จีโนมและเอ็กโซมของแฝดแท้ที่เป็นไตรโซมี 21 แต่มีโรคหัวใจแต่กำเนิดในแฝดเพียงคนเดียว

นายพงศธรณ์ ไชยทรัพย์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาชีวเวชศาสตร์ (สหสาขาวิชา)
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2557

GENOME AND EXOME SEQUENCES OF MONOZYGOTIC TWINS WITH TRISOMY 21, DISCORDANT FOR A CONGENITAL HEART DEFECT

Mr. Pongsathorn Chaiyasap

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Biomedical Sciences
(Interdisciplinary Program)
Graduate School
Chulalongkorn University
Academic Year 2014

Thesis Title	GENOME AND EXOME SEQUENCES OF MONOZYGOTIC TWINS WITH TRISOMY 21, DISCORDANT FOR A CONGENITAL HEART DEFECT
By	Mr. Pongsathorn Chaiyasap
Field of Study	Biomedical Sciences
Thesis Advisor	Professor Vorasuk Shotelersuk, M.D.
Thesis Co-Advisor	Associate Professor Kanya Suphapeetiporn, M.D., Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.....Dean of the Graduate School
(Associate Professor Sunait Chutintaranond, Ph.D.)

THESIS COMMITTEE

.....Chairman
(Professor Apiwat Mutirangura, M.D., Ph.D.)

.....Thesis Advisor
(Professor Vorasuk Shotelersuk, M.D.)

.....Thesis Co-Advisor
(Associate Professor Kanya Suphapeetiporn, M.D., Ph.D.)

.....Examiner
(Associate Professor Sunchai Payungporn, Ph.D.)

.....Examiner
(Assistant Professor Rachaneekorn Tammachote, Ph.D.)

.....External Examiner
(Sissades Tongsima, Ph.D.)

พงศธรณ์ ไชยทรัพย์ : จีโนมและเอ็กโซมของแฝดแท้ที่เป็นไตรโซมี 21 แต่มีโรคหัวใจแต่กำเนิดในแฝดเพียงคนเดียว (GENOME AND EXOME SEQUENCES OF MONOZYGOTIC TWINS WITH TRISOMY 21, DISCORDANT FOR A CONGENITAL HEART DEFECT) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ศ. นพ. วรศักดิ์ โชติเลอศักดิ์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. พญ. ดร.กัญญา ศุภปีติพร, หน้า.

ร้อยละ 40 ของผู้ป่วยกลุ่มอาการไตรโซมี 21 มีโรคหัวใจพิการแต่กำเนิด โดยร้อยละ 60 ที่เหลือจะมีโครงสร้างหัวใจปกติ การพบโรคหัวใจพิการแต่กำเนิดเพียงบางส่วนของผู้ป่วยนี้ บ่งว่าจำนวนชุดของยีนที่เพิ่มขึ้นบนโครโมโซมที่ 21 อาจมีปฏิสัมพันธ์กับอัลลีลของยีนบนโครโมโซมเส้นอื่นๆ ส่งผลให้ความเสี่ยงของการเกิดโรคหัวใจพิการแต่กำเนิดเพิ่มขึ้น ในงานวิจัยนี้เราได้ศึกษาผู้ป่วยแฝดแท้ที่เป็นไตรโซมี 21 ทั้งสองคน แต่พบโรคหัวใจบกพร่องแต่กำเนิดเพียงคนเดียว ผลการวิเคราะห์ไมโครแซทเทลไลท์ ยืนยันว่าแฝดคู่นี้เป็นแฝดเหมือน ผู้วิจัยตั้งสมมติฐานว่าการแสดงออกของโรคที่ต่างกันในแฝดคู่นี้น่าจะเกิดจากการกลายพันธุ์หลังไข่แยกตัวเป็นแฝด ดังนั้นผู้วิจัยจึงได้ใช้เทคโนโลยีเอ็นจีเอสซึ่งเป็นเครื่องวิเคราะห์หาลำดับสารพันธุกรรมในการศึกษาทั้งจีโนมและเอ็กโซมของดีเอ็นเอที่สกัดจากเลือดของแฝดคู่นี้ ผลการวิเคราะห์ข้อมูลลำดับพันธุกรรมแสดงให้เห็นตำแหน่งที่มีพันธุกรรมต่างกันระหว่างแฝดคู่นี้ 15 ตำแหน่ง อย่างไรก็ตามเมื่อได้ทำการตรวจสอบตำแหน่งพันธุกรรมเหล่านี้ด้วยวิธีการหาลำดับพันธุกรรมแบบแซงเกอร์โดยใช้สารพันธุกรรมจากเม็ดเลือดขาวพบว่าพันธุกรรมของแฝดคู่นี้ทั้ง 15 ตำแหน่งไม่แตกต่างกัน การที่ไม่พบพันธุกรรมที่ต่างกันนั้นบ่งชี้ว่าพันธุกรรมที่ต่างกันของแฝดเหมือนนั้นอาจพบได้น้อยมาก นอกจากนี้ยังชี้ให้เห็นถึงข้อจำกัดในการใช้เทคโนโลยีเอ็นจีเอสในการหายีนที่เป็นสาเหตุของการแสดงออกของโรคที่ต่างกันในแฝดเหมือน

สาขาวิชา ชีวเวชศาสตร์ ลายมือชื่อนิสิต .....

ปีการศึกษา 2557 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5387847320 : MAJOR BIOMEDICAL SCIENCES
KEYWORDS: GENOME / EXOME / TRISOMY 21 / DOWN SYNDROME / MONOZYGOTIC TWIN / IDENTICAL TWIN / NEXT GENERATION SEQUENCING / CONGENITAL HEART DEFECTS / VENTRICULAR SEPTAL DEFECT

PONGSATHORN CHAIYASAP: GENOME AND EXOME SEQUENCES OF MONOZYGOTIC TWINS WITH TRISOMY 21, DISCORDANT FOR A CONGENITAL HEART DEFECT. ADVISOR: PROF. VORASUK SHOTELERSUK, M.D., CO-ADVISOR: ASSOC. PROF. KANYA SUPHAPEETIPORN, M.D., Ph.D., pp.

Congenital heart defects (CHD) occur around 40% of individual with trisomy 21 while the remaining 60% have a structurally normal heart. This suggests that the extra copy of genes on chromosome 21 is a risk factor for abnormal heart development. Increased dosage of genes on chromosome 21 could interact with certain alleles of genes on other chromosomes and might contribute to CHD. Here, we identified a pair of Thai monozygotic twins with trisomy 21 but discordant for congenital heart defects. Twin-zygosity was confirmed to be monozygotic by microsatellite genotyping. We hypothesized that the discordant phenotype in these monozygotic twins could be resulted from post-twining mutation. Therefore, we applied next generation sequencing (NGS) technologies to sequence both genome and exome from their leukocytes DNA. The post-analysis of the sequencing data identified 15 discordant exonic variants between the twins. However, validation of all variants with conventional Sanger sequencing revealed no differences between both twins. Given that no discordant variants were found suggests that sequence differences of leukocytes' DNA of monozygotic twins might be extremely rare. This also suggests the limitation of the current NGS technology in identifying causative genes for discordant phenotypes in monozygotic twins.

Field of Study: Biomedical Sciences Student's Signature .....
Academic Year: 2014 Advisor's Signature .....
Co-Advisor's Signature .....

ACKNOWLEDGEMENTS

I would like to express my gratitude to everyone who participated in this work. First, I am deeply grateful to my advisor, Prof. Vorasuk Shotelersuk and my co-advisor, Assoc. Prof. Dr.Kanya Suphapeetiporn for all their guidance, understanding, patience, helps, suggestions and encouragement throughout the time during my graduate studies. I also would like to express my heartfelt thanks to all committee members, Prof. Dr.Apiwat Mutirangura, Assoc. Prof. Dr.Sunchai Payungporn, Asst. Prof. Dr.Rachaneekorn Tammachote and Dr.Sissades Tongsima for their helpful suggestion and correction.

Most importantly, I would like to give my special thanks to the patients and their family for participation in this study.

I am really grateful to my colleague, Mr.Supasak Kulawonganunchai for his helps and contributions. Furthermore, I would like to thank every member at the Center of Excellence for Medical Genetics, especially Mr. Chalurmpon Srichomthong and Miss Siraprapa Thongkobpetch, for their guidance, helps and support since the first time I started laboratory practice. I would like to thank Dr. Chupong Ittiwut and Dr. Rungnapa Ittiwut and Miss Yaowapa Raksadan for all their supports. The work in this thesis would have been impossible to accomplish without their helps.

Finally, I would like to express my deepest gratitude to my dear parents, Mr. Pakorn Chaiyasap and Mrs. Nathathai Chaiyasap for their love, guidance, understanding, encouragement and supports throughout my life. Moreover, I thank my dear friend, Tananya Taweephoncharoen, who provides me love and support.

Financial support was provided by Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0071/2552), Ratchadapiseksomphot Endowment Fund of Chulalongkorn University (RES560530177-HR), National Science and Technology Development Agency (NSTDA) and Thailand Research Fund (RTA5680003, RSA5480022, RSA5480026)

CONTENTS

Page

LIST OF ABBREVIATIONS

AD	average depth
AVSD	atrioventricular septal defect
BAM	binary alignment map
bp	base pair
BQS	base quality score
BWA	Burrows-Wheeler Alignment
CHD	congenital heart defect
DAH	duplicate array hybridization
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DS	Down syndrome
GATK	Genome Analysis Toolkit
Gb	gigabases
IGV	Integrative Genomics Viewer
Indel	small insertion or deletion

Mb	megabases
mRNA	messenger RNA
NEV	non-exonic variant
NGS	next generation sequencing
PCR	polymerase chain reaction
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RTA	real time analysis
SAM	sequence alignment map
SNV	single nucleotide variants
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
sstDNA	single-stranded template DNA
SV	synonymous variant
VACTERL	Vertebral anomalies, Anal atresia, Cardiac defects, Tracheoesophageal fistula and/or Esophageal atresia, Renal & Radial anomalies and Limb defects

VSD	ventricular septal defect
WES	whole exome sequencing
WGS	whole genome sequencing

CHAPTER I

INTRODUCTION

Background and rationale

Down syndrome (DS; OMIM 190685) is a human chromosomal disorder caused by an extra copy of genomic regions on chromosome 21. This type of chromosomal disorder is one of the most common causes of human genetic disorders, occurs at approximately 1 in 750 live births [1]. The extra copy of chromosome 21 is mostly due to failure of normal chromosomal segregation in maternal meiosis, which account for approximate 87% of Down syndrome patients [2]. Common facial characteristic includes oblique eyes, flat nasal bridge, protruding tongue and epicanthus. Other phenotypes include short and broad hands, intellectual deficit, hypotonia and other associated development disorders and congenital anomalies [3].

Some Down syndrome phenotype present in every DS patient, while other presents in only a fraction of DS patients. One of the severe phenotypes is the congenital heart defect (CHD). It occurs in approximately 40% of the patients with trisomy 21 while occurs only 0.8% in the general population [1]. This suggests that the extra genes on chromosome 21 is a necessary risk factor but not sufficient for abnormal heart development [4]. Increase dosage of genes on chromosome 21 could interact with certain alleles of genes on the other chromosomes and might contribute to CHD.

In the aspect of genetic disorders, phenotype discordance in monozygotic twins may provide a clearer understanding of relevant factors that are involved in disease etiology [5]. Despite the fact that monozygotic twins are generally considered

as genetically identical, genetic differences may arise during embryonic development. These genetic differences could be single nucleotide mutations, insertion or deletions, copy number variation, conversion, or postzygotic mitotic recombination. These genetic variations have been suggested as possible mechanisms which might cause discordant monozygotic twins [6]. A recent study had demonstrated that there are differences in single nucleotide polymorphism between monozygotic twins [7]. Therefore, we hypothesize that discordant phenotype in monozygotic twins could result from acquired genetic differences.

Usually, the detection of genetic variants can be done by comparison between a sequence of interest and the reference genome. However, in the case of monozygotic twin, we can compare their sequences against each other. Since the rest of the genome of monozygotic twins should be identical, mismatches identified could be selected for validation as possible mutations underlying discordant phenotype, For example, a study of monozygotic twin pair discordance for autosomal dominant neurofibromatosis type 1 (NF1) has demonstrated this idea. They found a *de novo NF1* mutation in all investigated cells of the affected twins, while cells from the unaffected twin were mosaic [7].

Since there should be only a small number of genetic differences between twins [8], a genome-wide coverage and highly sensitive method with high resolution should ideally be applied [9]. With the emerging of next generation sequencing (NGS) technology, detection of genetic differences at a single nucleotide level is possible [10]. In addition, a genome-wide coverage approach would allow for a non-biased, non-restricted to certain pre-select regions. Suggested discordant variants results from NGS can be validated with a conventional Sanger sequencing [11].

Here, we identified a pair of Thai monozygotic twins with trisomy 21 but discordant for congenital heart defects. We hypothesized that genetic differences from post-twinning mutations, such as single nucleotide variants (SNVs) or small insertions or deletions (Indels), may cause the discordant phenotypes. Therefore, we apply next-generation-sequencing in order to identify such genetic differences in these twins.

Research questions

1. What are the genetic differences between a pair of monozygotic twins with trisomy 21, discordant for a congenital heart defect?
2. Are these genetic differences associated with congenital heart defects in trisomy 21?

Objectives

1. To identify the genetic differences between monozygotic twins with trisomy 21, who were discordant for a congenital heart defect.
2. To study genetic variants associated with congenital heart defects in patients with trisomy 21.

Hypotheses

1. There are genetic differences between the monozygotic twins.
2. These differences are associated with the phenotypic discordance.

Key word

genome, exome, trisomy 21, Down syndrome, monozygotic twin, identical twin, discordance, next generation sequencing, congenital heart defects, ventricular septal defect

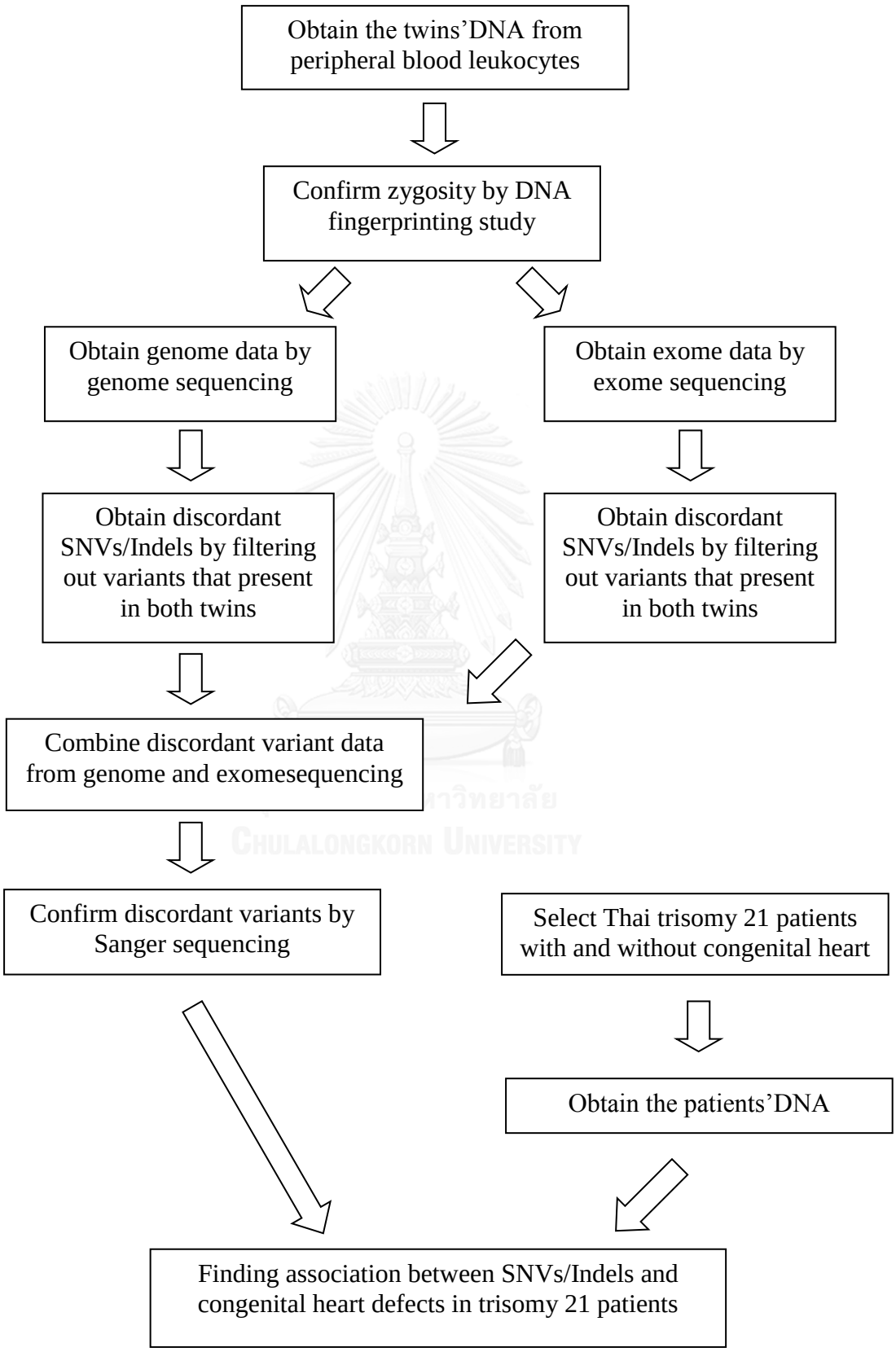
Ethical consideration

The study was approved by the institutional review board of Faculty of Medicine of Chulalongkorn University.Written informed consent was obtained from the parents of patients who participated in this study.

Expected benefit

This study could demonstrate the genetic differences in monozygotic twins with discordant phenotype. Furthermore, it may contribute to better understanding of the etiology and mechanism of congenital heart defects and basic knowledge of twins.

Conceptual Framework



CHARPTER II

REVIEW OF RELATED LITERATURE

Congenital heart defect and Down syndrome

Ventricular septal defect (VSD) is one type of congenital heart defects, which is described as openings in the wall between the left and right ventricles of the heart. It is one of the most common forms of congenital heart defects in childhood. VSD is mostly resulted from incomplete development of the heart during the first 6 weeks of pregnancy. The prevalence of VSD is around 0.8% per live births in the general population [12], and accounts for nearly 50% of all newborns with congenital heart diseases [13]. In addition, some chromosomal disorders including trisomy 21, can increase the risk of congenital heart defects to nearly 40% in these patients [1]. With this fixed pattern of congenital heart defects, it is possible that there may be a locus on chromosome 21 that is involved in the development of the cardiac defect [14]. However, not a single gene has yet been identified to be responsible for this phenotype. There is a study suggesting that overexpression of type VI collagen gene, which is located on chromosome 21, plays a role in the pathogenesis of atrioventricularseptal defect (AVSD) in Down syndrome [15]. It has been noted that not all patients with Down syndrome have an AVSD and this kind of heart defect is also seen in patients without Down syndrome. With this observation, it is indicated that other genes, not located on chromosome 21, play a role in the development of this cardiac defect and may be influenced by gene dosage imbalance from extra chromosome 21.

Next generation sequencing technology

All NGS platforms share a common feature which is massively parallel sequencing of DNA molecules that are separated in a flow cell. From this reason, the instrument can generate hundreds of megabases (Mb) to gigabases (Gb) of nucleotide sequence from a single run, depending on the platform [16]. There are many considerations when NGS approach will be applied in research. First, the depth of sequence coverage will depend on regions of interest. For example, 90 Gb of sequence is required to obtain 30-fold average coverage of the genome, while 3 Gb of sequence will be able to achieve 75-fold average coverage of the exome [17]. The reason is because the exome represents only about 30 Mb comparing to about 3 Gb of the genome. However, these are the theoretical number because there are inefficiencies in the process like sequence regions with low sequence coverage, and off-target hybridization. Furthermore, another important consideration is that NGS technologies have higher base calling error rates than conventional Sanger sequencing [18]. However, this can be relieved by increasing the depth of sequencing coverage to ensure minimal false positive [18]. From this reason, it is important to validate mutant or variant genotype by using conventional sequencing technique like Sanger sequencing.

Overview of Illumina NGS technology

The Illumina NGS technology applies a sequencing-by-synthesis approach which means all fours nucleotides are added simultaneously, along with DNA template, and DNA polymerase to the flow cell channels for incorporation into the oligo-primed cluster fragments. The surface of each flow cell is covered with

Illumina-specific adapter library. Each of all four nucleotides carries a base-unique fluorescent label and the sequence is read by each cycle of synthesis. With each nucleotide added, the fluorescent will be emitted and capture. The process of sequencing can be divided into template preparation, sequencing, and data analysis.

The process of DNA library preparing start with nebulization of genomic DNA to make the double strand DNA become short sequence approximately 200-500 base pairs. The size of the DNA fragment is controlled by pressure of the gas in the nebulizer. After that, the end of the DNA fragment will be repaired and ligated with adapter. These adapters contain sequencing primer binding site and region that is complementary to oligo on the flow cell. Illumina utilized the solid-phase amplification to produce randomly distributed, clonally amplified cluster from fragment on a flow cell. This process is called clustering. The flow cell is a glass slide which contains multiple lanes. The surface of each lane is coated with two types of adapter-specific adapter oligo. Hybridization is enabled by the first type of the oligo on the surface. This oligo is complementary to the adapter region on one of the fragment strand. When the template is flowed through the lane, the fragment strand will bind to the surface oligo. DNA polymerase then creates a complement sequence of the hybridized fragment. After that, the double strand molecule is denatured to a single strand. The original template then washed away, leaving only the complement sequence which attached on the glass slide. The strand is clonally amplified through bridge amplification. In this process, the strand folds over and the adapter region hybridizes to the second type of the oligo on the flow cell. DNA polymerase then generated the complementary strand which forms the double strand bridge. This bridge is denatured, resulting in two single strands copy of the DNA template that

attached to the flow cell. The process of bridge amplification then repeats in many cycles. It also occurs simultaneously for millions of clusters resulting in clonal amplification of all the fragments. After the bridge amplification, the reverse strands are cleaved from the oligo on surface and washed away, leaving only the forward strands. The 3'end is blocked to prevent bridge forming.

Sequencing begins by hybridization of the first sequencing primer on the complementary region on the adapter of the DNA template. With each cycle, the fluorescent tag nucleotide that complemented to the DNA template will corporate into the extending sequencing. After the addition of each nucleotide, the clusters are excited by a light source and the characteristic fluorescent signal is emitted. The number of the cycle determines the length of the read while the wave length of the emitted light and intensity determine base call. For a given cluster, all identical strands are read simultaneously which make each cluster give enough light for detection. Hundreds of millions of cluster are sequenced in a massively-parallel process and give data of the billions of forward sequence read. After the completion of the first read, the read product is washed away and the 3'end of the template is de-blocked. The template now can fold over to form bridge formation and DNA polymerase then synthesize the reverse DNA strand, forming the double strand bridge template. This double strand bridge then denatured to form single strand template. The original forward strand then cleaved off and washed away, leaving the reverse strand attached to the flow cell. The sequencing process then repeats again by hybridization of the second sequencing primer to the reverse strand. The DNA polymerase incorporates the specific fluorescent tag nucleotide and the machine read the emitted fluorescent light to determine base call. As with first read, the cycle is

repeated until the reverse strand is completely sequenced. The entire process generation billions of sequence read representing all the fragments of the investigating DNA sample. All of the read sequences are separate into a small group which is called local sequence clustering. The grouping is based on the similarity of the base call. After that, forward and reverse read are paired to create continuous read sequences. These continuous read sequences then are aligned to the reference genome for variant identification. [19-23]

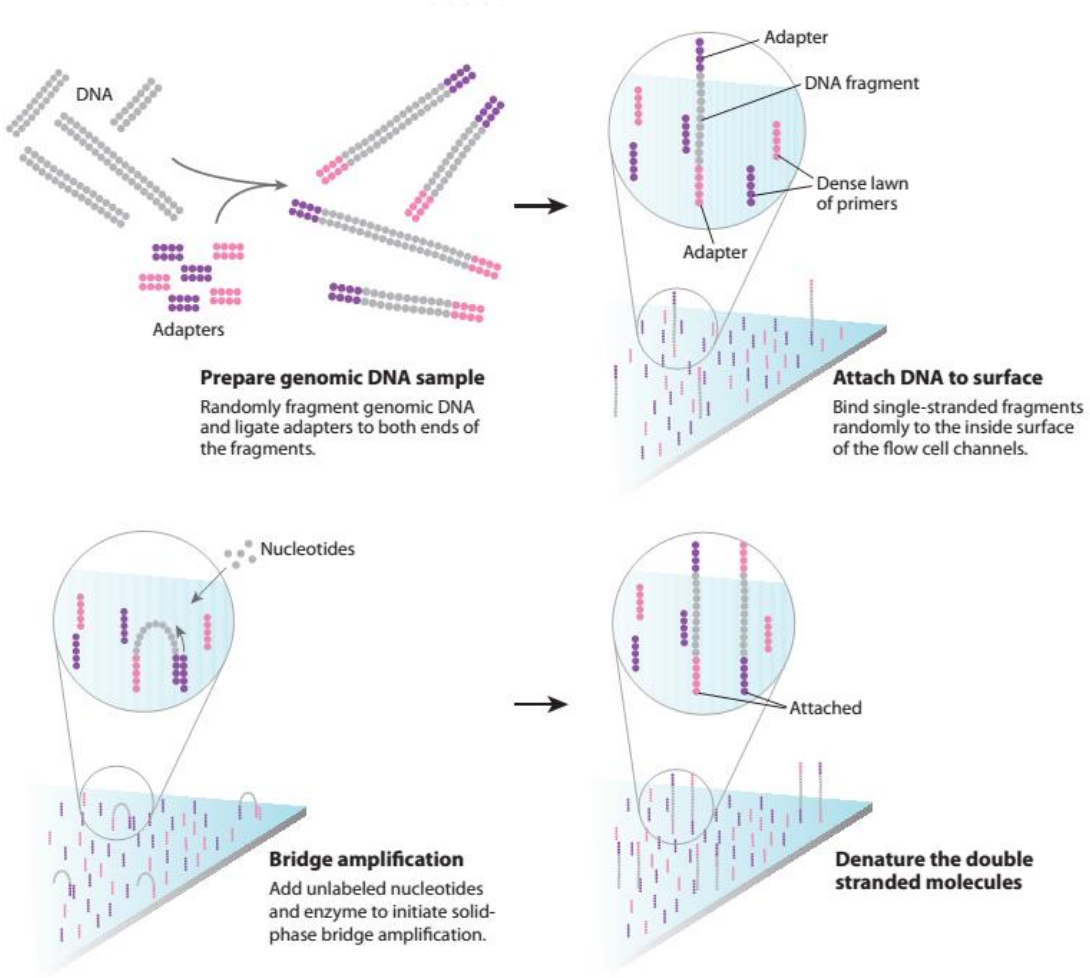


Figure 1.Illumina NGS technology: Template preparation and cluster amplification [19].

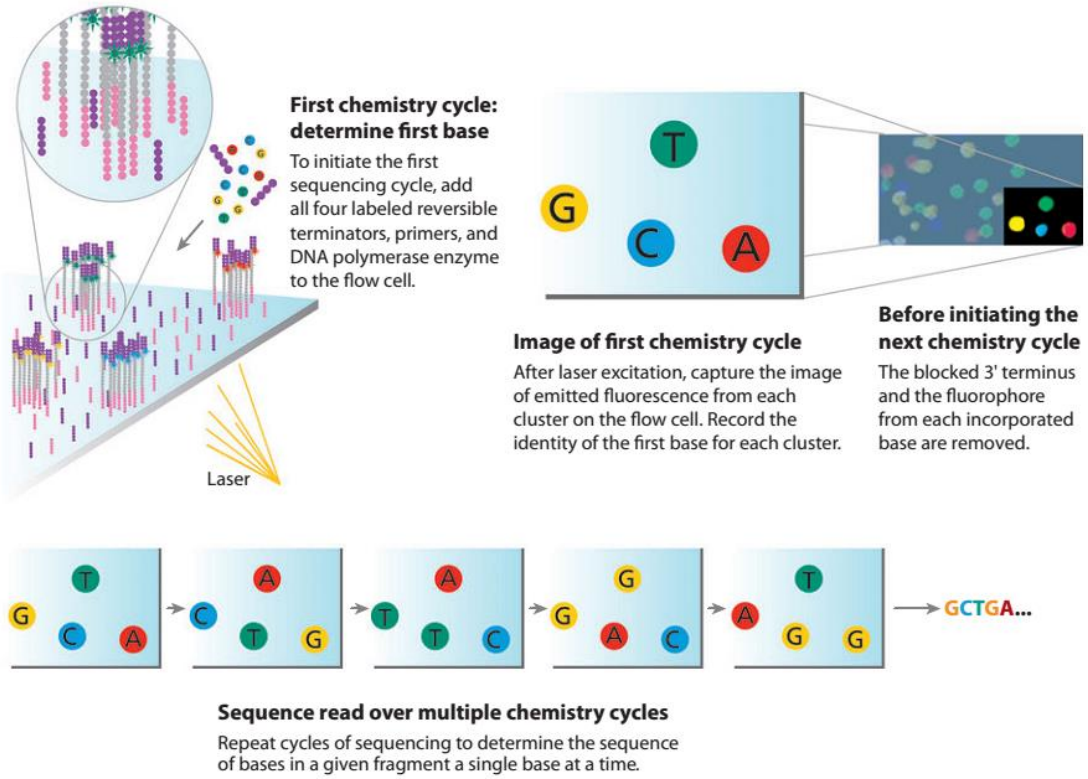


Figure 2.Illumina NGS technology: Sequencing-by-synthesis approach. DNA polymerase extends the synthesized strand with fluorescent tag nucleotide. The base is determined by the wavelength emitted from each cluster in each cycle [19].

Overview of Applied Biosystem NGS technology

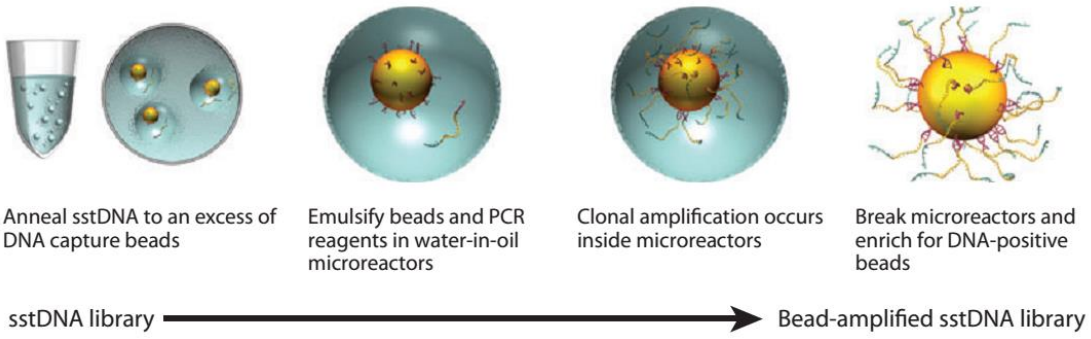
The Applied Biosystems SOLiD sequencer utilizes sequencing by oligonucleotide ligation and detection approach which is different from Illumina platform. The di-bases probe with fluorescence label is hybridized to the complementary template sequence. Then the endonuclease cleaves the last 3 bases at the 3’end of the probe. New di-bases probe then ligate to the template and the cycle is repeated. The system also uses unique two-base encoding for calling sequence base.

Template preparation begins with the similar step as in Illumina. The genomic DNA is sheared into smaller sequence and ligated with adapter to create fragment library. There are two types of template library in SOLiD system. First is the fragment library, which each template contain one fragment of the interesting DNA. Second one is the mate-paired library, which each template contain two fragments of the interesting DNA. The mate-paired library has an advantage that the distant between the two fragments is known. This will help increase the accuracy of the alignment process. The next step is the emulsion PCR, which is performed by put the DNA-library, DNA polymerase and 1 μm bead in the small droplets of water suspended in oil. These beads are coated on the surface with oligo that complement to the DNA fragment adapter. DNA library is diluted in this reaction to incorporate only one DNA fragment and one bead into a single droplet. The DNA fragment then hybridizes to the surface of the bead by annealing of adapter on template to the oligo on the bead. After that, the DNA polymerase synthesizes the complementary sequence from the template, resulting in the clonal amplification from the single template in the droplet. This process is called bead enrichment. Then beads are covalently attached to the glass slide and put in the SOLiD sequencer machine.

Sequencing begins with combining the template on the bead with the universal sequencing primer, ligase and large pool of di-base probes. The di-base probes are labeled with 4 fluorescent dyes, each dye represents 4 of 16 possible di-nucleotide sequences. The complementary probe hybridizes to the template sequence and is ligated. After that, the fluorescent is excited to emitted light and measured. Then the fluorescent is cleaved off and the next complementary probe hybridizes and ligated again. The cycle is repeated until the length of the desired read is acquired. When the

reading is finished, the double strand DNA is denatured and the synthesis strand is wash away, leaving the original template attached to the bead. A new sequencing primer is hybridized, and the complementary probe is ligated again. This new sequencing primer is offset from the first one by 1-base pair. The process of complementary probe hybridization, ligation, fluorescent measuring, and cleaved off are all repeated. After finish, the primer reset process is then repeated for total 5 rounds. This resulting in each base is measured twice by two independent ligation reactions by two different primers. Data analysis is done by collecting color image from each bead in each primer round and ligation cycle. With each color represents each represented di base pair; the color space for each sequence has to be decoded to determine the base sequence. [19-23]

Figure 3.Applied Biosystem NGS technology: Emulsion PCR [19]



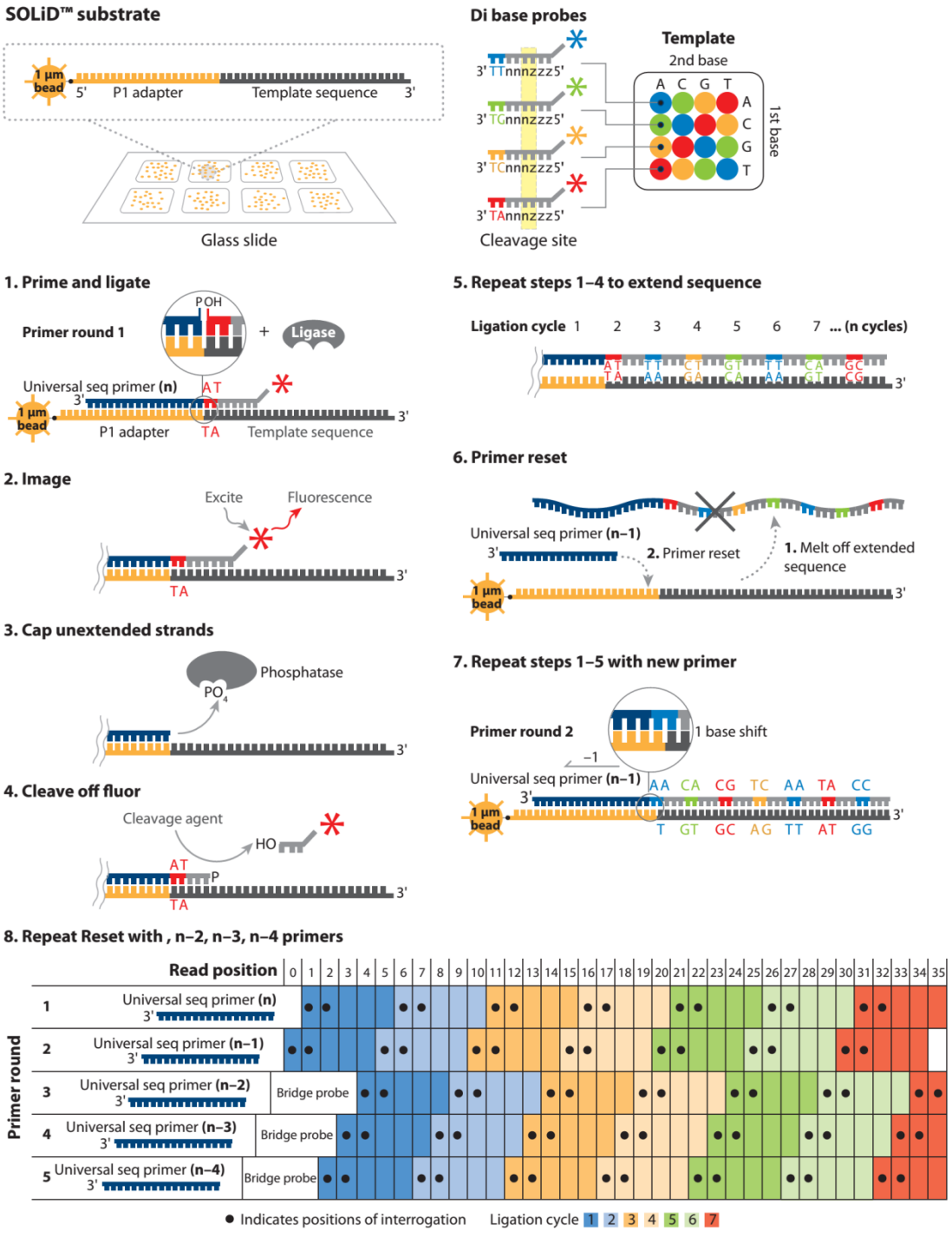


Figure 4. Applied Biosystem NGS technology: sequencing by oligonucleotide ligation and detection [19]

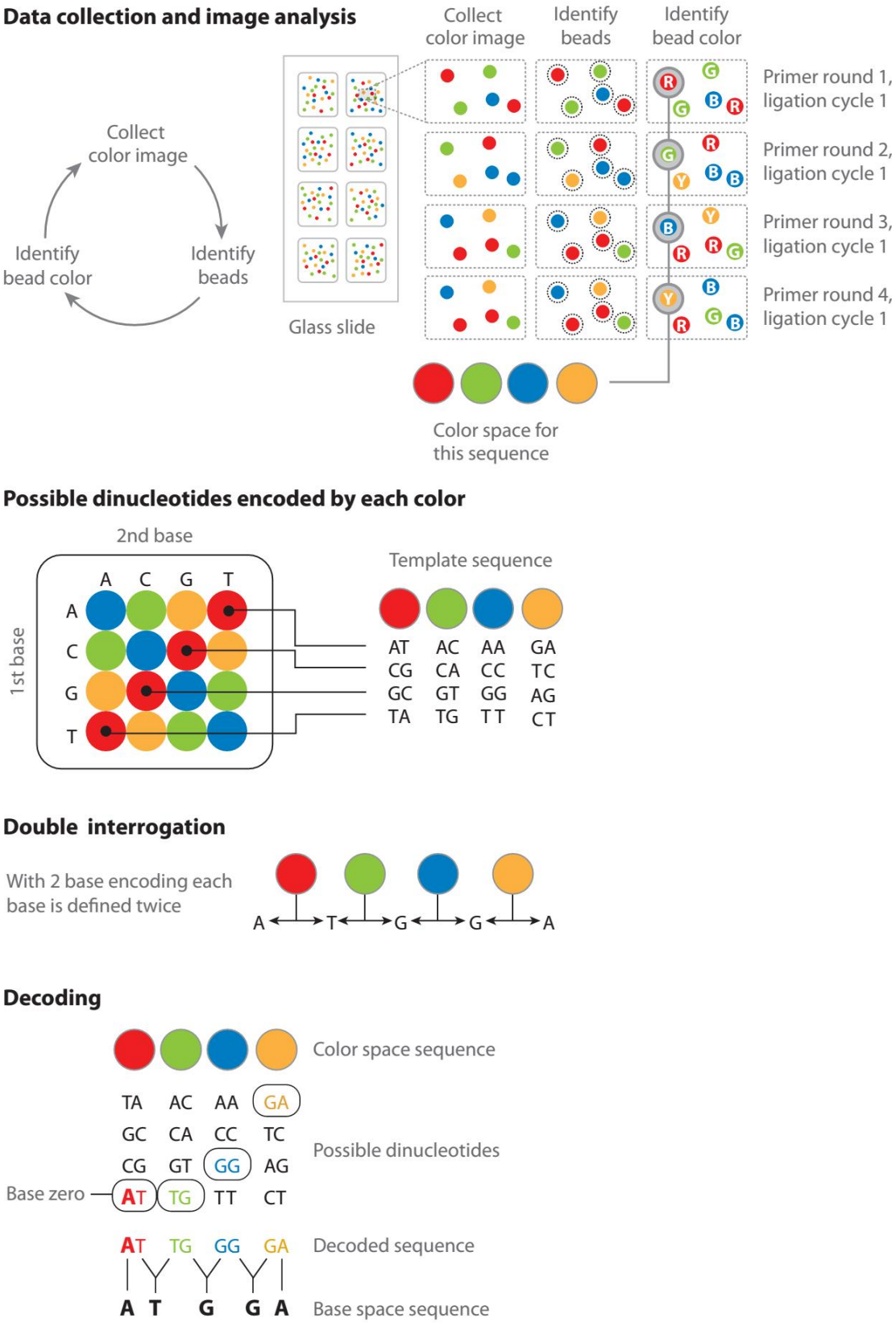


Figure 5. Applied Biosystem NGS technology: principles of two-base encoding.

Discordant twin study and next generation sequencing

Recently there are increasing numbers of research applying next generation sequencing with a discordant twin study. In one study, the researchers performed whole-genome sequencing of genomic DNA taken from a pair of monozygotic twins discordant for multiple sclerosis [24]. Approximately 3.6 million single nucleotide variants (SNVs) and 0.2 million small insertions and deletions (Indels) were detected in the genome of the twins. Among these SNVs and Indels, there was a small fraction that was found different between the twins. However, no discordant variant could be reproduced by mRNA sequencing or by SNP arrays, suggesting that this variation was due to sequencing or genotyping errors. In conclusion, this group of researchers did not find any evidence for genetic differences that could explain multiple sclerosis in monozygotic twins. One possible reason that this study failed to find true somatic mutations is the depths of coverage of sequencing. The average depth of coverage is approximately 22-fold for each twin, not reaching the 30-fold that can offer a reasonably high sensitivity to rule out both sequencing errors and germline mutations [25]. Though this study cannot find true discordant somatic mutations, it highlights the potential of next generation sequencing in a twin study approach.

Another research in 2010 demonstrated the usefulness of next generation sequencing [26]. In this study, the researchers identified a family with a recessive form of Charcot-Marie-Tooth disease. Then, they sequenced the whole genome of the proband to identify all potential functional variants and genotype these variants in the affected family members. The sequencing of DNA obtained from the proband had an average depth of coverage about 30-fold per base and detected approximately 3.4 million SNVs. After that, they cross-referenced the nonsynonymous SNVs with the

Human Gene Mutation database and resulted in 159 SNVs that revealed a clear association with a heritable trait. Two SNVs were associated with Charcot-Marie-Tooth disease in the database. To verify the mutation, they used PCR assay and directly sequenced *SH3TC2* gene in all members of the study family. With the result of segregation analysis in the family pedigree, they suggested that haploinsufficiency in *SH3TC2* was the underlying gene in their study. This publication successfully showed the possibility of gene identification using the next generation sequencing.

CHAPTER III

MATERIALS AND METHODS

Subject and clinical descriptions

We identified Thai twin boys who born at King Chulalongkorn Memorial Hospital, Bangkok. They presented with hypotonia, low-set ears, up-slant eyes, brachydactyly, clinodactyly, and flat nasal bridge. The physician had given a clinical diagnosis of Down syndrome. Later, the chromosome analysis showed the extra chromosome 21 in both twins, confirming the clinical diagnosis of Down syndrome. We assigned the older brother as the twin A and the younger brother as the twin B. Echocardiography was used to follow up on both twins. It revealed that twin A had a ventricular septal defect (VSD). He had given a corrective operation when he was one year and two months old. Although, twin B had a normal heart condition, he developed seizure when he was six months old. The seizure reoccurred even after he was treated with several antiepileptic drugs. At the age of one-year-old, he was given vigabatrin which stopped the seizure. Such seizure was not observed in twin A.

DNA Extraction

After informed consents were received, peripheral blood (6 ml) from both twins and their mother were obtained. Total genomic DNA extraction was extracted by using QIAamp® DNA blood mini kit, from their white blood cells. The procedure was performed according to the manufacturer's instruction (Qiagen, Valencia, CA).

1. Centrifuge approximately 3-6 ml. of blood at 1,000 x g for 10 minutes.
2. Remove supernatant plasma, then collect buffy coat to a new 15-ml tube.

3. Add 10 ml of cold lysis Buffer I and mix thoroughly.
4. Centrifuge at 1,000 x g for 5 min, discard supernatant and repeat this step one more time.
5. Add 200 µl of Buffer AL, and 20 µl of proteinase K to the sample.
6. Mix briefly with vortex mixer for 15 second and incubate at 56 °C for 10 min.
7. Add 200 µl of absolute ethanol to the sample, mix briefly with vortex mixer for 15 second.
8. Briefly centrifuge to collect all drops, then transfer the mixture to the QIAamp Mini spin column which place over 2 ml collecting tube.
9. Centrifuge at 6,000 x g for 1 min.
10. Discard the collecting tube then place the QIAamp Mini spin column into another clean 2 ml collecting tube.
11. Add 500 µl of Buffer AW1 and centrifuge at 6,000 x g for 1 min.
12. Discard the collecting tube then place the QIAamp Mini spin column into another clean 2 ml collecting tube.
13. Add 500 µl of Buffer AW2 and centrifuge at 20,000 x g for 3 min.
14. Discard the collecting tube then place the QIAamp Mini spin column into another clean 2 ml collecting tube.
15. Centrifuge at 20,000 x g for 1 min.
16. Discard the collecting tube then place the QIAamp Mini spin column into another clean 1.5 ml microcentrifuge tube.
17. Add 200 µl of Buffer AE or distilled water, wait for 5 min then Centrifuge at 20,000 x g for 2 min.

Zygosity analysis

ABI PRISM® Linkage Mapping Set v 2.5 was used to confirm that both twins were monozygotic. Thirteen microsatellite markers on thirteen different chromosomes were selected for this test.

Table 1. Selected ABI PRISM® markers

Chromosome	Locus	Panel & Dye	Dye label	Estimate size
1	D1S2785	P1V3	VIC	174-180
2	D2S206	P3V2	VIC	141-149
4	D4S424	P7V4	VIC	195-203
5	D5S408	P9F4	FAM	249-259
7	D7S657	P11N5	NED	248-262
9	D9S164	P15F1	FAM	83-91
11	D11S1314	P13V1	VIC	100-108
14	D14S74	P20V5	VIC	301-313
15	D15S127	P21N1	NED	128-154
16	D16S515	P21F5	FAM	335-353
18	D18S1161	P24N4	NED	224-230
20	D20S117	P25F2	FAM	149-181
21	D21S1914	P27V4	VIC	263-273

We performed PCR following the manufacturer's instructions. Each reaction included following reagents

Table 2. Reaction mix for ABI PRISM® Linkage Mapping Set

Reaction Components	**Volume (µl)**
Primer Mix (5µM each primer)	1.0
DNA (50 ng/µl)	1.2
True Allele PCR Premix	9.0
Sterile D.I. H_2O	3.8
Total Mix	15.0

Table 3. PCR condition for ABI PRISM® Linkage Mapping Set

Initial Step	**10 Cycles**			**20 Cycles**			**Final Step**
95 °C 12 min	94 °C 15 sec	55 °C 15 sec	72 °C 30 sec	89 °C 15 sec	55 °C 15 sec	72 °C 30 sec	72 °C 10 min

The DNA analysis was performed on an ABI Prism 3100 genetic analyzer, using fluorescently labeled selective primers (Applied Biosystems, Foster City). The data was analyzed with GeneMapper software (Applied Biosystems).

Genome sequencing and targeted capture exome sequencing

We sent genomic DNA from both twins to Beijing Genomic Institute (BGI), China, which offered service for whole genome sequencing (WGS). The sequencing

was performed by Applied BiosystemSOLiD 4.0 sequencing system. The primary sequencing data were analyzed by using standard SOLiD analysis workflow. After that, secondary analysis was performed by BioScope software, which aligned sequencing reads to human genome reference sequence (UCSC hg18).

For whole exome sequencing (WES), we sent both twins' genomic DNA to Macrogen, Inc., South Korea. The sequencing was done using Illumina HiSeq2000 Sequencer system. The primary sequencing data was analyzed by Real Time Analysis (RTA) software version 1.7, which performs base calling and quality scoring. After that, secondary analysis was done by Burrows-Wheeler Alignment (BWA) tool, which aligned reads to UCSC hg19.

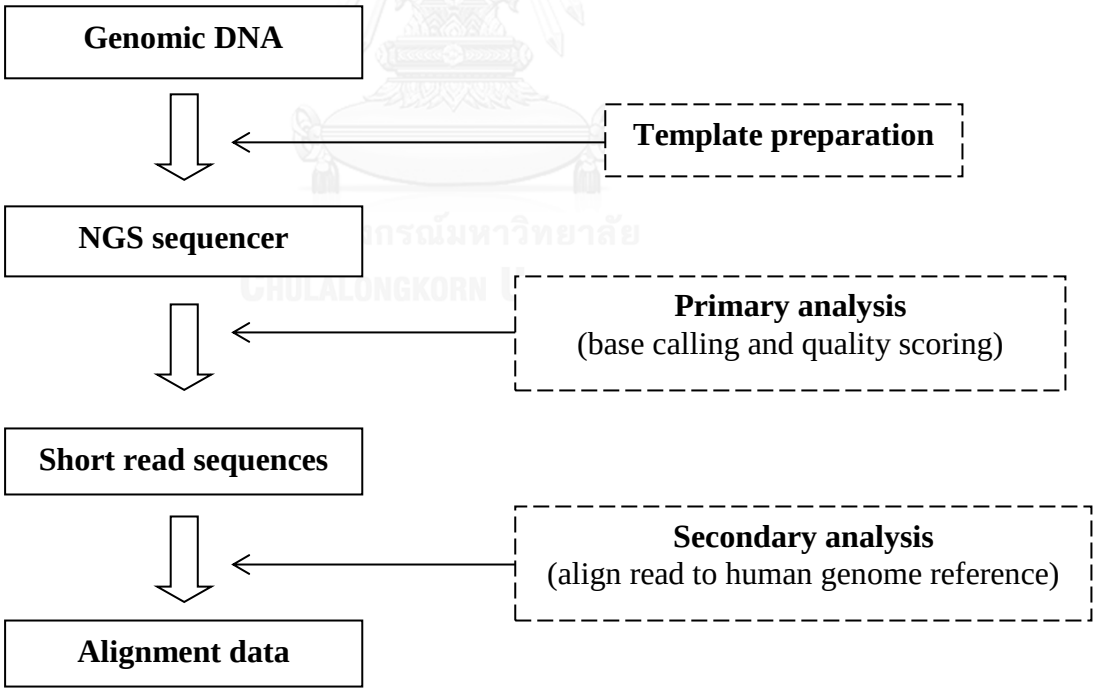


Figure 6. Genome and exome sequencing workflow

Discordant SNVs/Indels analysis

For genome sequencing data, the candidate single nucleotide variants (SNVs) and small insertions or deletions (Indels) were extracted by using SAMtools to compare the twins' alignment data (BAM files) with genome reference. In order to avoid false positive variants, the variant that could only be observed in one twin, but missing due to no coverage in the other twin were excluded from the candidate discordant variant set.

To identify discordant variant, VarScan version 2.2.5 was used for identifying SNVs by comparing, base quality, read counts and allele frequency between the twins' data. The discordant SNVs were called with minimum variant base quality score more than 15 and must have at least three variant-supporting reads. They also had sequencing read depth greater than or equal to 10X. To detect Indels, Genome Analysis Toolkit (GATK) version 1.0.5974 was used to with Somatic Indel Detector command line. The discordant Indels were called if Indels were detected in only one twin. Then we filtered both discordant SNVs and Indels by excluding synonymous variants and/or variants located outside the exonic regions to exclude likely non-functional variants.

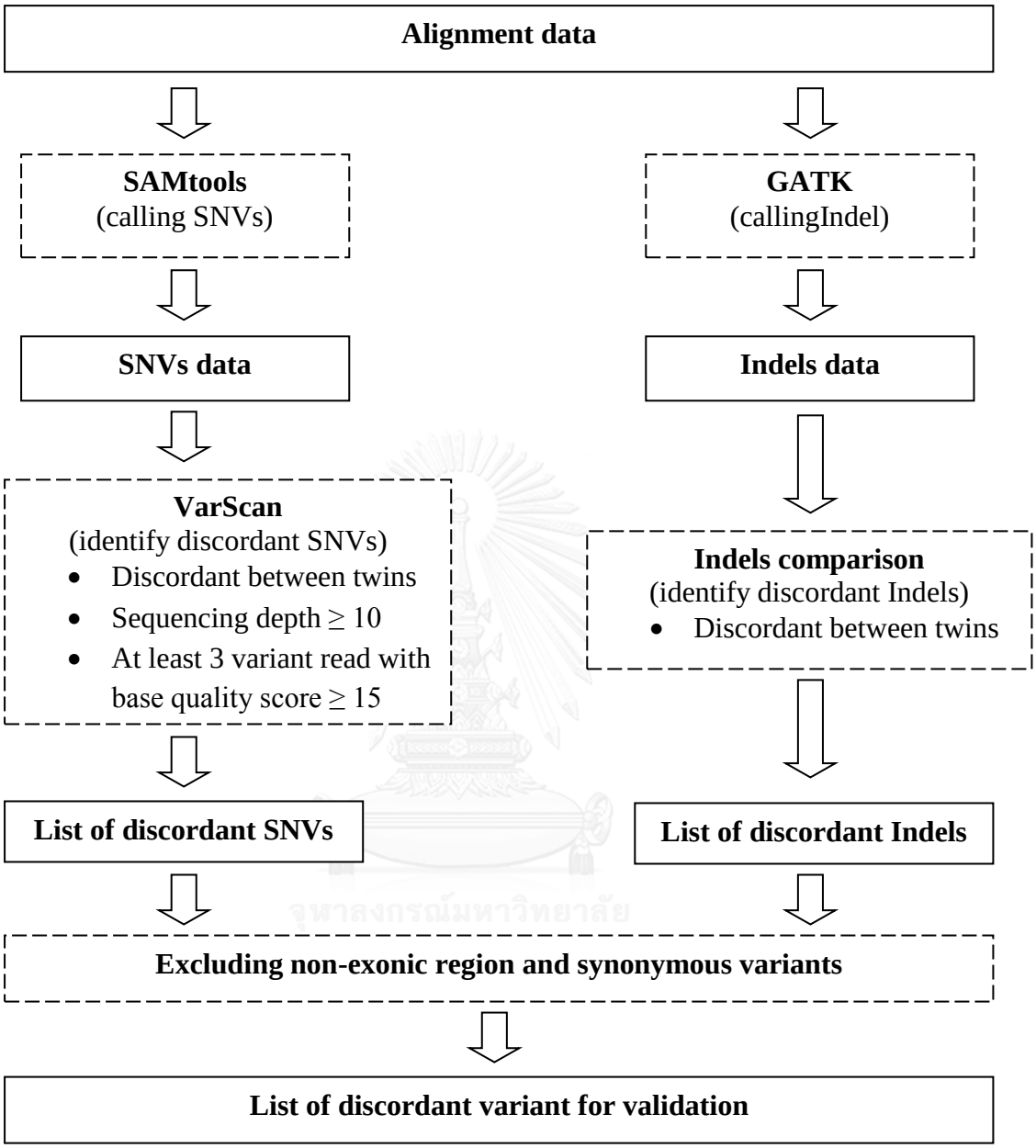


Figure 7. Discordant SNVs/Indels analysis workflow from genome data

For exome sequencing data, alignment files for each twin were created using SAMtools as standard analysis service from Macrogen. Then we use in-house variant calling script to call SNVs and Indels from each twin's alignment data. We compare side by side at each locus of the variants to detect discordant variants. Each locus must have depth coverage at least 30X and have base quality score more than 15 for

both twins. After that, we picked variants with variant-supporting reads more than four reads in one twin but have less than or equal to one read in the other.

After we got discordant data from previous steps, we use Integrative Genomics Viewer (IGV) software version 2.1 to manually visualize exome alignment data at each discordant variants locus. They were selected when they were not located at the start or end position of reads, and when only two haplotypes were found.This was performed to exclude obvious false positive variants.

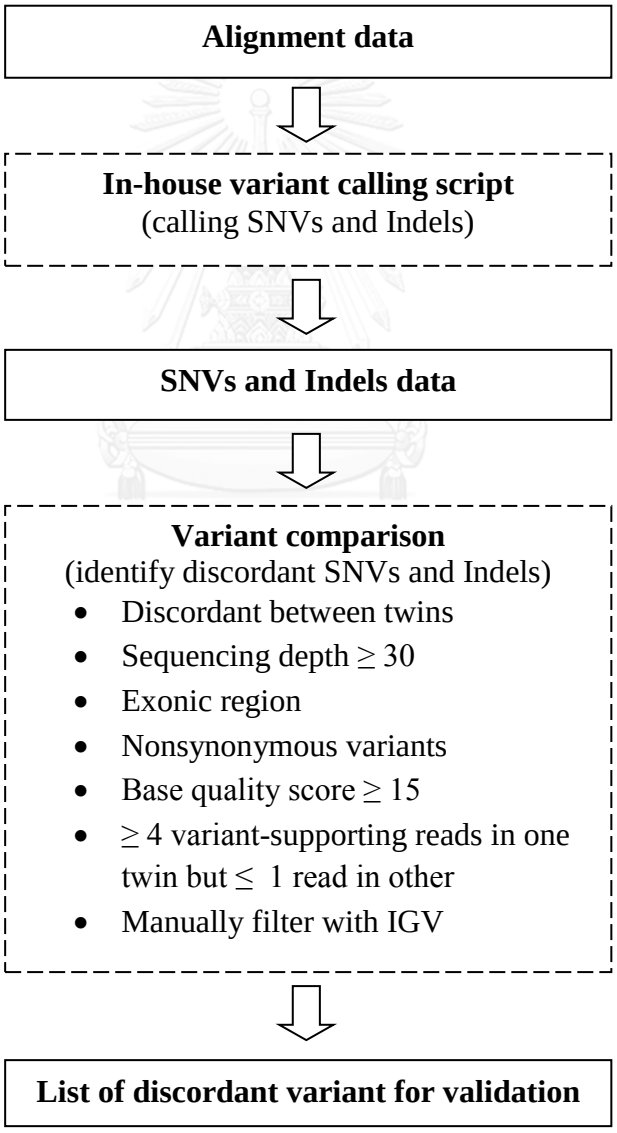


Figure 8. Discordant SNVs/Indels analysis workflow from exome data

Discordant SNVs/Indels validation

We amplified twins' genomic DNA by polymerase chain reaction (PCR), using primers specific to locus of discordant SNVs/Indels from previous steps. The following table showed the PCR mixture and condition. Duration of extension time was chosen based on the product size of the PCR reaction. PCR products were treated with ExoSAP-IT (USP Corporation, Cleveland, OH) at 37 °C for 40 minutes, then inactivation at 80 °C for 30 minutes. They were sent out for Sanger sequencing by Macrogen, Inc., South Koran. Finally, sequencing results were compared between the two twins to verify the discordant variants.

Table 4. Reaction mix for Sanger sequencing

Reaction Components	**Volume (µl)**
10X PCR buffer (KCl)	2.0
25mM $MgCl_2$	1.2
10mM dNTP	0.4
10µM forward primer	0.4
10µM reverse primer	0.4
5U/µl Taq polymerase	0.1
50ng/µl genomic DNA	1.0
Sterile deionized H_2O	14.5
Total Mix	20.0

Table 5. PCR condition

	Step	PCR product size	
		<500 bp	**500-1,000 bp**
	Initial step	94 °C / 5 min	94 °C / 5 min
19 cycles	Denature	94 °C / 30 sec	94 °C / 30 sec
	Annealing	65 °C (-0.5°C each cycle) / 30 sec	65 °C (-0.5°C each cycle) / 30 sec
	extension	72 °C / 30 sec	72 °C / 1 min
15 cycles	Denature	94 °C / 30 sec	94 °C / 30 sec
	Annealing	55 °C / 30 sec	55 °C / 30 sec
	extension	72 °C / 30 sec	72 °C / 1 min
	Final extension	72 °C / 5 min	72 °C / 5 min

CHAPTER IV

RESULTS

Zygosity analysis

We performed thirteen microsatellite markers analysis on both twins and their mother. The results showed that both twins have the same size of all thirteen microsatellite markers, which highly suggesting that they were monozygotic twins. They also share half of the microsatellite markers with their mother (table 6, figure 9 and 10).

Table 6. Microsatellite marker size at each locus of the twins and their mother.

Chromosome	**Locus**	**Microsatellite marker size (bp)**					
		Twin A		**Twin B**		**Mother**	
1	D1S2785	173	182	173	182	182	188
2	D2S206	143	143	143	143	143	147
4	D4S424	198	198	198	198	198	203
5	D5S408	251	259	251	259	259	259
7	D7S657	249	263	249	263	249	263
9	D9S164	91	95	91	95	95	99
11	D11S1315	102	102	102	102	102	102
14	D14S74	304	313	304	313	306	313
15	D15S127	130	156	130	156	130	152
16	D16S515	333	342	333	342	342	342
18	D18S1161	230	230	230	230	230	230
20	D20S117	180	180	180	180	178	180
21	D21S1914	269	271	269	271	269	271

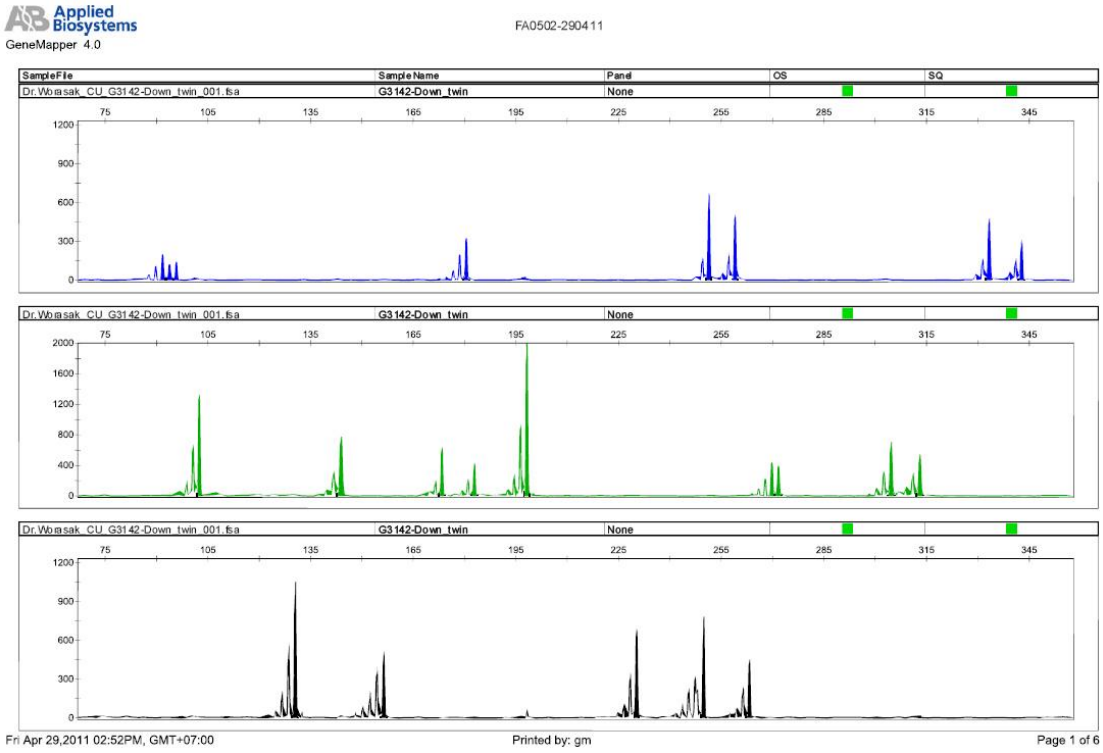


Figure 9. Twin A microsatellite marker size plot graph.

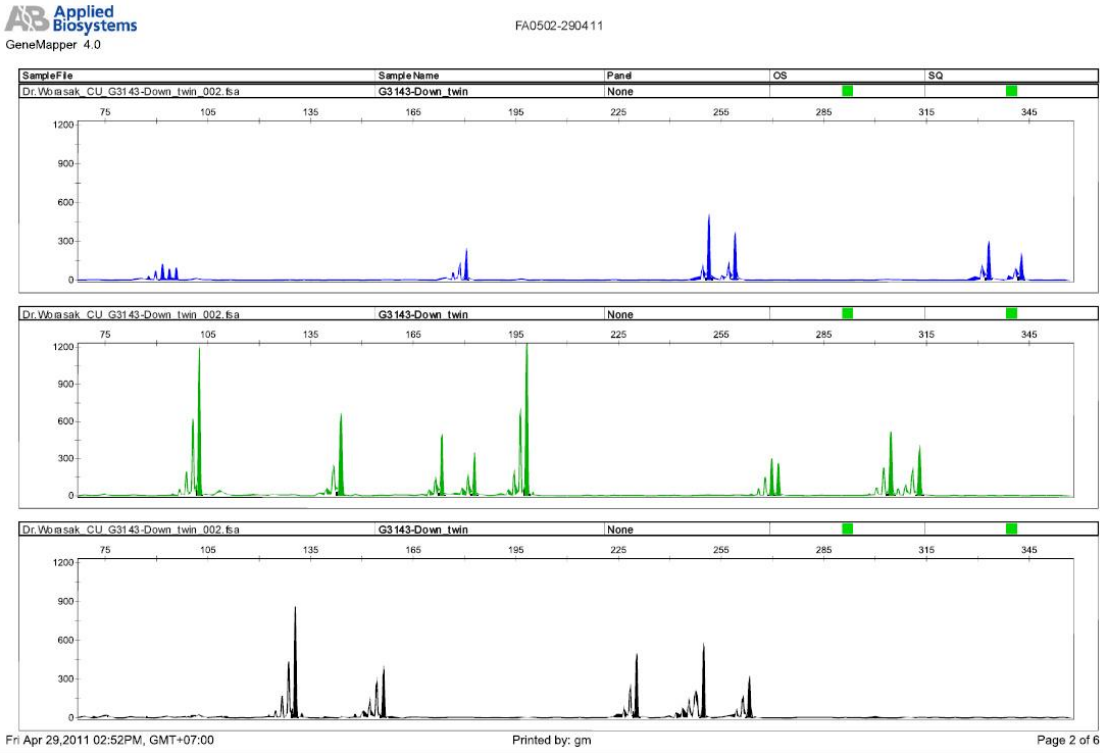


Figure 10. Twin B microsatellite marker size plot graph.

Discordant SNVs/Indels analysis

The genome sequencing dataset from twin A and B showed 27.22X and 28.67X of average read depth with 95.4% and 96.3% of the genome covered with at least 10X, respectively. VarScan and GATK gave the discordant SNVs and Indels in total 5,701 variants. After excluding synonymous and non-exonic variants, two SNVs and six Indels were obtained for further validation from genome sequencing data.

The exome sequencing dataset from both twins showed 44.8X and 36.7X of average read depth with 83.6% and 82.5% of the exome covered with at least 10X, respectively. Prior to the discordant analysis, exome data from both twins resulted in a total of 226,983 variants. By selecting only discordant variants with average sequencing depth ≥ 10X, we obtained 34,226 discordant variants. We applied the stringent filtering criteria which resulted in 7 discordant SNVs and no Indels. These discordant variants from exome dataset share no common with the ones from genome sequencing dataset.

Table 7.Summaryof sequencing results from whole genome sequencing data

SOLiD 4.0 whole genome sequencing	**Twin A**	**Twin B**
Size of genome (UCSC hg18)	2.8 Gb	2.8 Gb
Data mapped to genome (base pair)	78.4 Gb	82.6 Gb
Mean read depth of whole genome	27.22X	28.67X
% Coverage of target regions (> 10X)	95.4%	96.3%

Table 8.Summaryof sequencing results from exome sequencing data

HiSeq 2000 whole exome sequencing	**Twin A**	**Twin B**
Size of exome (UCSC hg19)	63 Mb	63 Mb
Total yield (base pair)	10,451 Mb	8,464 Mb
On-target base pair (mapped to target regions)	3,439 Mb	2,812 Mb
Mean read depth of target regions	44.8X	36.7X
% Coverage of target regions (> 10X)	83.6%	82.5%

Table 9.Number of discordant variants after applying different exclusion criteria for Whole genome sequencing experiments

WGS filtering criteria (using VarScan and GATK to detect variants)	**Number of discordant variants**
No filtering	5,701
Excluding non-exonic variants (NEV)	11
Excluding NEV and synonymous variants (SV)	8

Table 10.Number of discordant variants after applying different exclusion criteria for whole exome sequencing experiments

WES filtering criteria (using in-house variant calling)	**Number of discordant variants**
No filtering	34,226
Excluding non-exonic variants (NEV)	8,640
Excluding NEV with Average depth from both twins≤ 30 (AD ≤30)	1,070
Excluding NEV, AD ≤30 and synonymous variants (SV)	703
Excluding NEV, AD ≤ 30, SV and Base Quality Score ≤ 15 (BQS≤ 15)	302
Excluding NEV, AD ≤30, SV, BQC≤ 15 and variant-supporting read < 4 reads (V4)	59
Excluding NEV, AD ≤ 30, SV, BQC≤ 15, V4 and false discordant variants eye inspected from IGV visualization (IGV)	13
Excluding NEV, AD ≤ 30, SV, BQC≤ 15, V4, IGV and variants that have variant frequency < 5%	7

Discordant SNVs/Indels validation

Total 15 variants, eight from genome dataset analysis and seven from exome dataset analysis, were selected for validation by conventional Sanger sequencing. Each pair of primer was designed to cover the locus of each discordant variant. Electropherograms of all 15 variants showed no differences between the twins.

Table 11.Primer sequences used for variant validations from genome data analysis

Gene	Primer	Sequence 5'→ 3'	Product (bp)
ARHGAP11A	Forward	GGA GAA GAA TTT GGG AAG CC	585
	Reverse	GTG CCA CCA CAC TCA GCT C	
OVGP1	Forward	TTG GAG CAG GTA CTT CAG CC	1083
	Reverse	GTT TTG ATG CCT GCT TTG C	
ELK3	Forward	AAC ATT TTG GTG CTT CTG CC	1001
	Reverse	AGG ACA ACG TTT TGA GGG AC	
NEFL	Forward	CTG CTC CTG CTT GCC TTT G	501
	Reverse	TGA GCA AGG CTT CAT TTG TC	
NCKAP1L	Forward	CAG GGT TTT CAG GAA TCT GG	275
	Reverse	GAT TCC CGC AAT ACT TCC AC	
INO80E	Forward	TTG TAG TCC TGA CGG CAC AG	525
	Reverse	ACT TGA AAG AGC CCA TGC AC	
SGK110	Forward	GAA GGA ACC AAC TGG TGC AG	957
	Reverse	TGC AAT GTG TTC CCT CTC TG	
KRTAP7-1	Forward	GCC CAC CAG CTT GAG GTA T	468
	Reverse	GCA CAT GGG AAG GTA GGA AG	

Table 12.Primer sequences used for variant validations from exome data analysis

Gene	Primer	Sequence 5'→ 3'	Product (bp)
CAMLG	Forward	TGA GTT CTA GCT GCA ACA GCA T	399
	Reverse	TCT GGA AAG GTA CTG CTC TAG G	
CCT5	Forward	AAA ATG CAG GTT GAG ATG GC	706
	Reverse	TGC CTT TCT ATG CAA TTT GG	
FAF2	Forward	CCT GTT TTG GAA TAC CAC GC	375
	Reverse	AAA TAA TAC ACC CAA CGG GC	
ZNF649	Forward	AAA CCC ATC AGC ACC AAG TC	535
	Reverse	TGT TGA GAT TGC CCT TCT GA	
RICTOR	Forward	CAG GCA TTC TTC CCT TTT CA	435
	Reverse	GAG CAT GAA GAT TGG CCT AAA	
TTN	Forward	TCA TAG CTC CTT GCC ACA GT	443
	Reverse	GCC ATT TTA GCC CTC GAT TT	
ZNF557	Forward	AAC AGA ATT CCT GGG AGA GGA	670
	Reverse	TCA GAT TCG ACC TCC TCC TG	

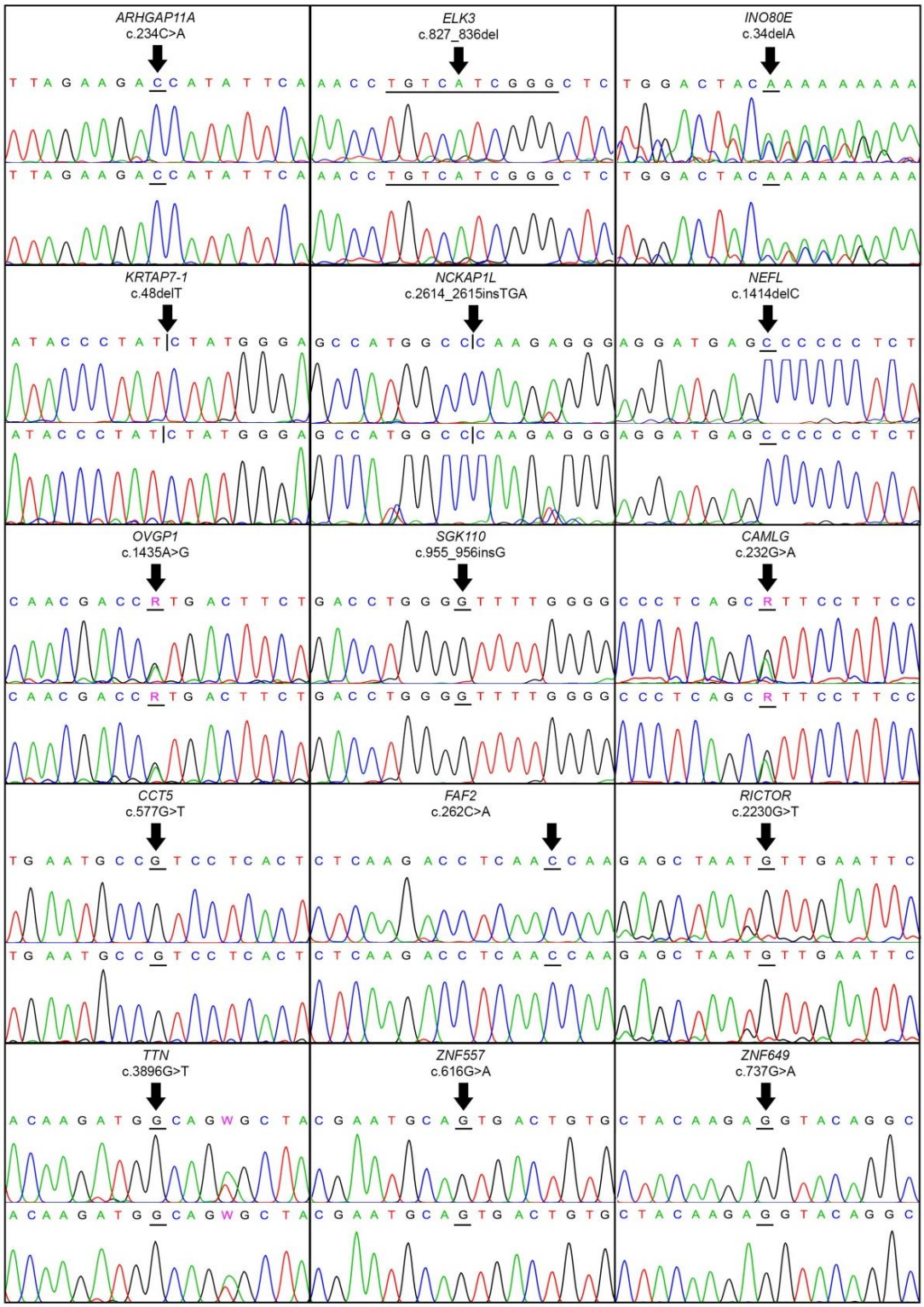


Figure 11.Electropherograms of Sanger sequencing. Upper and lower electropherograms of each panel represented twin A and B, respectively.

Table 13.Details of discordant variants from genome and exomesequencing data.

	Gene	Type	Chromosome	Position	Genotype		Depth of coverage		Variant Frequency	Depth of coverage		Variant Frequency	Sanger sequencing result	
					Reference	Variant	Twin A Reference	Twin A Variant	Twin A	Twin B Reference	Twin B Variant	Twin B	Twin A	Twin B
Genome data	*ARHGAP11A*	missense	15	30703018	C	A	34	12	26%	27	1	4%	C/C	C/C
	OVGP1	missense	1	111759211	A	G	11	13	54%	22	2	8%	A/G	A/G
	ELK3	deletion	12	95165468	TGTCAT CGGG	del	8	2	20%	10	0	0%	TGTCATC GGG	TGTCAT CGGG
	NEFL	deletion	8	24866982	C	del	6	0	0%	6	2	25%	C/C	C/C
	NCK4P1L	insertion	12	53211859	ref	TGA	9	0	0%	10	6	38%	ref	ref
	INO80E	deletion	16	29915166	A	del	11	0	0%	7	2	22%	A/A	A/A
	SGK110	insertion	19	60744149	ref	G	5	0	0%	5	5	50%	G/G	G/G
	KRTAP7-1	deletion	21	31123841	T	del	7	0	0%	4	3	43%	del	del
Exome data	*CAMLG*	missense	5	134076812	G	A	14	16	53%	1	18	95%	G/A	G/A
	CCT5	missense	5	10258269	G	T	31	4	11%	30	0	0%	G/G	G/G
	FAF2	missense	5	175913485	C	A	42	4	9%	36	0	0%	C/C	C/C
	ZNF649	missense	19	52394652	C	T	64	6	9%	56	0	0%	T/T	T/T
	RICTOR	missense	5	38958882	G	T	62	5	7%	64	0	0%	G/T	G/T
	TTN	missense	2	179643775	G	T	52	4	7%	49	0	0%	G/G	G/G
	ZNF557	missense	19	7083178	G	A	69	5	7%	61	0	0%	G/G	G/G

CHAPTER V

DISCUSSION

In this study, we applied next generation sequencing to sequence genome and exome of the monozygotic twins with trisomy 21, discordant for a ventricular septum defect. By applying strictly screening criteria, we obtained 9 SNVs and 6 Indels that potentially causing the discordant phenotype. However, validation of all 15 variants by direct sequencing of the corresponding locus revealed no differences between the twins.

Taking into account that we investigated only non-synonymous discordant variants in the coding regions, it is possible that one of the 5,690 discordant variants, located in the non-coding regions from the genome analysis [27] or one of 367 synonymous coding variants from exome analysis [28] could contribute to the twin discordance.

These negative results support the idea that genetic differences between monozygotic twins are very rare, even with discordant phenotypes. Some of previous studies also failed to identify discordant variants in discordant monozygotic twins. For example, a study by Baranzini (2010) applied three approaches to identify discordant SNVs for monozygotic twins discordant for multiple sclerosis. They use whole genome sequencing (WGS), duplicate array hybridization (DAH) and RNA sequencing (RNA-Seq) which they found 3,241, 126, and 322 discordant SNVs, respectively. Interestingly, no discordant SNVs from one method were replicated by a second method, even though 98% of concordant SNVs could be replicated by at least

two methods. They validated a set of 15 discordant SNVs by direct Sanger sequencing and found that all selected SNVs showed identical genotypes between discordant twins [24]. Another example is a study of monozygotic twins discordant for VACTERL association by Solomon (2013). They applied high-density microarray and WES approaches which also failed to identify discordant variants that could explain the discordant phenotypes [29]. However, a study by Reumers (2012) demonstrates that discordant variants between monozygotic twins do exist. They applied genome sequencing to investigate monozygotic twins discordant for schizophrenia. Their results showed two discordant SNVs which were confirmed as actual differences by Sanger sequencing [30].

In contrast to finding discordant variants, NGS approach has successes in finding concordant variants from patients with the same disease. It showed some successes in identification of the causative genes in a complex disorder like autism. Exome sequencing of 16 probands identify candidate homozygous recessive mutations in four unrelated families [31]. WGS also successfully identified the causative gene in a study of four family members, with unaffected parents and two siblings affected with Miller syndrome and primary ciliary dyskinesia [32].

One possible reason why NGS give high false positive rate for discordant variants is that NGS technology yields high error rate results. Even in leading NGS technologies like Illumina platform, the overall miscall error rate is typically around 1% [33]. Taking into account that the human genome size is around 3 billion base pairs, we could assume that there are approximately 30 million base calling errors in a genome sequencing of a person. This high error rate of genome sequencing process is demonstrated by the study of monozygotic twins discordant for schizophrenia. They

validated all 846 discordant SNVs from genome sequencing with Sanger sequencing but found only 2 SNVs as actual differences [30].

One possible explanation for a pair of monozygotic twins with identical leukocytes' DNA but have discordant phenotype could be mosaic genomic alteration. If a somatic mutation occurs before twinning, both twins will have the same mutation variant. However, level of mosaicism in the relevant tissues of the unaffected twin might not reach the necessary level for clinical expression, which result in discordant phenotypes [34]. In addition to mosaic state, epigenetics could also play an important role for the discordant phenotype in monozygotic twins. It was demonstrated in a genome-wide methylation study in monozygotic twins discordant for schizophrenia and bipolar disorder. Their result showed several disease-associated DNA methylation differences between discordant twin [35]. Another genome-wide DNA methylation analysis in twins discordant for autism also showed a similar result [36].

Of note, it is possible that genomic DNA extracted from leukocyte may not be suitable for NGS studies of discordant twins. Approximately 70% of all monozygotic twins are monochorionic and share blood circulation during pregnancies. Therefore, the hematopoietic stem cells could be transferred between them during embryogenesis and results in chimeric hematopoietic systems. From this reason, it is possible that mutations in one twin could be detected in the co-twin's blood and masks the underlying mutations that cause the disease in the affected twin [37]. Therefore, it is best to sample the tissue with discordant phenotype for DNA extraction. Unfortunately, heart tissues of these twins were unobtainable.

In conclusion, we applied NGS technology to study a pair of monozygotic twin with trisomy 21, discordant for congenital heart defect. By using strictly

screening criteria, we identify 15 discordant variant. However, the validation by Sanger sequencing of all variants showed no differences. The false positive results emphasized the limitation of current NGS technology in identification of rare mutations causing the discordant phenotypes in monozygotic twins.

ADDITIONAL PROJECT

Identification of a mitochondrial 12S rRNA A1555G mutation causing aminoglycoside-induced deafness in a large Thai family

Abstract

Hearing loss is one of the most frequent sensory disorders. It is preventable in genetically susceptible individuals by avoiding certain type of medications. Many families with an A1555G mitochondrial mutation who became profoundly deaf after receiving aminoglycoside were described around the world. However, none had been reported in Thailand. In order to identify the cause of hearing loss of a large Thai family with 11 members, reportedly turned deaf after receiving antibiotics, we obtained blood samples from five members; four of whom had hearing loss. Mutation analyses were performed using molecular techniques including polymerase chain reaction, direct Sanger sequencing and restriction fragment-length polymorphism. The result showed that all four affected members were found to harbor the same A1555G mitochondrial mutation while the unaffected had only the wild-type. In conclusion, identification of the mitochondrial mutation leading to aminoglycoside-induced hearing loss in a Thai population encourages a rise in awareness of medical practitioners in Thailand of this genetic susceptibility. Avoidance of certain medications in these individuals would prevent this acquired permanent hearing loss.

Key words

aminoglycoside, antibiotics, deafness, hearing loss, mutation analysis, mitochondrial DNA, *MT-RNR1*, A1555G, Thai.

Introduction

Hearing loss is one of the most frequent sensory disorders. It affects approximately one in 700-1000 newborns [38,39]. Possible factors included genetic mutations or environmental factors such as ototoxic drugs, for instance, aminoglycoside antibiotics [39]. This type of antibiotics is widely used due to its effectiveness in the treatment against gram-negative infections. They are also easily accessible and inexpensive to produce [40]. Examples of aminoglycoside antibiotics are Amikacin, Gentamicin, Kanamycin, Neomycin, Netilmicin, Paromomycin, Streptomycin, Spectinomycin and Tobramycin [41]. They are called ototoxic drugs because they are well known to be toxic to the vestibular system and auditory organs [41]. Generally, it is safe to use these antibiotics within recommendation dosage. However, for certain genetically susceptible individuals, receiving these medications could result in permanent sensorineural hearing loss. Many families with an A1555G mitochondrial mutation who became profoundly deaf after receiving aminoglycoside were described around the world [42-45]. These patients were found to harbor a maternally inherited A1555G mutation located at a highly conserved region of *mitochondrial encoded 12S rRNA* (*MT-RNR1*) gene in the mitochondrial genome. In these individuals, the aminoglycoside antibiotic causes irreversible hearing loss.

Here, we applied mutation analysis to five family's members from a large Thai family with 11 members reported having post-lingual deafness after receiving Streptomycin.

Methodology

Patients

We identified a large Thai family with 11 members reported having post-lingual deafness after receiving Streptomycin. We had opportunities to examine and obtain blood samples from five members; four were deaf (Fig. 1). At examination III-1, III-3, III-4, III-5 and IV-1 were of 59, 56, 55, 53 and 12 years of age, respectively. The first four members all had severe sensorineural hearing loss immediately after they received prescribed injections to treat infections at the age of 9, 3, 8 and 8 years of age, respectively.

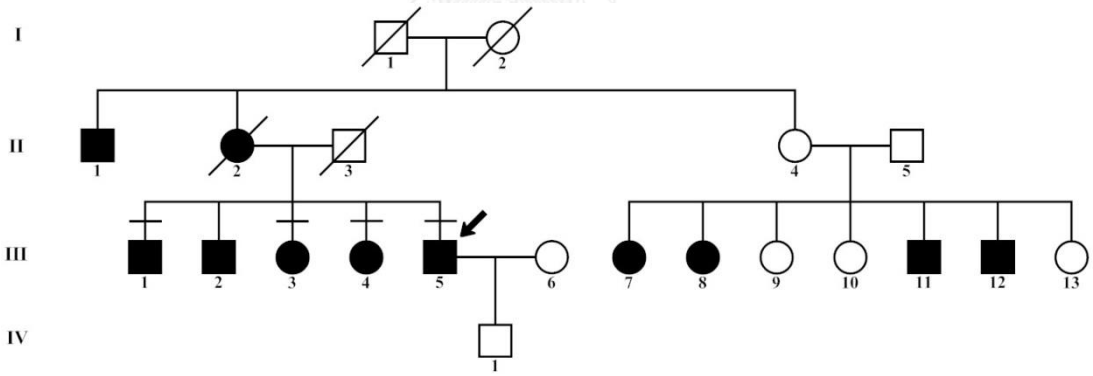


Figure 1. A pedigree of a Thai family with 11 members with aminoglycoside-induced deafness. Arrow, proband; blackened symbols, affected individuals; horizontal bars above symbols, individuals clinically examined.

MT-RNR1 gene analysis

Six milliliters of peripheral blood were obtained from each individual after the informed consent was obtained. Genomic DNA was isolated from leukocytes by using QIAamp® DNA blood mini kit according to manufacturer's instruction (Qiagen, Valencia, CA). PCR-amplification of the *MT-RNR1* from genomic DNA was performed using primers MTRNR1-F 5'-TGG CCA CAG CAC TTA AAC AC-3' and MTRNR1-R 5'-ACT ATA TCT ATT GCG CCA GG-3'.We used 50ng of genomic DNA, 1x PCR buffer (Fermentas, Thermo Fisher Scientific, Inc, Waltham, MA), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM of each primer, and 0.5 U of Taq Polymerase (Fermentas, Thermo Fisher Scientific, Inc.) in a total volume of 20 µl. Following parameter were used in PCR reaction: Initial denaturation at 94°C for 3 min, 35 cycles of denaturing at 94°C for 45s, annealing at 62°C for 45s and extension at 72°C for 1:30 min, then final extension at 72°C for 5 minutes. PCR product of III-5 was treated with ExoSAP-IT (USP Corporation, Cleveland, OH), according to the company recommendations, and sent for direct Sanger sequencing (Macrogen INC., Seoul, Korea). Sequence data were analyzed using Sequencher (version 4.2; Gene Codes Corporation, Ann Arbor, MI).

Presence of the A1555G mutation in other family members was determined by restriction fragment-length polymorphism (RFLP) analysis. The PCR products were digested with the restriction enzyme *BsmA*I (New England Biolabs, Beverly, MA), according to the manufacturer's instructions. Size of PCR product was analyzed by using gel electrophoresis. We used 1% agarose gel with Tris-borate buffer and visualized by Geldoc (Bio-Rad Laboratories Inc., CA).

Results

Sanger sequencing of III-5 revealed a homoplasmic A>G mutation at nucleotide position 1555 of the *MT-RNR1*(Fig. 2). PCR-RFLP showed that the other three affected members, III-1, III-3 and III-4 were all harbored the homoplasmic A1555G mutation (Fig. 3) while the unaffected IV-1 had only the wild-type nucleotide.

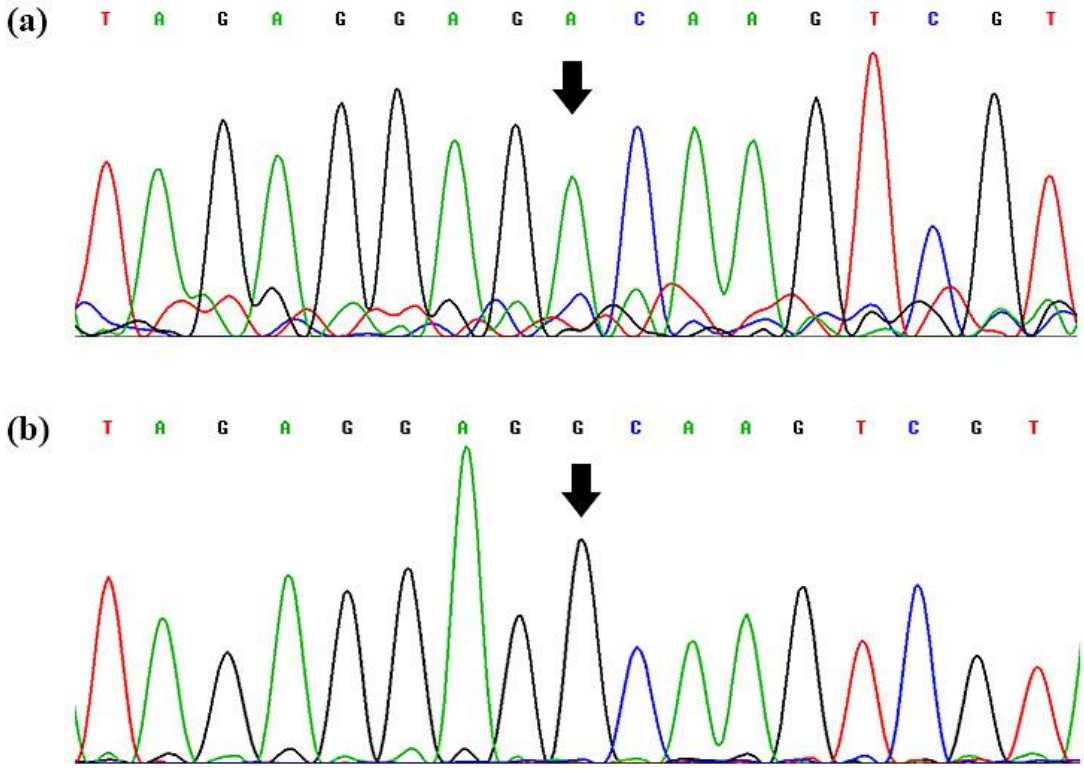


Figure 2. Mutation analysis of the *MT-RNR1* gene. The sequence electropherogram of the *MT-RNR1* gene from (a) a Thai healthy control and (b) III-5 which shows homoplasmic A>G mutation at position 1555 of human mitochondrial DNA.

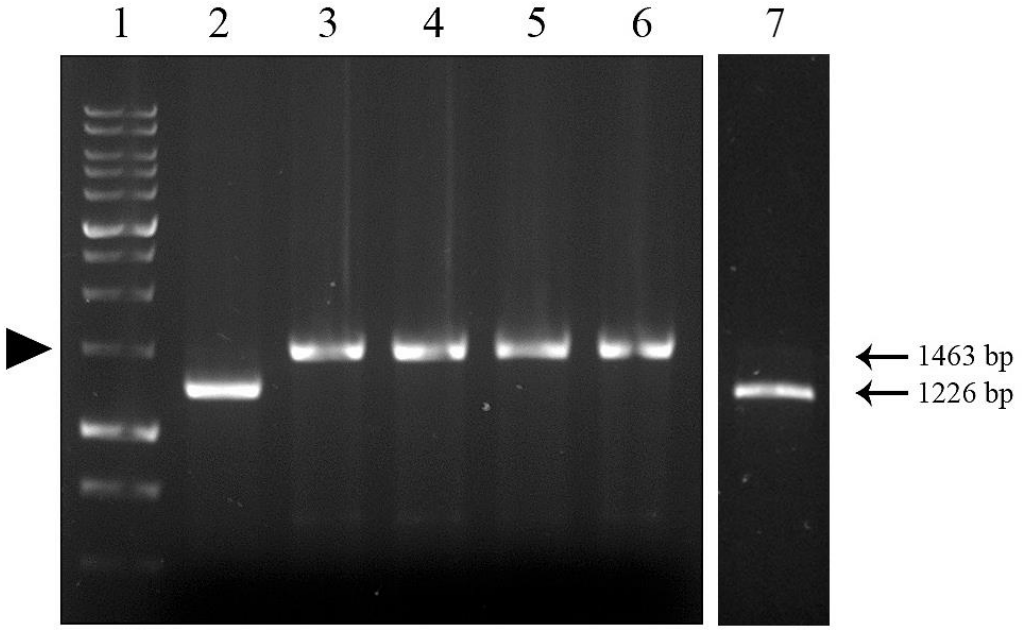


Figure 3. Restriction fragment length polymorphism analysis of the *MT-RNR1*gene. Lane 1 represents a 1000-bp marker with the band 1500-bp indicated by an arrow head. Lane 2 is a Thai healthy control showing a 1266-bp fragment, resulting from a restriction enzyme *BsmA*I digestion. Lanes 3-6 are III-1, III-3, III-4, III-5 which show undigested 1463-bp PCR products, indicating that all four patients have homoplasmic A1555G mutation. Lane 7 is IV-1, which show a 1266-bp fragment indicate normal genotype.

Discussion

In this study, we analyzed the A1555G hotspot of *MT-RNR1* gene by performed PCR of the proband, follow by Sanger sequencing. We performed RFLP analysis of this gene to determine the mutation in other family members. The mutation analyzes showed that this family had the mitochondrial genetic susceptibility for aminoglycoside-induced deafness.

This genetic susceptibility is located on human mitochondrial DNA; therefore it is maternally inherited transferring from mothers to their children. With a prevalence of almost 1 in 400 individuals in British populations, [46] the A1555G mitochondrial mutation could be considered common in some ethnics. These people could remain healthy with normal hearing throughout their entire life as long as they are not given aminoglycosides [46]. Despite an observation that the A1555G mutation could cause hearing loss even without exposure to aminoglycosides, the hearing loss in these cases tends to be late onset and less severe compared to those exposed to such antibiotics [47]. On the other hand, a child with the A1555G mutation was recently reported to have normal hearing despite repeated exposure to aminoglycosides [48]. Therefore, factors affecting penetrance of this genetic susceptibility require further studies.

Aminoglycosides perform their antibacterial effects by binding to the 30S ribosomal subunit of bacterial ribosomes, which alter their conformation. This result in codon misreading of RNA, induces errors in protein synthesis, and leads to bacterial death [49]. Because structural differences lower the drug's affinity for eukaryotic ribosome, aminoglycosides do not normally bind to human ribosome and therefore are generally safe for human use [41]. Current hypothesis is that, mutation at

position 1555 from adenine to guanine of human *MT-RNR1* gene causes a structural rearrangement, which might increases structural similarity of human mitochondrial rRNA to bacterial rRNA, promoting aminoglycoside binding. From this reason, it can lead to a misreading in mitochondrial protein synthesis, which decreases mitochondrial ATP synthesis, compromises ion pump activity, and progressively decreases endocochlear potential, ultimately resulting in hearing loss [50].

Conclusion

This study demonstrates that the A1555G mitochondrial mutation of familial aminoglycoside-induced deafness, which found worldwide, is also presented in a Thai population. Therefore, it is important for medical practitioners in Thailand to be aware of this mitochondrial DNA mutation. This factor should be considered before aminoglycosides prescription, especially when there is deafness reported in a patient's family history. Patients with this mutation should be prescribed with an alternative antibiotic to avoid preventable permanent hearing loss.

Acknowledgements

This study was supported by Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0071/2552), Ratchadapiseksomphot Endowment Fund of Chulalongkorn University (RES560530177-HR), and the Thailand Research Fund.

REFERENCES

1. Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nature reviews. Genetics.* Oct 2004;5(10):725-738.
2. Antonarakis SE. 10 years of Genomics, chromosome 21, and Down syndrome. *Genomics.* Jul 1 1998;51(1):1-16.
3. Langdon J, Down H. Observations on an ethnic classification of idiots1. *Heredity.* 1966;21(4):695-697.
4. Roizen NJ, Patterson D. Down's syndrome. *Lancet.* Apr 12 2003;361(9365):1281-1289.
5. Brickell KL, Leverenz JB, Steinbart EJ, et al. Clinicopathological concordance and discordance in three monozygotic twin pairs with familial Alzheimer's disease. *Journal of neurology, neurosurgery, and psychiatry.* Oct 2007;78(10):1050-1055.
6. Ketelaar ME, Hofstra RMW, Hayden MR. What monozygotic twins discordant for phenotype illustrate about mechanisms influencing genetic forms of neurodegeneration. *Clinical Genetics.* 2012;81(4):325-333.
7. Maiti S, Kumar KHBG, Castellani CA, O'Reilly R, Singh SM. Ontogenetic <italic>De Novo</italic> Copy Number Variations (CNVs) as a Source of Genetic Individuality: Studies on Two Families with MZD Twins for Schizophrenia. *PLoS ONE.* 2011;6(3):e17125.
8. Nachman MW, Crowell SL. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics.* September 1, 2000 2000;156(1):297-304.
9. Aparicio SAJR, Huntsman DG. Does massively parallel DNA resequencing signify the end of histopathology as we know it? *The Journal of Pathology.* 2010;220(2):307-315.
10. Kuhlenbäumer G, Hullmann J, Appenzeller S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Human Mutation.* 2011;32(2):144-151.
11. Imelfort M, Batley J, Grimmond S, Edwards D. Genome Sequencing Approaches and Successes

Plant Genomics. In: Gustafson JP, Langridge P, Somers DJ, eds. Vol 513: Humana Press; 2009:345-358.

12. Sadowski SL. Congenital Cardiac Disease in the Newborn Infant: Past, Present, and Future. *Critical Care Nursing Clinics of North America.* 2009;21(1):37-48.
13. Minette MS, Sahn DJ. Ventricular septal defects. *Circulation.* Nov 14 2006;114(20):2190-2197.
14. Korenberg JR, Bradley C, Disteche CM. Down syndrome: molecular mapping of the congenital heart disease and duodenal stenosis. *American journal of human genetics.* Feb 1992;50(2):294-302.
15. Gittenberger-De Groot AC, Bartram U, Oosthoek PW, et al. Collagen type VI expression during cardiac development and in human fetuses with trisomy 21. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology.* 2003;275A(2):1109-1116.

16. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Human molecular genetics.* Oct 15 2010;19(R2):R119-124.
17. Bainbridge MN, Wang M, Burgess DL, et al. Whole exome capture in solution with 3 Gbp of data. *Genome biology.* 2010;11(6):R62.
18. Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. *Briefings in bioinformatics.* Sep 2010;11(5):484-498.
19. Mardis ER. Next-generation DNA sequencing methods. *Annual review of genomics and human genetics.* 2008;9:387-402.
20. Metzker ML. Sequencing technologies - the next generation. *Nature reviews. Genetics.* Jan 2010;11(1):31-46.
21. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* Nov 2008;92(5):255-264.
22. Mardis ER. Next-generation sequencing platforms. *Annual review of analytical chemistry.* 2013;6:287-303.
23. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology.* 2012;2012:251364.
24. Baranzini SE, Mudge J, van Velkinburgh JC, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature.* Apr 29 2010;464(7293):1351-1356.
25. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 2008;456(7218):53-59.
26. Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *The New England journal of medicine.* Apr 1 2010;362(13):1181-1191.
27. Makrythanasis P, Antonarakis SE. Pathogenic variants in non-protein-coding sequences. *Clin Genet.* Nov 2013;84(5):422-428.
28. Kimchi-Sarfaty C, Oh JM, Kim IW, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science.* Jan 26 2007;315(5811):525-528.
29. Solomon BD, Pineda-Alvarez DE, Hadley DW, et al. Exome Sequencing and High-Density Microarray Testing in Monozygotic Twin Pairs Discordant for Features of VACTERL Association. *Molecular syndromology.* Feb 2013;4(1-2):27-31.
30. Reumers J, De Rijk P, Zhao H, et al. Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nature biotechnology.* Jan 2012;30(1):61-68.
31. Chahrour MH, Yu TW, Lim ET, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS genetics.* 2012;8(4):e1002635.
32. Roach JC, Glusman G, Smit AF, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science.* Apr 30 2010;328(5978):636-639.
33. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nature reviews. Genetics.* Jun 2011;12(6):443-451.

34. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. *Nature reviews. Genetics.* May 2013;14(5):307-320.
35. Dempster EL, Pidsley R, Schalkwyk LC, et al. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Human molecular genetics.* Dec 15 2011;20(24):4786-4796.
36. Wong CC, Meaburn EL, Ronald A, et al. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular psychiatry.* Apr 2014;19(4):495-503.
37. Erlich Y. Blood ties: chimerism can mask twin discordance in high-throughput sequencing. *Twin research and human genetics : the official journal of the International Society for Twin Studies.* Apr 2011;14(2):137-143.
38. Mehl AL, Thomson V. The Colorado newborn hearing screening project, 1992-1999: on the threshold of effective population-based universal newborn hearing screening. *Pediatrics.* Jan 2002;109(1):E7.
39. Morton CC. Genetics, genomics and gene discovery in the auditory system. *Human molecular genetics.* May 15 2002;11(10):1229-1240.
40. Schacht J, Talaska AE, Rybak LP. Cisplatin and aminoglycoside antibiotics: hearing loss and its prevention. *Anat Rec (Hoboken).* Nov 2012;295(11):1837-1850.
41. Xie J, Talaska AE, Schacht J. New developments in aminoglycoside therapy and ototoxicity. *Hearing research.* Nov 2011;281(1-2):28-37.
42. del Castillo FJ, Rodriguez-Ballesteros M, Martin Y, et al. Heteroplasmy for the 1555A>G mutation in the mitochondrial 12S rRNA gene in six Spanish families with non-syndromic hearing loss. *Journal of medical genetics.* Aug 2003;40(8):632-636.
43. Li R, Xing G, Yan M, et al. Cosegregation of C-insertion at position 961 with the A1555G mutation of the mitochondrial 12S rRNA gene in a large Chinese family with maternally inherited hearing loss. *American journal of medical genetics. Part A.* Jan 15 2004;124A(2):113-117.
44. Yuan H, Qian Y, Xu Y, et al. Cosegregation of the G7444A mutation in the mitochondrial COI/tRNA(Ser(UCN)) genes with the 12S rRNA A1555G mutation in a Chinese family with aminoglycoside-induced and nonsyndromic hearing loss. *American journal of medical genetics. Part A.* Oct 1 2005;138A(2):133-140.
45. Wang X, Lu J, Zhu Y, et al. Mitochondrial tRNAThr G15927A mutation may modulate the phenotypic manifestation of ototoxic 12S rRNA A1555G mutation in four Chinese families. *Pharmacogenetics and genomics.* Dec 2008;18(12):1059-1070.
46. Rahman S, Ecob R, Costello H, et al. Hearing in 44-45 year olds with m.1555A>G, a genetic mutation predisposing to aminoglycoside-induced deafness: a population based cohort study. *BMJ open.* 2012;2:e000411.
47. Estivill X, Govea N, Barcelo E, et al. Familial progressive sensorineural deafness is mainly due to the mtDNA A1555G mutation and is enhanced by treatment of aminoglycosides. *American journal of human genetics.* Jan 1998;62(1):27-35.
48. Al-Malky G, Suri R, Sirimanna T, Dawson SJ. Normal hearing in a child with the m.1555A>G mutation despite repeated exposure to aminoglycosides. Has

the penetrance of this pharmacogenetic interaction been overestimated? *International journal of pediatric otorhinolaryngology*. Jun 2014;78(6):969-973.

49. Kaul M, Barbieri CM, Pilch DS. Aminoglycoside-induced reduction in nucleotide mobility at the ribosomal RNA A-site as a potentially key determinant of antibacterial activity. *Journal of the American Chemical Society*. Feb 1 2006;128(4):1261-1271.
50. Huth ME, Ricci AJ, Cheng AG. Mechanisms of aminoglycoside ototoxicity and targets of hair cell protection. *International journal of otolaryngology.* 2011;2011:937861.

REFERENCES

[1]

APPENDIX

VITA

Name Mr. Pongsathorn Chaiyasap

Date of Birth July 8th, 1985

Place of Birth Bangkok, Thailand

EDUCATION:

2004-2007 B.Sc. (1st Class Honors, Gold medal), Genetics science (Honors Program), Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok, Thailand.

2010-2015 Ph.D. Student in Biomedical Sciences program, Faculty of Graduate School, Chulalongkorn University, Bangkok, Thailand

SCHOLARSHIP

2010-2015 The Royal Golden Jubilee Ph.D. Program. Chulalongkorn University and Thailand Research Fund (TRF).

HONORS AND AWARDS:

2007 Graduated with 1st Class Honors, Gold medal Department of Botany, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

2014 Certificate of Appreciation as an Intramural Research Trainee, Medical Genetics Branch of the National Human Genome Research Institute, National Institutes of Health, Bethesda, USA

INTERNSHIP

2006 Center of Excellence for Medical Genetic, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

2013-2014 Medical Genetics Branch of the National Human Genome Research Institute, National Institutes of Health, Bethesda, USA

Laboratory training: DNA, RNA, protein extraction; Protein assay; Real-time PCR; Sanger sequencing; Western-blot; Firbroblast Cell culture; Fluorescent staining; Zebrafish handling, mating, embryo collecting, fin-clipping.