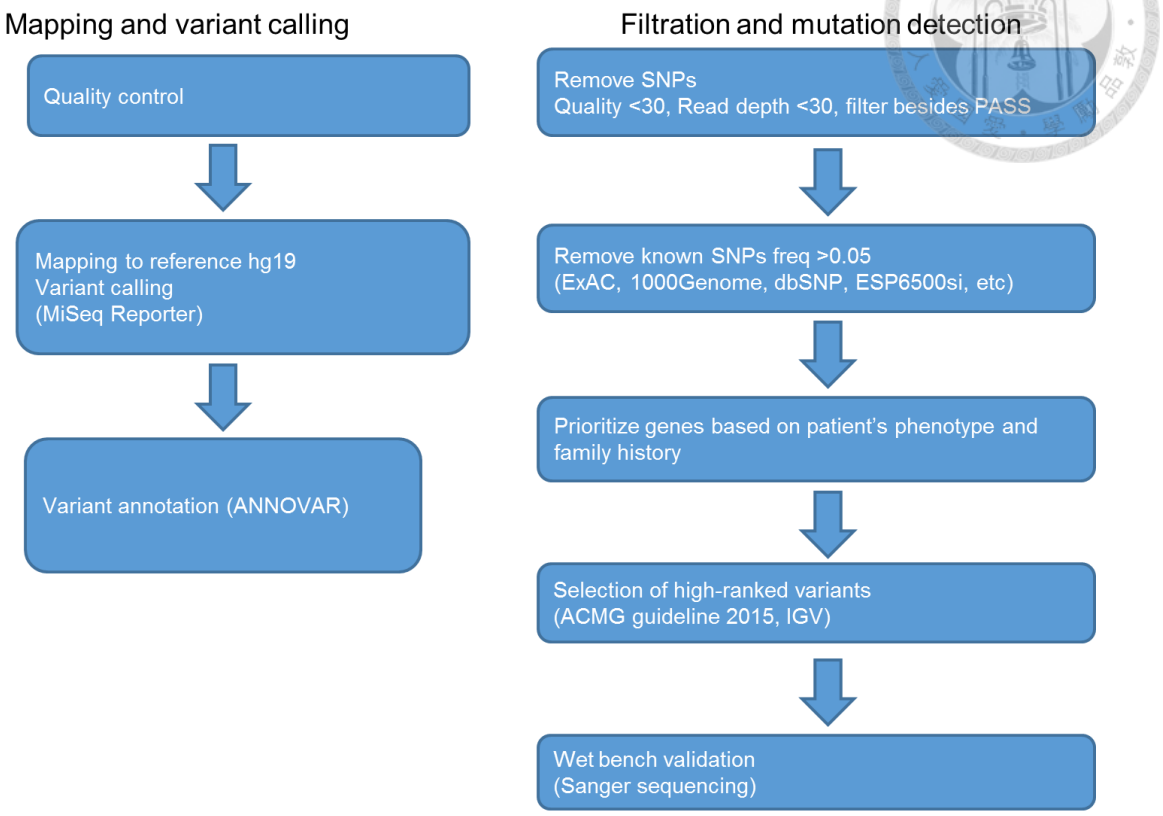
doi:10.6342/NTU201704070

Table 6. (Cont'd) Pathogenic variants found in each patient

Patient ID	Gene	Nucleotide change	Zygosity	Protein change	Molecular diagnosis	Inheritantce
V	CSF1R	c.2330G>A	Het	p.Arg777Gln	Leukoencephalopathy, diffuse hereditary, with spheroids	AD
AG	WNT1	c.104+1G>A	Het			
		c.105G>A	Het	p.Trp35*	OI, type XV	AR
AH	COL1A1	c.661G>C	Het	p.Gly221Arg	OI, type I	AD
AI	TRPV4	c.1781G>A	het	p.Arg594His	SED, Maroteaux type	AD
AK	SLC12A3	c.488C>T	Het	p.Thr163Met	Gitelman syndrome	AR
		c.2612G>A	Het	p.Arg871His		
		c.734T>C	Het	p.Leu245Pro		

doi:10.6342/NTU201704070

Figure 1. The bioinformatics analysis pipeline for the NGS data.

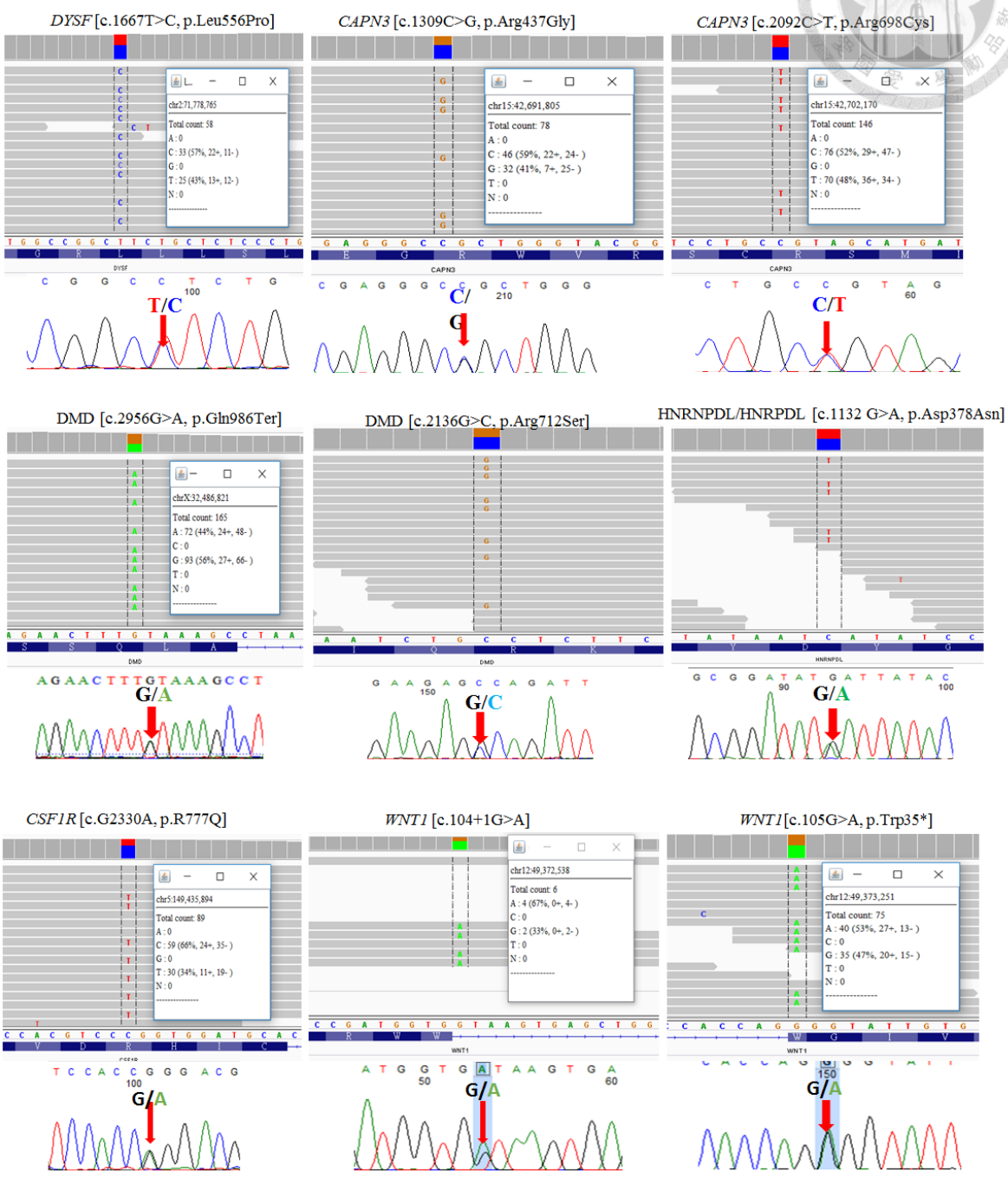


doi:10.6342/NTU201704070

Figure 2. Sequencing data of the pathogenic mutations found in patients depicted by IGV and confirmation by Sanger sequencing.

doi:10.6342/NTU201704070

Figure 2. (Cont'd) Sequencing data of the pathogenic mutations found in patients depicted by IGV and confirmation by Sanger sequencing.



doi:10.6342/NTU201704070

Figure 2. (Cont'd) Sequencing data of the pathogenic mutations found in patients depicted by IGV and confirmation by Sanger sequencing.

doi:10.6342/NTU201704070