WHOLE EXOME SEQUENCING ANALYSIS OF CEREBRAL PALSY PATIENTS WITH UNDERLYING GENETIC FACTORS

NUR ATIKAH BINTI NOR AZHAR

UNIVERSITI SAINS MALAYSIA

2021

WHOLE EXOME SEQUENCING ANALYSIS OF CEREBRAL PALSY PATIENTS WITH UNDERLYING GENETIC FACTORS

by

NUR ATIKAH BINTI NOR AZHAR

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

March 2021

ACKNOWLEDGEMENTS

In the name of Allah, the most Gracious. All praise to Him for His blessings and gave me strength to complete this research and Peace be upon Prophet Muhammad (s.a.w). My sincere appreciation to the following individuals for their support and constructive critics that inspired me to completing research; Prof. Dr. Nik Soriani Yaacob (main supervisor), Chemical Pathology department, Universiti Sains Malaysia (USM), Assoc. Prof. Dr. Teguh Haryo Sasongko (co-supervisor), Perdana University RCSI School of Medical Sciences, Assoc. Prof. Dr. Salmi Abd Razak, (co-supervisor), Paediatric Department (USM School of Medical Sciences), Dr. Ong Hui San (co-supervisor), Centre of Bioinformatics of Perdana University, Dr. Saidi Jaafar (co-supervisor), School of Dental Sciences, USM and Dr. Nor Azni Yahaya (site supervisor), Paediatric Department of Hospital Raja Perempuan Zainab II (HRPZ II). Thanks to USM for funding the project under RU Top Down (1001.CSKK.870020) grant, Ministry of Education, Malaysia for 1 year MyBrain15 (2017) scholarship scheme and *Perbadanan Tabung Pengajian Tinggi* (PTPTN). Also thanks to those who helped me, namely Dr. Tengku Damitri, Mdm. Nuremi Zahrina, Mdm. Mazlita, Ms. Noor Shahida, Mr. Muhammad Zaman Shafie and Mr. Haris Amirul. Last but not least, many thanks to my beloved parents Mdm. Siti Nariah Binti Mat Salleh and Mr. Nor Azhar Bin Mohd Yusoff, Nur Adnin (sister), Wan Sharifah Azma (cousin), Normala Haron (cousin) and my friends, Dr. Ritzzalina and Dr. Nasyitah for motivating and supporting my journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACMG	American College Medical Genetics Standard Guidelines
AD	Alzheimer's Disease
ADD3	Adducin 3 Gamma Gene
ADGRG3	Adhesion G Protein-Coupled Receptor G3 Gene
AE	Elution Buffer
AFR	African
ALS	Amyotrophic Lateral Sclerosis
aMCI	Amnestic Mild Cognitive Impairment
AMR	Admixed American
ANKRD15	Ankyrin Domain 15 Gene
ANKRD36	Ankyrin Repeat Domain 36 Gene
ANNOVAR	Annotate Variations
ANO5	Anoctamin 5 Gene
AP-4	Adaptor Protein 4 Complex Gene
ApoE	Apolipoprotein E Gene
AR	Autosomal Recessive
ARMadillo	Smaller Sequence Motifs Comprised Of A Series ARM
ASTE1	Asteroid Homolog 1 Gene
BEST3	Bestrophin 3 Gene
BL	Lysis Buffer
BQSR	Base Quality Scores Recalibration
BW	Washing Buffer Solution B
BWA-MEM	Burrows-Wheeler Aligner Software

CA125	Cancer Antigen Gene
CADD	Combined Annotation-Dependent Depletion
CD4+ T	Cluster Of Differentiation A Of T Helper Cell
CDKL2	Cyclin Dependent Kinase Like 2 Gene
cDNA	Complementary DNA
CEP164	Centrosomal Protein 164 Gene
cGMP	Cyclic Guanosine Monophosphate -Dependent Calcium-
CHS1	Chediak-Higashi Syndrome 1 Gene
cM	Centimorgan unit
CNS	Central Nervous System
CNV	Copy Number Variant
COL6A6	Collagen Type VI Alpha 6 Chain Gene
COL7A1	Collagen Type VII Alpha 1 Chain Gene
COOH	Carboxyl Terminus
CpG	Cytosine-Guanine Site
CTBP2	C-Terminal Binding Protein 2 Gene
CTDSP2	Carboxy-Terminal Domain Small Phosphatase 2
CXCR4	Chemokine Receptor Type 4
CYHR1	Cysteine And Histidine Rich 1 Gene
DCHS2	Dachsous Cadherin-Related 2 Gene
DEB	Dystrophic Epidermolysis Bullosa
DLL1	Delta Like Canonical Notch Ligand 1 Gene
DNA	Deoxyribose Nucleic Acid
DNAH17	Dynein Axonemal Heavy Chain 17 Gene
DP	Sequencing Depth/Coverage

DS	Down Syndrome
EAS	East Asian
EB	Epidermolysis Bullosa
EDTA	Ethylenediaminetetraacetic Acid
EE1	Enrichment Elution Buffer 1
EEX	Expanded Exome Oligos
EGF	Epidermal Growth Factor
EHB	Enrichment Hybridization Buffer
EMT	Epithelial-Mesenchymal Transition
ET2	Elute Target Buffer 2
EtBr	Ethidium Bromide
EtOH	Absolute Ethanol
EWS	Enrichment Wash Solution
FAM104B	Family With Sequence Similarity 104 Member B Gene
FAM163A	Family With Sequence Similarity 163 Member A Gene
FASTA	FAST-AII Reported Sequenced Format
FASTQC	Reported Sequence Format With Quality Control
FIN	Finnish
FMR1	Fragile X Mental Retardation 1 Gene
FS	Fisher Strand
FXR2	FMR1 Autosomal Homolog 2 Gene
GABA	Gamma Aminobutyric Acid Gene
GAD1	Glutamate Decarboxylase 1 Gene
GATK	Genome Analysis Toolkit
gDNA	Genomic DNA

GERP++_RS	Genomic Evolutionary Rate Profiling Rejected Substitution tool
GLYR1	Glyoxylate Reductase 1 Homolog Gene
GPCR	Adhesion G Protein-Coupled Receptor
GPR97	Adhesion G Protein-Coupled Receptor Gene
GRCh37	Genome Reference Consortium Human Genome Build 37
GVCF	Genome Variant Calling Format
HapMap	Haplotype Mapping
HEAT	Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A) regulatory A subunit, yeast-kinase (TOR1)
hg 19	Reference Human Genome 19 Database
HGVS	Human Genome Variation Society Database
HLA-DRB1	Human Leukocyte Antigen-Major Histocompatibility Complex Class II Beta Chain 1 Gene
HLA-DRB5	Human Leukocyte Antigen-Major Histocompatibility Complex Class II Beta Chain 1V Gene
hnRNPG	Heterogeneous Nuclear Ribonucleoprotein G gene
HRPZ II	Hospital Raja Perempuan Zainab II
HSP	Hereditary Spastic Paraplegia
HUSM	Hospital Universiti Sains Malaysia
ID	Identification Code
IFT	Intraflagellar Transport
IgG1	Immunoglobin G1
IL6	Interleukin-6 Gene
Indels	Insertion Or Deletion Mutation
iNOS	Inducible Nitric-Oxide Synthase Gene
ITPR1	Inositol Triphosphate Receptor Gene

KANK1	Kank N-Terminal Motif And Ankyrin Repeat Domain 1 Gene
KCNC3	Voltage-Gate Potassium Channel Activity Kv3.3 Gene
KCNK18	Polymorphism Phenotyping Potassium Two Pore Domain Channel Subfamily K Member 18 Gene
KCNQ2	Potassium Voltage-Gated Channel Subfamily Q Member 2 Gene
KCNQ3	Potassium Voltage-Gated Channel Subfamily Q Member 3 Gene
KRTAP19-6	Keratin Associated Protein 19-6 Gene
LECs	Lymphatic Endothelial Cells
LGMD2L	Limb-Girdle Muscular Dystrophy Type 2L
LOVD	Leiden Open Variation Database
LSD2	Histone-Lysine-Specific Demethylase
LSDBs	Locus-Specific Databases
LTA	Lymphotoxin A Gene
LYST	Lysosomal Trafficking Regulator Gene
MAPRE3	Microtubule Associated Protein RP/EB Family Member 3 Gene
MAPs	Microtubule-Associated Proteins
maxGaussians	Maximum Gaussians
MBL	Mannose-Binding Lectin Gene
MBL-54	Mannose-Binding Lectin Of Exon 1 At Codon 54 Gene
MGI	Malaysian Genome Institution
MHC	Major Histocompatibility Complex
miRNA	Micro RNA
MMD3	Distal-Miyoshi Myopathy

MQ	Mapping Quality
MQRankSum	Mapping Quality Rank Sum Test
MR	Mental Retardation
MREC	Medical Research Ethics Committee, Ministry Of Health Malaysia
MTOCs	Microtubule-Associated Proteins
MUC16	Mucin 16 Of Cell Surface Associated Gene
NDSP	Neuroblastoma Derived Secretory Protein Gene
NEC2	Nextera Enrichment Capture 2
NEM	Nextera Enrichment Amplicon Mix
NEW1	Nextera Enrichment Well 1
NEW2	Nextera Enrichment Well 2
NFE	Non-Finnish European
NLM	Library Amplicon Mix
NLT	Nextera Library Tagment
NO	Nitric Oxide
NRC HYB	Nextera Rapid Capture Hybridization Program
OD	Optical Density
OS	Operating System
PCDH	Protocadherin Gene
PCR	Reverse Transcription Polymerase Chain Reaction
PPC	Primer Cocktail
pPolyphen	Polymorphism Phenotyping
PVL	Periventricular Leukomalacia
QD	Quality By Depth
RBMX	RNA Binding Motif Protein X-Linked Gene
ReadPosRankSum	Read Position Rank Sum Test
RSB	Resuspension Buffer

SAMtool	Sequence Alignment/Map Tool
SAS	South Asian
WES	Whole Exome Sequencing
SIFT	Sorting Intolerant From Tolerant
SMB	Streptavidin Magnetic Beads
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SOR	Strand Odds Ratio
SPB	Sample Purification Beads
SPTBN2	Spectrin Beta Non-Erythrocytic 2 Or Beta-III Spectrin Gene
SSH	Secure Shell
ST	Stop Tagment Buffer
SVD	Spontaneous Vaginal Delivery
TBE	Tris-Borate-Edta
TCRs	T Cell Receptors
TD	Tagment DNA Buffer
Th1	Type 1 Helper T Cell
TNF-2	Tumor Necrosis Factor-2 Gene
TNF-ß	Tumor Necrosis Factor-Beta Gene
TNF-α	Tumor Necrosis Factor-Alpha Gene
TRESK	TWIK-Related Spinal Cord Potassium (K+) Channel
TTC13	Tetratricopeptide Repeat Domain Protein 13 Gene
TW	Washing Buffer Solution T
uBAM	Unmapped Binary Alignment Map
UCSC	University Of California Santa Cruz
UMBI	UKM Medical Molecular Biology Institute

USM-CPRC	Universiti Sains Malaysia-Cerebral Palsy Research Cluster
UTR	Untranslated Region
VQSR	Variant Quality Score Recalibration
VQSRLOD	VQSR Log Odds
VSMC	Vascular Smooth Muscle Cells
YOKUK	Yayasan Orang Kurang Upaya Kelantan

ANALISIS PENJUJUKAN EKSOM MENYELURUH KE ATAS PESAKIT "CEREBRAL PALSY" DENGAN FAKTOR GENETIK

ABSTRAK

"Cerebral palsy" (CP) adalah gangguan disebabkan disfungsi neurologi yang tidaklah bertambah buruk seiring bertambahnya usia namun akan tertampil sebagai kerosakan otot sehingga dewasa. Beberapa faktor risiko CP adalah jangkitan kuman semasa kehamilan, malformasi otak, kelahiran secara instrumental dan lain-lain yang berlaku pada tiga kejadian berbeza (antenatal, intrapartum dan postpartum). Sementara itu, CP idiopatik adalah merujuk kes-kes CP tanpa faktor-faktor risiko yang jelas. Kajian berbentuk penerokaan tanpa cadangan hipotesis bagi lokus spesifik ini dilakukan untuk meneroka genetik CP pada pesakit Kelantan dengan faktor genetik tersembunyi, menggunakan Penjujukan Eksom Menyeluruh (WES). Berbeza dengan sindrom-sindrom gangguan saraf lain, pengaruh ketidaknormalan genomik terhadap kejadian CP belum banyak dikaji walaupun ia menyumbang 70- 80% bagi kes dengan faktor pranatal. Tambahan pula, Malaysia belum membangunkan pangkalan data kelaziman dan genetik CP. Justeru, pengenalpastian punca genomik akan meningkatkan pemahaman punca CP dalam keluarga-keluarga terlibat. Keseluruhannya, terdapat 20 subjek iaitu 10 individu CP dan 10 ibu bapa normal. Sampel DNA yang dijujuk kemudiannya dianalisis menggunakan protokol bioinformatik Genome Analysis Toolkit (GATK) untuk mendapatkan semua varian berkemungkinan. Kami menemui 37 varian unik iaitu varian de novo (n = 29) dan varian perwarisan (n = 8) daripada sejumlah 29 gen dan mengklasifikannya mengikut garis panduan piawai American College Medical Genetic (ACMG). Terdapat sembilan gen berkait sistem saraf (*CDKL2 CEP164, FAM10B, FAM163A, FXR2, KCNK18,*

KCNQ3, MAPRE3 dan *RBMX*) dan mutasi mereka disyaki menyumbang kepada kerosakan sistem saraf pusat (CNS), di mana CP adalah gangguan yang berasal daripada kerosakan "CNS". Sembilan gen berkait imuniti bermutasi (*COL6A6, COL7A1, CYHR1, DLL1, GPR97, HLA-DRB1, HLA-DRB5, LYST* dan *MUC16* pula hanya terdapat pada subjek CP, ini menunjukkan individu CP berpotensi menghidapi masalah imuniti. Justeru, CP berpotensi dikategorikan sebagai gangguan neurologi yang terkait dengan disregulasi imuniti. Seterusnya adalah tiga gen yang berpotensi sebagai gen berkait imuniti (*ANKRD36, DNAH17* dan *TTC13*) di mana fungsi molekul mereka tidak diketahui, namun terlibat dalam sistem imuniti. Terdapat juga lapan gen lain-lain kategori (*ANO5, ASTE1, BEST3, CTBP2, CTDSP2, DCHS2, GLYR1* dan *KRTAP19-6*). Berdasarkan data varian, adalah dihipotesis bahawa genetik CP kekeluargaan adalah gangguan genetik secara heterogen yang mana kedua-dua varian perwarisan dan varian de novo yang terlibat berkemungkinan mengandungi maklumat ramalan genetik penting bagi CP. Ini menggalakkan berbagai kajian di masa hadapan dengan menggunakan sampel yang lebih besar, analisis pengesahan yang lebih banyak dan kajian pengkelasan fungsi untuk memahami lebih lanjut sumbangan gen-gen terlibat ini terhadap CP. Hal ini berkemungkinan membantu pembangunan molekul dadah sasaran dan intervensi terapi CP.

WHOLE EXOME SEQUENCING ANALYSIS OF CEREBRAL PALSY PATIENTS WITH UNDERLYING GENETIC FACTORS

ABSTRACT

Cerebral palsy (CP) is a result of neurological dysfunctions that do not worsen with age but will be manifested as muscle impairments till adulthood. Some CP risk factors are intrauterine infection, brain malformation, instrumental delivery, etc. that occur at three different events (antenatal, intrapartum and postpartum). Meanwhile, idiopathic CP refers to CP cases with unknown risk factors. This exploratory study with no proposed hypothesis of specific loci was carried out to explore CP genetic in Kelantan patients with underlying genetic factors using Whole Exome Sequencing (WES). As opposed to other neurodevelopmental disabilities, contribution of genomic abnormalities to CP occurrence has not been extensively researched despite probably accounting for 70–80% of cases with prenatal causes. Plus, Malaysia has yet to establish CP prevalence and genetic database. Therefore, identification of genomic causes would improve CP causal insight in selected families. In total, 20 subjects consisting of 10 CP individuals and 10 unaffected parents were analysed in the current study. The sequenced DNA samples were analysed using bioinformatics Genome Analysis Toolkit (GATK) to generate all possible variants. We found 37 rare de novo (n=29) and inherited variants (n=8) from a total of 29 genes and classified them according to American College Medical Genetic (ACMG) standard guidelines. There are nine central nervous system (CNS) related genes (*CDKL2, CEP164, FAM104B, FAM163A, FXR2, KCNK18, KCNQ3, MAPRE3* and *RBMX*, CP is a disorder of CNS origin and these genetic mutations are anticipated to contribute to CNS defect. Nine mutated immune-related genes (*COL6A6, COL7A1, CYHR1, DLL1, GPR97, HLA-*

DRB1, HLA-DRB5, LYST and *MUC16*) are present in CP subjects only, which possibly indicates that CP individuals may have potential of developing immune-related problems. Thus, CP may be categorized as immune dysregulation-related neurological disorder. Next are three potentially immune-related mutated genes (*ANKRD36, DNAH17* and *TTC13*) with unknown molecular function, but are associated with the immune system. There are also eight mutated genes of other categories (*ANO5, ASTE1 BEST3, CTBP2 CTDSP2, DCHS2, GLYR1* and *KRTAP19-6*). Based on the variant data we hypothesize that familial genetic CP is genetically heterogenous, consisting of discovered de novo and inherited variants that may contain important predictive information on CP. This encourages future studies with bigger sample size, more validation analysis and functional characterization to further understand the contribution of the related genes to CP. This could help in the future development of molecular targeted drugs and CP therapeutics intervention.

CHAPTER ONE

INTRODUCTION

1.1 Cerebral Palsy Definition

By terminology “Cerebral” refers to the brain, while “Palsy” refers to the weakness or paralysis or lack of muscle control. Therefore, cerebral palsy (CP) is a condition of any brain part anomalies that occur during brain development resulting in muscle control disorders (Parveen et. al., 2018). In 2004, the International Working Groupon the Definition and Classification of Cerebral Palsy defined CP as a group of permanent disorders of movement and posture development resulting in activity limitation of non-progressive disturbances that occurs in the developing foetal or an infant brain. The motor disorders of CP are accompanied by epilepsy, secondary musculoskeletal problem and disturbances of sensation, perception, cognition, communication, and behaviour (Rosenbaum et al., 2007).

Basically, CP is a result of any neurological dysfunctions that do not get worse with age but will be manifested in movement and coordination impairments till adulthood. As a chronic disability of Central Nervous System (CNS) origin (Grether et. al., 1992) this “cerebral palsy” term in a clinical descriptive is also applied to the heterogeneous group of neurodevelopmental disorders in which motor impairments usually co-occur with a range of medical disorders or comorbidities (Moreno-de-Luca et al. 2012). Recently, it is suggested for CP to be defined as a spectrum disorder due to CP itself as a condition of heterogeneous nature in all its aspects of aetiology, severity, treatment options, individual’s trajectories and outcomes (Shevell, 2019).

1.2 Epidemiology of Cerebral Palsy

CP is the most common cause of physical disability in childhood (Kirby et al., 2011). To date, the CP prevalence in Malaysia is not available, this might be due to lack of CP studies locally. However, based on direct communication (via e-mail) with the Department of Social Welfare Malaysia, up until February 2018 a record of 12908 CP cases (two years old to 19 years old) have been registered from all around Malaysia (Malaysia Social Welfare, 2018). Several studies from Asian countries showed different CP prevalence. In Bangladesh, the CP prevalence for a time-frame of 1998 to 2016 was 3.4 per 1000 live-births which was considered high due to lack of early intervention as a result of substantially delayed diagnosis (Khandaker et. al., 2019).

A study by Liu and team (1999) showed CP prevalence of 1.6 per 1000 live-births in China for 1990 to 1997 time period, which was considered low. This was in accordance with low survival rate of low-birth weight (LBW) children in China (Liu et. al., 1999), whereby LBW is one of the CP risk factors. Multiple studies in Japan reported different CP prevalences, ranging from 1.4 to 2.9 per 1000 CP children (Takeshita et. al., 1989; Touyama et. al., 2008; Koterazawa et. al., 2016; Toyokawa et. al., 2017). These different CP prevalence values vary by study variables or parameters such as children's age, specified period of time, and location-based surveillance. Another study involving the Asian population who were residing in California, United States of America (USA) showed CP prevalence in those Asian Americans was lower than in whites, which was 1.09 in 1000 live-births compared to 1.36 per 1000 live-births (Lang et. al., 2012). The studied Asian American ethnicities were East Asian (Chinese, Japanese, Koreans), Filipino, Indian, Pacific Islander (Guamanians, Hawaiians, and Pacific Islanders),

Samoan or Southeast Asian (Cambodian, Laotian, Thai, Vietnamese) (Lang et. al., 2012). On top of all, a study showed a worldwide prevalence of 2–3 per 1000 live-births that has remained stable for more than four decades, despite substantial improvements in obstetric and neonatal care (Clark et. al., 2003). Besides that, CP is shown to be common in males compared to females through multiples studies of different populations and races across the world (Izuora et. al., 1981; Nelson et. al., 1985; Blair et. al., 1988; Laisram et. al., 1992; Jarvis et. al., 2005; O'Callaghan et. al., 2011). This possibly happens due to recessive X-linked chromosome variants in males in which they are more vulnerable to genetic mutation since male has one X chromosome only (Jacquemont et. al., 2014). Indirectly, this finding tells us that genetic factor may also play a role in CP incidence.

1.3 General Causes of Cerebral Palsy

There are three different timings of CP causal events which are antenatal, intrapartum and postpartum. In addition to that, there is also idiopathic causal category sincemore than 30% of reported CP cases are due to unknown risk factors (Rosenbaum et. al., 2003). Antenatal (from inception of pregnancy to the onset of labour) risk factors are abdominal trauma, birth defect, chorioamnionitis, foetal brain malformation, hypoxia, intrauterine growth restriction, intrauterine infection, low birth-weight, maternal fever, multiple gestation, thrombophilia and viral infection (Murphy et. al., 1995; Stanley et. al., 2000; Gibson et al., 2003; Jones et. al., 2007; McIntyre et. al., 2013;). Examples of intrapartum risk factors are abnormal foetal presentation, birth asphyxia, meconium aspiration, prematurity, vaginal breech delivery, prolonged and difficult labour, delivery, instruments, placental abruption and uterine rupture (Murphy et. al., 1995;

Gibson et al., 2003; Jacobsson et. al., 2004; McIntyre et. al., 2013). Risk factors for postpartum are asphyxia, hypoglycaemia, neonatal seizures, cerebral infarction, respiratory distress, neonatal infection (sepsis and meningitis), jaundice and head injury (Stanley et. al., 2000; Han et. al., 2005; Jones et. al., 2007; McIntyre et. al., 2013). Aforementioned risk factors prove that, CP disorder is actually a result of multiple aetiologies. A recent systematic review study had shown ten risk factors that were consistently reported as statistically significant to be causing CP in term born infants. These are birth asphyxia, hypoglycaemia, instrumental/emergency caesarean delivery, LBW, major and minor birth defects, meconium aspiration, neonatal infections, neonatal seizures, placental abnormalities and respiratory distress syndrome (McIntyre et. al., 2013).

1.3.1 Antenatal Risk Factors

According to above mentioned systematic review study by McIntyre et. al., 2013, analyses showed that birth defect risk factor is the highest relative risk for CP with a wide range of central estimates between studies which are 2.6 — 20.7 for all CP types and 2.2 — 27.4 for term CP infant (> 2500 g) category (McIntyre et. al., 2013). Other studies showed a higher incidence of birth defect that is associated with maternal age (≥ 35 years old) and maternal age is also one of CP risk factors (Salem et. al., 2011; Johnson et. al., 2012). The congenital abnormalities include chromosomal abnormalities and other organ anomalies (Johnson et. al., 2012). The congenital anomalies that are highly reported in CP children involve cerebral defects of microcephaly and hydrocephaly, cardiac defect, musculoskeletal defect and urogenital defect (Odding et. al., 2006; Andersen et. al., 2008; Rankin et. al., 2010; Moreno-de-

luca et. al., 2012). Besides that, CP risk increases by 50% in newborns with low birth-weight (< 2500g) compared to newborns with normal weight (Odding et. al., 2006). The risk of having LBW newborn is significantly higher in a pregnant mother with older age (≥ 35 years old) (Johnson et. al., 2012). A case control study was done between the LBW survivors at young adulthood age and control group of young adulthood of normal weight-born that showed the LBW individuals have higher neurosensory impairment ($P < 0.001$) compared to the control group. This neurodevelopmental condition is persistent until their adulthood (Hack et. al., 2002). LBW is also reported to be occurring due to intrauterine growth restriction (IUGR), one of the CP risk factors (Williams et. al., 1997). The IUGR describes a condition of foetus that does not grow at expected rate during pregnancy, which later resulting in the infant with LBW.

The IUGR is also influenced by placental abnormalities risk factor when a foetus does not receive adequate oxygen and nutrient (McIntyre et. al., 2013). Based on the reports, maternal age, IUGR and LBW may together contribute to the development of CP. Normal placenta is essential for a healthy pregnancy by transporting the nutrient and oxygen, removing waste products, protecting from infection and modulating maternal immune system and hormone production to maintain the pregnancy (Heazell A., 2012). However, placenta with abnormal structures of area, thickness, size and volume causes placenta insufficiency (Ohgiya et. al., 2016). This indicates that placenta abnormalities do affect its functions which later disturb the foetus' growth. Small placenta or low weight placenta increases the CP risk in term infants (Strand et. al., 2016)

1.3.2 Intrapartum Risk Factors

Birth asphyxia which is also known as birth trauma describes the difficult vagina delivery event. Over the past three decades, a consistent trend shows that 17% to 24% of CP cases are related to birth asphyxia (Volpe, 2008). Theoretically, birth asphyxia is a clinical situation of foetal distress when a foetus is deprived of oxygen during labour. This contributes to irreversible brain damage effect in CP (Blair et. al., 1988; Stanley et. al., 2000). The criteria for birth asphyxia diagnosis are low Apgar score (≤ 3) for over 5 min, abnormal umbilical artery (pH < 7) and newborn cerebral dysfunction (seizure) (Gilstrap et. al., 1989). However, one in every six to nine CP cases is estimated to reduce brain injury risk by 10% if the asphyxiated newborn receives hypothermia intervention within six hours of the birth asphyxia event (Jacobs et. al., 2013; McIntyre et. al., 2013).

Instrumental deliveries occur secondary to a prolonged second stage of labour, meanwhile the emergency caesarean sections occur due to failure to progress or as a result of foetal stress and possible hypoxia (Philpot et. al., 2018). These show that instrumental and emergency caesarean deliveries are common alternatives taken during a difficult vagina delivery. Multiple association studies had shown that these two risk factors are significantly associated with the increased risk of CP in term infants (Walstab et al. 2004; Stelmach et al. 2005; Thorngren-jerneck et. al., 2006; O'Callaghan et. al., 2013). Meconium aspiration during intrapartum is another statistically significant reported risk factor of CP (Beligere et. al., 2008; McIntyre et. al., 2013). This is considered as presence of foetal distress during the labour. Meconium consists of a mixture of blood, cellular debris, gastrointestinal, hepatic

secretion, lanugo, pancreatic secretions, swallowed amniotic fluid and vernix caseosa. Meconium that presents in the foetus' tracheobronchial can cause airways blocking, interference with alveolar gas exchange, chemical pneumonitis and surfactant dysfunction.

1.3.3 Postpartum Risk Factors

Glucose is important for brain development. An infant with low blood glucose within few days after birth is considered hypoglycaemic. Koh et. al., (1988) had shown that full-term infants with blood glucose level < 2.6 mmol/L resulted reversible injury to the nervous system. This threshold is also proposed for pre-term infant cases (Lucas et. al., 1988). Regardless of different gestational ages and infant birth-weights, blood glucose level below 2.6 mmol/L (< 2.6 mmol/L) has been used as a hypoglycaemia intervention threshold value for infants. Persistent or recurrent hypoglycaemia can cause permanent brain injury called neonatal hypoglycaemia encephalopathy which later results in many impaired conditions such as long-term vision impairment, hearing impairment, cognitive abnormalities, epilepsy (Su et. al., 2012) and central nervous system disorders including CP (Lou et. al., 2010) as available at *http://en.cnki.com.cn/Article_en/CJFDTotal-YXZS201013031.htm*. In fact, these impairments are all CP comorbidities. Hypoglycaemia also contributes to neonatal seizure (Montassir et. al., 2009). The positive correlation between these two CP risk factors of hypoglycaemia and neonatal seizure has made neonatal seizure as a prognostic factor for neonatal hypoglycaemia condition in which both lead to neurological outcome (Koivisto et. al., 1972; Pildes et. al., 1974; Yager et. al., 2002). Neonatal seizure occurrence is estimated at 1 in 4000 term births (Nelson et. al., 2004).

Neonatal infection is developed in infants with cerebral abnormalities (Blair et. al., 1993; Walstab et. al., 2004) that is verified during the Neonatal Intensive Care Unit (NICU) stay.

The example of neonatal infection are pneumonia, septicaemia, meningitis, encephalitis, neonatal urinary tract infection (UTI) or ventriculitis (Kimberlin D, 2004; Ganatra et. al., 2010; Emamghorashi et. al., 2012; Shimol et. al., 2012; Camacho-Gonzalez et. al., 2013). Respiratory distress syndrome in infant is caused by lack of surfactants in the lung, which leads to atelectasis, decreased gas exchange and hypoxia (Thygesen et. al., 2016). This infant respiratory syndrome causes intracerebral haemorrhage/ intraventricular haemorrhage (ICH/IVH) and also periventricular leucomalacia (PVL), which are also risk factors of CP (Kinney et. al., 2006; Sun et. al., 2013). Based on these explanations, it is clear that some risk factors are influencing each other impact, which later contributes to CP pathway.

1.3.4 Genetic Factors As Possible Causal of Cerebral Palsy

Many CP cases with unknown causes are found to be associated with genetic alterations which either directly cause CP or contribute to CP susceptibility (Moreno--de-luca et. al., 2012; McMichael et. al., 2015). As opposed to other neurodevelopmental disabilities, the genomic abnormalities contribution to CP occurrence has not been researched extensively, despite probably accounting for a proportion of 70–80% of cases with prenatal causes. (Moreno-De-Luca et al., 2012). Several lines of evidence support the theory that multiple genetic factors contribute to

the cause of CP. First, mutations in multiple genes result in mendelian disorders that are present with CP-like features (Moreno-De-Luca et al., 2012), and several single-gene mutations have been identified in idiopathic (ie, non-syndromic) CP pedigrees (Lynex et al., 2004; Lerer et al., 2005; Verkerk et al., 2009; Abou Jamra et al., 2011; Moreno-De-Luca et al., 2011). These mutations, however, represent too limited cases to assure generalization. There is a great clinical heterogeneity between cases. This may be due to genetic heterogeneity, whereby different genetic abnormalities as a result of mutations at two or more genetic loci contribute to the same outcome phenotype in different individuals (McGinniss, et. al., 2013).

Second, the prevalence of congenital anomalies in individuals with CP (11–32%) is significantly higher than in the general population (2–3%) (Blair et al., 2007; Garne et al., 2008). Third, register-based studies have reported a significantly higher concordance rate for CP in monozygotic twins than in dizygotic twin pairs ($p = 0.0026$) (Petterson et al., 1990). Fourth, the risk of CP in consanguineous families is about 2.5 times higher than the risk in outbred families (Al-Rajeh et al., 1991; Erkin et al., 2008). Fifth, several studies have reported familial aggregation of CP, including identical CP syndromes in the same family (Gustavson et. al., 1969; Palmer et al., 1994; Amor et al., 2001; Hemminki et al., 2007). Sixth, a paternal age effect has been described in some forms of CP (Fletcher et. al., 1993).

1.4 Early Signs of Cerebral Palsy

As a neurological condition that affect movement and coordination, CP is usually apparent within the first few years of life (Illingworth et. al., 1966). CP diagnosis is made before the age of 2 years for infants or toddlers with CP symptoms (Ashwal et. al., 2004). The major signs for CP diagnosis are delayed motor milestones, abnormal neurological, primitive reflexes persistence and abnormal postural reactions (Jones et. al., 2007). The child is considered delayed in motor milestones if he/she does not show physical abilities such as no rolling over at six months old, not sitting even with support by eight months old and not walking by 18 months old. Other than that, the delayed motor also causes a fisting after five months of age and there are also discrepancies between intellectual and motor development (Bennett et. al., 1999; Jones et. al., 2007).

Neurological abnormalities test is to examine for persisted ankle clonus which is a series of involuntary, rhythmic, muscular contractions, and relaxations beyond 12 months old; brisk deep tendon reflex (hyperreflexia) examination that describes the increase speed in reflex. Examination for muscle tone which is also known as muscle tension control. Two states of muscle tension that are noticeable at rest are called hypertonia and hypotonia which are the increasing and decreasing muscle tone respectively. Another muscle tension is dystonia, a condition of increased muscle tone that is noticeable through repetitive involuntary movement with abnormal posture (Bennett et. al., 1999; Jones et. al., 2007). Abnormal muscle tone leads to abnormal posture reactions such as head lag, poor head control, adducted thumbs, hyperextension trunk, opisthotonus (abnormal posture due to muscle spasm) and lower extremities with hyperextension and scissoring (Bennett et. al., 1999; Jones et. al.,

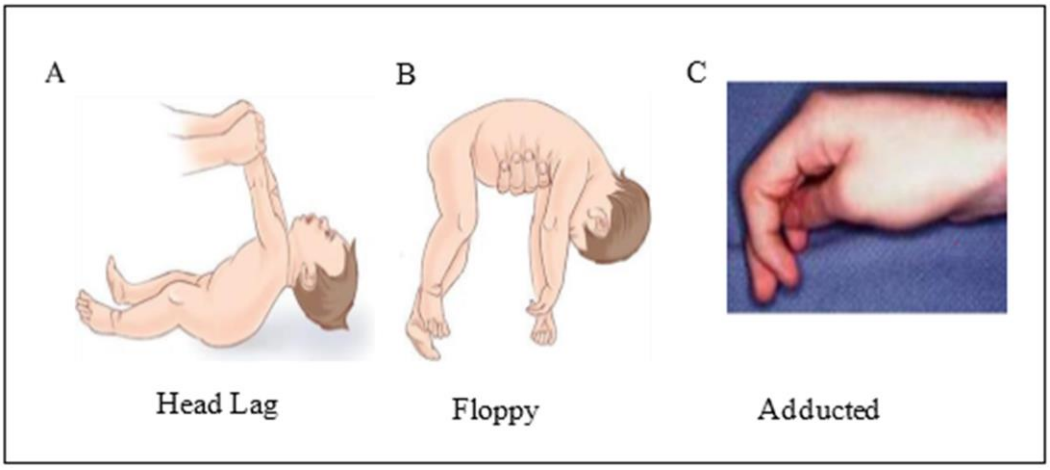


Figure 1.1: The early signs that are obvious in infants or toddlers. A and B were extracted from Carney et. al., 2007 (*https://neurology.mhmedical.com/content.aspx?bookid=459&sectionid=41027558 &jumpsectionid=41030165&caclickthru=129824)*. Meanwhile C was from *https://www.orthobullets.com/pediatrics/4129/cerebral-palsy--upper-extremity-disorders*

The diagnosis also involves primitive reflexes persistence assessment to observe any clinical reflection signs that occur due to malfunction in nervous system control. Instead of supposedly integrated within four to eight months old, the primitive reflexes are observed to be persisted until adulthood. The primitive reflexes are asymmetrical tonic-neck reflex, Moro reflex, extensor thrust on vertical suspension and absence of normal parachute reaction after 11 months (Jones et. al., 2007; Agarwal Et. al., 2012). This persistence later influences the muscle tone change and limbs which causes interference with voluntary motor movement (Bennett et. al., 1999; Jones et. al., 2007).

Both delayed motor milestones and primitive reflexes persistence signs are helpful for diagnosis on CP children with no brain damage incident (Jones et. al., 2007). The non-progressive condition of CP means that the original malfunction in the brain does not get worsen by age. However, CP children's physical manifestations will change with age. This physical changing may be misunderstood to be worsen by time, while it is actually the result of physical feature deficit signs that are detected during diagnosis phase and become more obvious as the CP children grow (Jones et. al., 2007).

1.5 Types of Cerebral Palsy

Motor dysfunction in CP is divided into three major types of CP physiology which are spastic, dyskinetic and ataxic (Figure 1.2). CP is classified according to the damage on different parts of the brain (cerebral cortex, basal ganglia and cerebellar) affecting different parts of the body (Figure 1.2). The detail explanation on Figure 1.2 can be obtained in the following subsection 1.5.1, subsection 1.5.2 and subsection 1.5.3.

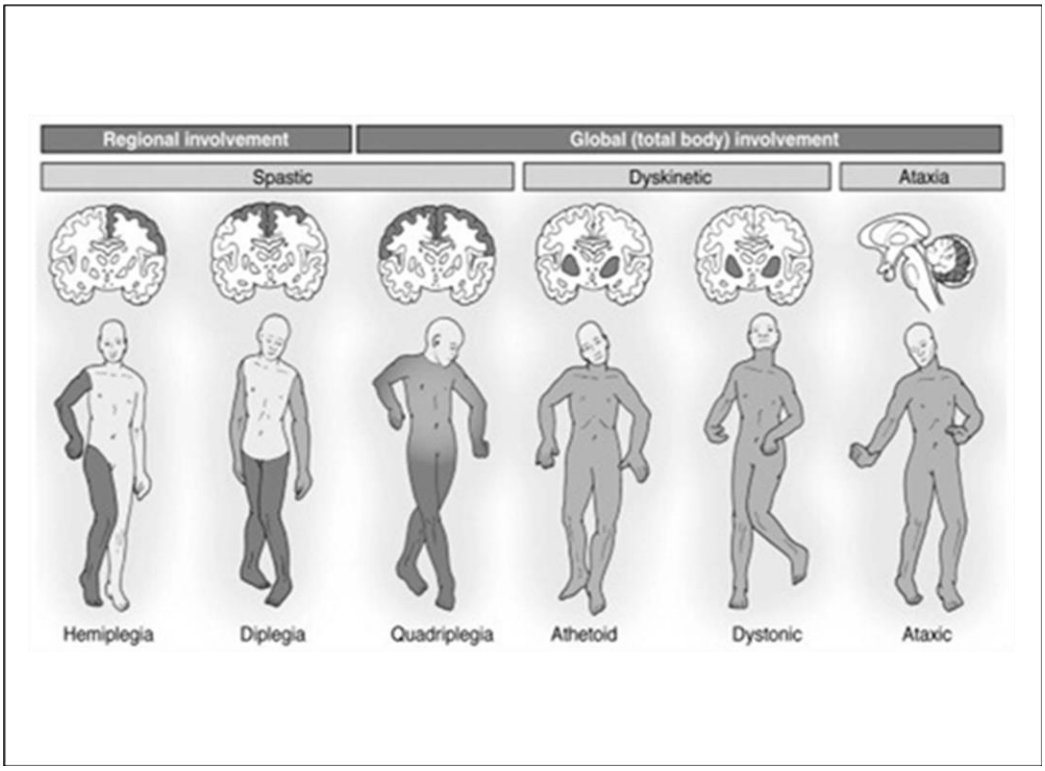


Figure 1.2: The classification of CP according to its affected area distribution (Extracted from *Parvizi, 2010*).

1.5.1 Spastic CP

Spastic CP is the most common CP, diagnosed in 75% cases (Jan et. al, 2006). It is a condition of muscle stiffness due to incorrect relayed messages to the muscle from the malfunction brain parts of cerebral cortex and corticospinal tracts (Norton, 2007) as available at *https://www.sciencedirect.com/topics/pharmacology-toxicology-and pharmaceutical- science/cerebral-palsy*. While performing an activity by a normal muscle state, a group of skeletal muscles is contracting while the opposite skeletal muscle group is relaxing when a normal person bending his/her arm. On a contrary, in spastic CP child, both groups of muscles are contracting together while performing any activities. This is referred as increasing muscle tone which will result in exaggerated reflexes that lead to stiffness, jerky and awkward movement (Ranatunga et. al., 2011) (Figure 1.3 (A)).

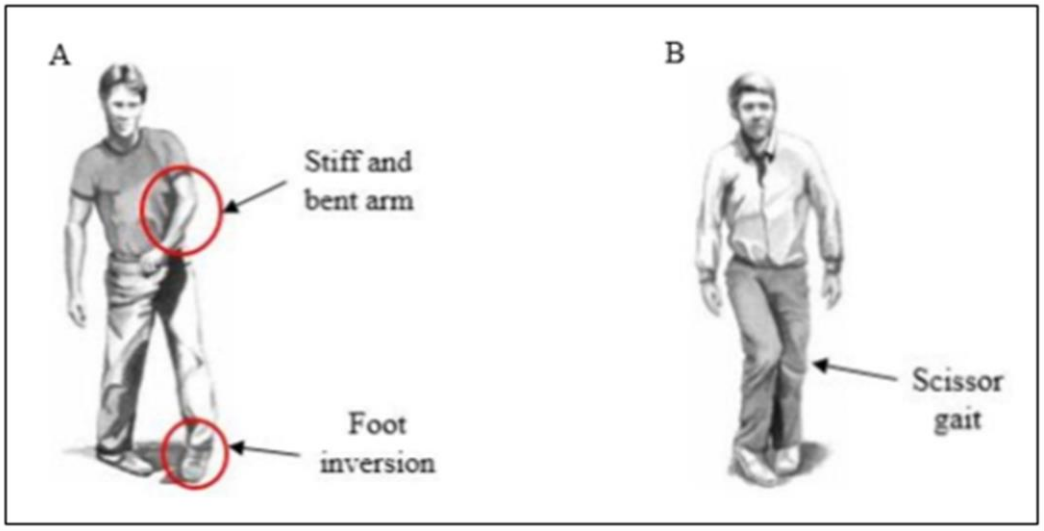


Figure 1.3: The spastic diplegia CP individuals with stiffness and foot inversion (A) and scissor gait (B). (Extracted from Borchers, 2015 at *https://studfile.net/preview/7605613/page:38/*.

Spastic CP is divided into three subtypes according to topographical pattern of affected limbs (refer Figure 1.2). For spastic hemiplegia subtype one side of the body limbs, either left or right side, is affected and the upper limb is worse affected than the lower limb. This subtype is unnoticeable until the child tries to grasp objects (Rapin, 2000; Norton, 2007). Spastic hemiplegic child has a poor motor coordination that causes the unstable walking with tiptoeing, slower speed, obvious swing and longer gait cycle (Wang et. al., 2012). Spastic diplegia subtype affects all four limbs. However, the lower limbs are predominantly affected compared to upper limbs that show slightly affected signs. Diplegic children walk with "scissors gait" but some are unable to walk or immobilized without ambulatory devices (Rapin, 2000) (Figure 1.3 (B)). Spastic quadriplegia is the most severe subtype for a CP child with paresis of all four limbs. The quadriplegic children are unable to mobilize or sit without the help of wheelchair with neck and trunk support (Rapin, 2000).

1.5.2 Dyskinetic CP

Dyskinetic CP (Figure 1.2) is a condition of abnormal involuntary movements due to affected basal ganglia that controls automatic movement, movement related to posture and fine digital movement (Norton, 2007). Dyskinetic child has a probability of unable to speak due to facial dyskinesia, has hearing impairment and less or no use of hands at all, yet may have normal intelligence (Rapin, 2000). Dyskinetic CP is divided into two subtypes which are dystonia and athetosis. Dystonia is a condition of sustained muscle contraction that frequently causes twisting and repetitive movements or maintained-abnormal postures or either both. The dystonic features are foot inversion, wrist ulnar deviation or lordotic trunk postures (Sanger et. al., 2010).

The severity and quality of these involuntary movements are influenced by body position, consciousness level, emotional state and any tasks being attempted by the child (Jones et. al., 2007). Opposite to dystonia, athetosis is a condition of unable to maintain a stable posture. It is also accompanied with a slow and continuous involuntary extra movement, which occur particularly at the arms, feet, hands, face, neck and trunk (Sanger et. al., 2010). It is noticeable when a CP child is particularly starting to move and also will appear floppy when being carried by another person (Aguiar et. al., 2013).

1.5.3 Ataxic CP

The third CP main classification is another non-spastic called ataxic CP (Figure 1.2) which is the least common type of CP accounting for less than 1 in 10 CP individuals (Chen et. al., 2020). The ataxic CP is due to abnormalities in cerebellum (Rapin, 2000). An ataxic child has body imbalance issue which leads to shaky and tremor movement (Aguiar et. al., 2013) and it is apparently observed in gait and trunk (Rapin, 2000). Besides these three main types, there are also cases of CP children with mixed CP types due to having more than one type with several movement patterns.

1.6 Gross Motor Function Classification System (GMFCS)

Palisano's team (1997) built up GMFCS to define how severe the movement problem by describing the motor function range from mild to severe in a CP child. There are five levels of GMFCS classification system based on parameters of gross motor skill, truncal control (sitting ability), walking, mobilize with wheelchair and assistive

devices dependent. The CP children within level I are able to walk, climb the stairs without limitation and are also capable of performing gross motor skills such as running and jumping (Figure 1.4). However, the speed, balance and coordination are impaired. Level II category is for CP children who are also able to walk independently however need to hold onto a rail while climbing stairs.

They also experience limitation when walking on uneven surface, in crowds or in a confined space (Figure 1.4). Level III categorizes CP children who walk indoors or outdoors on even surface with assistive mobility devices. They are able to climb stairs while holding onto a rail and also able to propel a wheelchair manually (Figure 1.4). The CP children within level IV walk in a short distance and rely more on walk assistive devices (Figure 1.4). Whereas, level V is for CP children with all motor function areas being affected and are completely dependent on aid devices with no means of independent mobility. The voluntary control of movement, trunk and head are all affected (Figure 1.4).

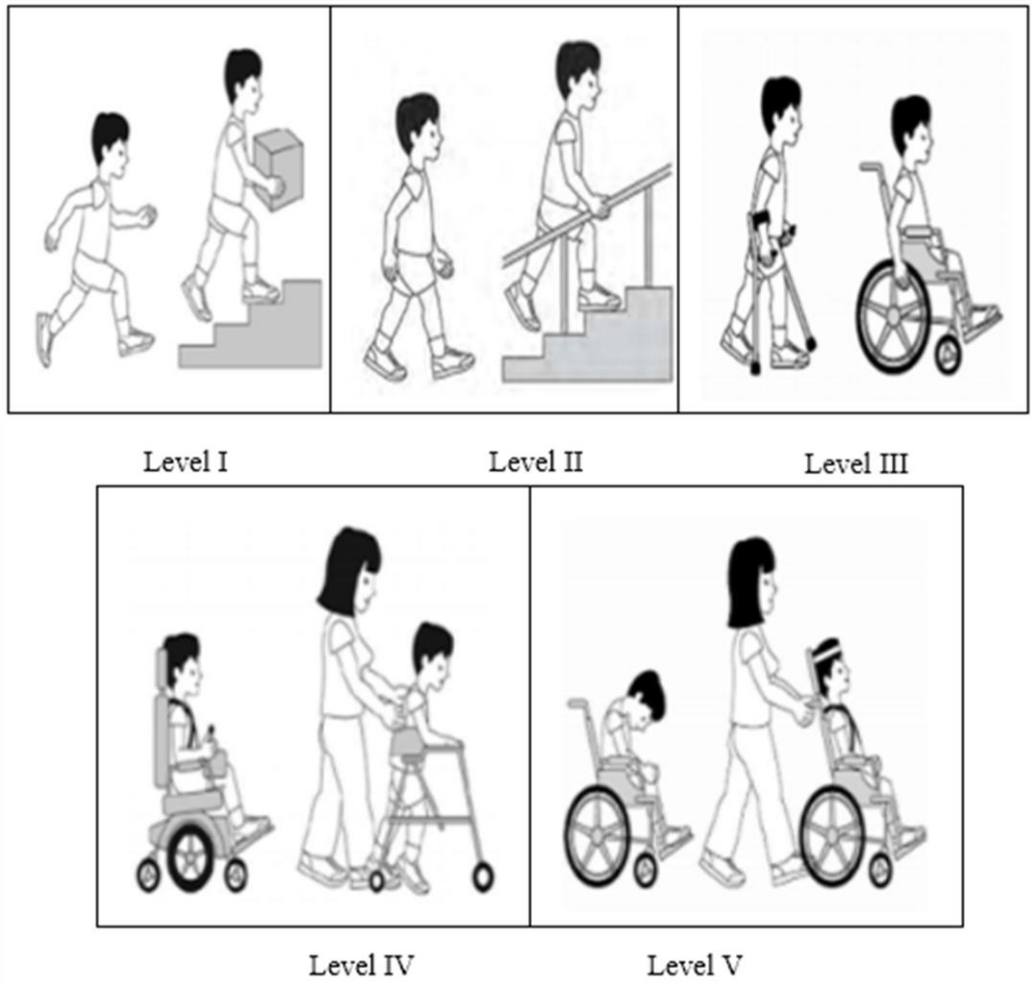


Figure 1.4: The levels of GMFCS classification system for CP children (Palisano et. al., 1997).

1.7 Cerebral Palsy Comorbidities

Besides clinical spectrum of movement difficulties as discussed earlier, CP individuals are also accompanied with other several comorbidities such as epilepsy, visual and hearing impairment, speech and communication problem, intellectual disability, autism, cardiac anomalies, urogenital abnormalities, skin problems with pressure sore or friction area, nose of air-way space problem and musculoskeletal anomalies including hip dislocation, contractures and scoliosis (Figure 1.5) (Odding et. al., 2006; Jones et. al., 2007; Andersen et. al., 2008; Pakula et. al., 2009; Moreno-de-luca et. al., 2012).

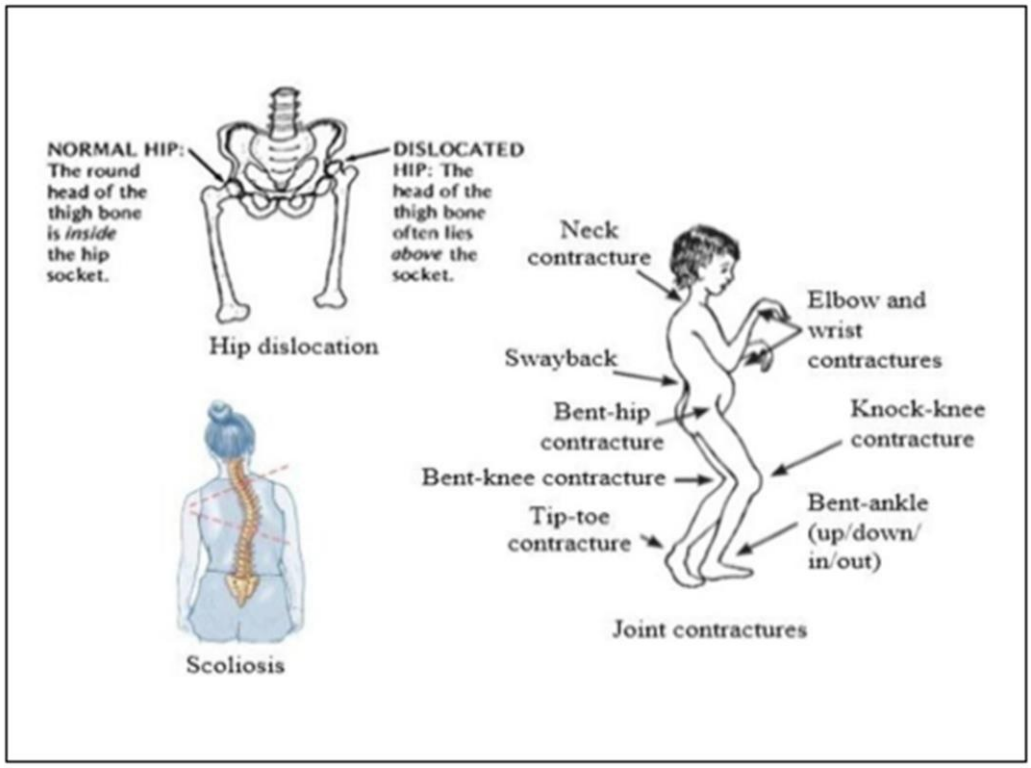


Figure 1.5: The musculoskeletal anomalies observed in CP individuals. (Extracted from *https://www.dinf.ne.jp/doc/english/global/david/dwe002/dwe00220.html*, *https://en.hesperian.org/hhg/Disabled_Village_Children:Contractures_in_Cerebral_P alsy* and © Mayo Foundation For Medical Education And Research).

Besides those, muscle impairments of pharynx and larynx are reported to lead to aspiration with feeding that causes the CP children having difficulty in swallowing, whereas oral muscle impairment causes drooling (Jones et. al., 2007; Amoghimath et. al., 2016). Apart from that, there are several CP comorbidities that are commonly reported. These include intellectual disability in 30-65% (Pakula et. al., 2009), epilepsy in 20-40% (Odding et. al., 2006), visual impairment in 40%, hearing impairment in 5-15%, seizures in 30-50% and speech deficit in 40% of CP cases (Moreno-de-luca et. al., 2012). All of these comorbidities are compiled and illustrated in a bar chart as in Figure 1.6.

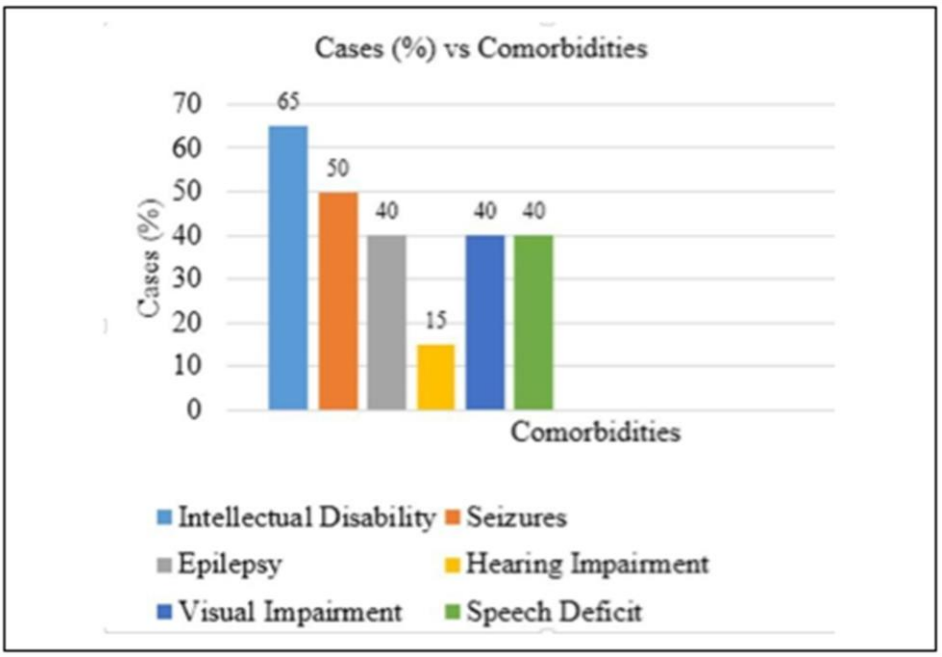


Figure 1.6: The prevalence and comorbidity types in CP individuals. These are compiled from several studies as (Odding et. al., 2006; Pakula et. al., 2006 and Moreno-de-luca et. al., 2012).

1.8 Management and Therapy For Cerebral Palsy Children

As a lifelong condition, CP is not curable but it is treatable in order to improve the CP children's quality of life. The treatments for CP children are based on the aspects of movement problems, associated medical problems and functional and cognitive abilities.

1.8.1 Movement Problem Treatment

In order to encourage motor development, treatments such as physiotherapy (body movement) and occupational therapy (ability to perform daily activities) should be done. There are several devices, alternatives and medications available to counter movement problems. Orthosis or also known as braces is an orthotic device that is used for lower limbs in CP children. It is used for standing, walking and preventing muscle contracture and deformity (Knutson et. al., 1991; Morris C., 2002). There are many types of orthosis such as foot orthosis to oppose foot pronation (collapsed foot arch), supramalleolar orthosis for malleoli support and plantar flexion and dorsiflexion, ankle-foot orthosis to provide support from fibula in children with excessive flexible knee flexion posture, and knee-ankle-foot orthosis provides support from knee to calcaneus to make the knee joint be in the gait line progression. There are also multiple of hip-gait orthosis support such as reciprocating gait orthosis to hold pelvis into butterfly-shape pelvic band which maximizes reciprocal action by both hip sides thus helps in sitting ability, hip guidance orthosis to help in walking faster however shows higher energy expenditure and parapodium orthosis for a child with swivel walking (Knutson et. al., 1991). There is also plaster cast alternative to calf the muscle and improve foot position during walking (Zachazewski et. al., 1982; Domalgaska et. al.,

2006). For upper limb extremities, splint devices are used to facilitate better grasp and maximize the arm and hand functions in CP children. Splinting prevents contracture and deformity by supporting the wrist and thumb in the functional position (Imms et. al., 2011). Next treatment for movement problems is Botulinum Toxin A (BoNT-A) injection for either lower limb or upper limb with the spastic or tight muscles problem. This treatment reduces muscle spasticity and lasts for between 3 and 6 months. In some CP children, there is also gait improvement but in a small magnitude. However, in ambulatory CP children, BoNT-A injection causes muscle atrophy (Multani et. al., 2019). Few oral medications are also prescribed for spasticity that restrict the movement. These include diazepam, dantrolene and baclofen. Diazepam is the most cost-effective medication and applicable to initiate afterward physical therapy in a newly CP diagnosed infant and young CP children (Mathew et. al., 2005).

Dantrolene decreases the stiffness by acting directly on contractile mechanism of skeletal muscle. Long term use of dantrolene causes liver problem (0.07% to 1.0%). It causes less sedation compared to diazepam (Pinder et. al., 1977). Baclofen is applied to CP children with severe spasticity to reduce the spasticity in CP children as well as spinal spasticity in adult. Diazepam, dantrolene and baclofen are similar in efficacy but have different spectrum of adverse effects (Lataste et. al., 1994). The orthopaedic surgery is another undertaken treatment to overcome the movement problems in CP children. There are various surgical techniques for different form of deformities and musculoskeletal problems suffered by the CP children (Tonkin et. al., 1995).

1.8.2 Associated Medical Conditions

Epilepsy, constipation, gastro-oesophageal reflux and drooling are the common medical problems associated with CP children. Epilepsy is a common medical condition in CP children with 20-40% occurrence (Moreno-de-Luca et. al., 2012). There are multiple established anticonsulvants used for this medical problem, such carbamazepine, sodium valproate, lamotrigine, phenytoin, phenobarbitone, clonazepam, topiramate and levetiracetam. These medications have various side effects such as drowsiness, nausea and sedation. Constipation is also a common problem in CP children (74%) and other disorders with neurological impairment. It may be influenced by disruption of the neural modulation of colonic motility (Veugelers et. al., 2010). Besides constipation there are few examples of intestinal motility problems which include oral-motor dysfunction, gastro-oesophageal reflux, delayed gastric emptying and slow intestinal transit (Staiano et. al., 1994; Ravelli et. al., 1998). The medical treatments depend on different categories of how severe the constipation form is.

Next is gastro-oesophageal reflux, another intestinal motility problem in which the lower oesophageal sphincter (LES) is affected. This results in an involuntary passage of gastric contents into the oesophagus. It has been found in a higher prevalence which is up to 75% in CP patients (Su et. al., 2003). To lower the reflux there are non-pharmacological alternatives such as raise the head of the bed, reduce weight, limit secondary exposure to smoke, take high-pectin diet, avoid taking diet containing caffeine, spicy foods, fatty foods, and chocolate. These also prevent further complications (Espocito et. al., 2015; Fernando et. al., 2019). Pharmacological

treatments to significantly reduce acid reflux are proton pump inhibitors, H2 antagonist baclofen medication, and prokinetic agents like mosapride (Kawai et. al., 2004; Whitworth et. al., 2012). A medical condition of uncontrolled saliva that is resulted from impaired swallowing and poor lip closure. The incidence is 48.7% to 58% in which 17.7% are having severe drooling (Tahmassebi et. al., 2003; Hedge et. al., 2009). Speech pathologists can help improve the ability to control mouth closure to overcome drooling problem (Hedge et. al., 2009). However, if the speech therapy is not effective, the benzhexol hydrochloride is suggested (Reddihough et. al., 2008a). The severity of drooling is reported to be reduced with age due to maturation of the oral muscle. However, persistent drooling problems can be fixed by surgical treatment to remove the sublingual glands (Burton, 2008).

1.8.3 Functional and Cognitive Abilities

Multiple therapy programs are to improvise CP children's progresses. As the early intervention programs for CP disorder, these therapy programs also involves the caregivers to make them knowledgeably participate in maximizing CP children abilities. Neuro-developmental therapy or Bobath Therapy is to maximize the CP children's functional abilities. It focuses on making any desirable movements more possible and preventing any undesired movements. It improves functional motor level and functional independency by improving postural control and balance (Tekin et. al., 2018). Second is Conductive Education program to overcome the learning disabilities. This program educates on how to develop life skills including daily living, physical, emotional, social and cognitive and communication skills. This program also educates parents on how to assist their CP children with mentioned skills (Reddihough et. al.,

2008b). Next is Constraint Induced Movement (CIMT) therapy which is applicable for hemiplegic CP children. This therapy encourages them to use their hemiplegic arms and hands in doing activities by constraining their non-hemiplegic arms and hands (Hoare et. al., 2007). It successfully sustain the improvement in motoric function in young hemiplegic CP children (Taub et. al., 2004). Another functional abilitytraining program is called goal directed training. The families and the CP children themselves are trained to identify which specific task that CP children want to perform at a particular time. The principle is repeating the attempt for the same task. Families and therapists can assess the children's ability in performing the task.

1.9 Genomic Approaches to Identify Gene Alterations in Cerebral Palsy

Multiple studies had shown that CP is a complex disorder that is genetically heterogeneous. It is not a single gene disorder and DNA variants are discovered to be associated with CP. By definition, "DNA variant" means any gene alterations that occur in comparison with the "reference genome" sequences. This "variant" term is further categorized into two different types which are single nucleotide polymorphisms (SNP) and mutation (Karki et. al., 2015). "SNP" is referring to any gene alterations in DNA base sequences that occur at any single position in the genome, and it is abundantly (>1%) found in a population (Brookes, 1999). Meanwhile, "mutation" which is found in less than 1% in a population, is referring to any kind of gene alterations in the DNA sequence and it also refers to gene copy number variation (CNV) (Condit et. al., 2002; Brooker, 2012). In order to identify gene alterations in CP, several approaches had been applied such as Homozygosity Mapping

(HM) (section 1.9.1), analyses of candidate genes through association study (section 1.9.2) and next generation sequencing (NGS) (section 1.9.3).

1.9.1 Homozygosity Mapping (HM) and Linkage Analysis Approaches

It is also known as autozygosity mapping, a type of genome-wide analysis that involves genotyping and linkage analysis. It is mostly applicable for identifying recessive variants by mapping for DNA markers identical by descent in consanguineous families (Seelow et. al., 2009), for Mendelian diseases with a homozygous autosomal recessive pattern of inheritance and for an extended family with more than one affected family members (Vahidnezhad et. al., 2018). The principle of HM is to search for genetic regions of homozygosity (ROH) in patient's DNA that is varying from a few to several megabase pairs (Mb). Thus minimizes the overall cost and time (Alkuraya et. al., 2010). The HM technique involves procedures of polymerase- chain reaction (PCR) followed by genotyping to identify any ROHs presence (McHale et. al., 2000).

ROH containing mutated genes of rare recessive traits will be detected from that SNP-based genotyping. However, ROH also can be detected from NGS data (Vahidnezhad et. al., 2018). This HM analyses can be useful for both research and clinical purposes with the case of multiple affected relatives. Associated with HM is the linkage analysis which localizes a region in genome where a locus or loci that regulate the expression of a trait may be harboured (Ferreira, M. A R., 2004). Linkage analysis determines whether or not the regions are linked to the disease (Seelow et. al., 2009).

Several CP studies which used the HM method were successfully discovering several gene alterations that are implicated in CP such as single locus mutation on chromosome 9p12-q12 (refer subsection 1.9.1(a)), GAD1 gene mutation Adaptor Protein 4 Complex gene mutations (refer subsection 1.9.1(b)), Adducin 3 Gamma (*ADD3*) gene mutation (refer subsection 1.9.1 (d)) and *ANKRD15* or *KANK1* deletion (refer subsection 1.9.1(e)).

1.9.1(a) Single Locus Mutation of Chromosome 9p12-q12

A study by McHale's team (2000) was done involving an Asian family pedigree with two consanguinities sub-families producing two different sibships respectively. The first sibship consists of two ataxic CP of son and daughter out of four children meanwhile the second sibship consists of two ataxic CP daughters out of three daughters. All the diagnosed CP children were idiopathic CP cases and were born at term on spontaneous vagina delivery (SVD). They were also delayed in motor milestones, however, were in normal range of intelligence. The genome search using the linkage mapping kit with 343 autosomal dinucleotide markers was done for all affected children and their unaffected parents respectively. Another genotyping was applied to the other three unaffected children to check for any homozygosity regions. This was followed by statistical analysis of Linkage analysis which later discovered a mutation at single locus of chromosome 9p12-q12. A homozygous region of 23 cM near to centromere on Chromosome 9 was present in all affected individuals. This study confirmed the finding as single gene disorder and proposed that the pattern of inheritance for this familial CP case as autosomal recessive (AR).

1.9.1(b) Missense Mutation of GAD1 Gene

Lynex & colleagues (2004) studied two different consanguineous families with 2 and four individuals of congenital spastic CP of unknown specific cause, respectively. All of the six CP individuals had global developmental delay, moderate to severe intellectual disability, poor or absent speech. The second family consists of two pairs of twin children. Only one son of the eldest pair of dizygotic twins that were born by Caesarean section is CP affected. The linkage mapping was done for 290 polymorphic DNA markers of which 0.5cM region was found on chromosome 2q24-25. This is followed by direct sequencing for all affected and unaffected individuals which later confirmed the presence of homozygous missense mutation in glutamate decarboxylase 1 (*GAD1*) gene segregating with CP individuals. The GAD1 gene encodes for the brain-expressed isoform of glutamate decarboxylase that is responsible for the production of gamma aminobutyric acid (GABA) from its excitatory counterpart glutamate. GABA is the major inhibitory neurotransmitter in CNS (Bu et. al., 1992) and is crucial for normal brain development (Martin et. al., 2000) and synaptic plasticity (Hyde et. al., 2011).

1.9.1(c) Mutations in Adaptor Protein 4 Complex Genes (AP-4)

The AP-4 complex is expressed in CNS during embryologic and postnatal development (Yap et. al., 2003; Verkek et. al., 2009). The AP-4 complex genes with subunits of M1 (μ1), E1 (ε1), B1 (ß1) and S1 (σ1) are basically involved in intracellular trafficking of glutamate receptors (Yap et. al., 2003). The AP-4 proteins play a rolein membrane protein vesicular trafficking to selectively sort the proteins from trans-Golgi network to the postsynaptic somatodendritic domain in order to establish the

neuronal polarity (Matsuda et. al., 2009). In 2009, Verkek and team reported a study on a consanguineous Moroccan family with 5 CP siblings of infantile hypotonia which later evolved to spastic quadriplegia within first year of life. The HM was done on the candidate region of 14Mb of chromosome 7q22. They found the homozygous mutation in intron of AP4M1 gene (c.1137+1G→T) in all CP affected children. This *AP4M1* encodes for μ subunit. The mutation leads to truncated carboxyl (COOH) terminus, a cargo binding site. Verkek's team hypothesized that this mutation contributes to defect during brain development which leads to congenital spastic quadriplegia and its pattern of inheritance is AR. In 2011, another mutation was discovered in another AP-4 complex gene, *AP4E1,* in Palestinian-Jordanian inbred kindred with two CP spastic tetraplegic children (Moreno-De-Luca et. al., 2011). Through copy number array analysis, they identified a homozygous deletion on chromosome 15q12.2 specifically on exons 1-11 of AP4E1 gene that is presents in both CP children.

The *AP4E1* encodes for ε1 subunit. Jamra et. al., 2011 studied eight individuals with spastic paraplegia from three consanguineous Syrian families. Using autozygosity mapping, Sanger sequencing and next-generation exome sequencing they discovered mutations in AP-4 complex genes in those affected individuals. The mutations are a nonsense mutation (c.124C>T) in AP4S1 gene (encodes for σ1 subunit), a frameshift mutation (c.487_488insTAT) in AP4B1 gene (encodes for ß1 subunit) and a splice mutation (c.542þ1_542þ4delGTAA) in AP4E1 gene (encodes for ε1 subunit) (Jamra et. al., 2011). Disruption of any one of the aforementioned four subunits of AP-4 complex genes would result in the AP-4 complex dysfunction resulting in AR of CP (Moreno-De-Luca et. al., 2012).

1.9.1(d) Mutation in Adducin 3 Gamma (ADD3) Gene

In 2013, Kruer and team studied a consanguineous family that consisted of three spastic quadriplegia children and one spastic diplegia child from unaffected parents. Using the HM, exome sequencing and Sanger sequencing, a homozygous mutation of c.1100G>A (p.G367D) in *ADD3* was discovered. They also did an in vitro study to observe the loss of function of *ADD3* in drosophila and found out that the mutation in *ADD3* disrupts the normal actin-capping function of adducin, leading to abnormal proliferation and migration. This has confirmed a critical role for adducin in locomotion including to regulate the actin cytoskeleton. They suggested that disrupted adducin function may lead to neuromotor impairment and implicating abnormalities of the dynamic cytoskeleton as a pathogenic mechanism contributing to CP.

1.9.1(e) Deletion on KANK1 Gene

A study by Lerer's team (2005) involved an extended family of four large-generation pedigree with seven CP children who are in third generation and two who are in fourth generation. The CP individuals' clinical features are intellectual disability (moderate to severe), nystagmus, brain atrophy, ventriculomegaly and congenital hypotonia that evolved to spastic quadriplegia. Linkage analysis identified a 225 kb deletion on chromosome 9p24.3 involving KN motif and ankyrin repeat domain 1 (KANK1) gene (previously known as ankyrin domain 15 (*ANKRD 15*) gene) in all affected individuals in an extended family, but was absent in 210 control individuals. KANK1 is ubiquitously expressed during foetal development and also adult life. This gene involves brain development and also important for cytoskeleton structure and plays roles in protein-protein interaction and signalling complexes (Lerer et. al., 2005).

1.9.2 Candidate-Genes Association Study Approach

This approach focuses on the candidate genes that are known for their functions which are relevant to the particular studied disease or disorder (Zhu et. al., 2007). Several association studies with different techniques discovered multiple SNPs such as cytokine SNPs (refer subsection 1.9.2(a)), *induciblenitric- oxide synthase and lymphotoxin α* SNPs (refer subsection 1.9.2(b)), *Interleukin-6* SNPs (refer subsection 1.9.2(c)) and *apolipoprotein E* (refer subsection 1.9.2(d)) which contribute to CP susceptibility during foetal development.

1.9.2(a) Cytokine Single Nucleotide Polymorphisms (SNPs)

A case control study by Gibson et. al., (2006) involving 443 CP white infants and 883 control white infants using DNA isolation and amplification had shown the association of two inherited cytokine polymorphisms with CP. The two polymorphic groups of cytokines are *mannose- binding lectin* of exon 1 at codon 54 (MBL-54) and *tumor necrosis factor-2 (TNF-2)* polymorphism in the promoter region. These two polymorphic groups had been genotyped to determine the polymorphic base at the interest site based on the genotyping standard reference samples. TNF-alpha (TNF-α) acts as proinflammatory cytokine to promote inflammatory and immunology response. However, its *TNF-2* polymorphism at codon 308 increases the circulating level of TNF-alpha (TNF-α) protein which becomes toxic to neurons, specifically contributing to pathogenesis of white matter brain damage. Meanwhile, the MBL protein acts as anti-infectious agent by identifying and removing the pathogens from the body (Kilpatrick et. al., 1999). The *MBL-54* polymorphism causes the low level of MBL protein resulting in deleterious effect on the foetus. This study proved that these two

cytokine polymorphisms contribute to CP pathogenesis by affecting the foetal brain and these polymorphisms are associated with an increased risk of CP occurrence.

1.9.2(b) SNPs of *iNOS* and *LTA*

An association study by Gibson et. al., 2008 was carried out to re-evaluate 28 candidate SNPs that had been shown to be associated with CP as reported previously. In this study, Gibson and team applied the technique of 5' nuclease assay with 2-fluorescent dye oligonucleotide probes for amplification and genotyping purposes. However, from a total of 28 SNPs, only two SNPs were shown to be associated with CP in the total white Australians population of 413 CP children and 856 control children who were born in specified time range of year 1986 to 1999. These two SNPs are inducible nitric-oxide synthase (*iNOS*) gene SNP on chromosome 17q11.2-12 and a SNP of lymphotoxin α (*LTA*) gene or also known as *tumor necrosis factor-beta* (*TNF-ß*) on chromosome 6p21.3. Both genes are involved in inflammatory response. Overexpression of *iNOS* causes injury in white matter region in both infant and adults' brains (Boullerne et. al., 2006). Gibson et. al., 2006 proposed that heterozygous T allele of *iNOS* and homozygous variant of *LTA* influence the risk of CP.

1.9.2(c) SNPs in *Interleukin-6* (*IL6*)

Khankhanian and team (2013) studied 250 CP infants randomly selected from the Kaiser Permanente Medical Care Program (KPMCP) against 305 controls. The subjects were African American, Asian, Hispanic, white, and others including mixed and unknown races. The team executed the Sanger sequencing for the targeted IL6

gene since it had been proposed to be associated with increased risk of CP (Djukic et. al., 2009; Wu et. al., 2009; Wu et. al., 2011). This study discovered seven SNPs in IL6 gene which are rs1800795 in promoter region, rs2069832 (intron), rs2069833 (intron), rs1474348 (intron), rs1474347 (intron), rs1554606 (intron) and rs2069845 (CpG site). These SNPs were shown to be associated with CP pathogenesis by increasing the IL6 protein level. High IL6 protein level in amniotic fluid, cord plasma and brain lesions lead to CP risk of periventricular leukomalacia (PVL).

1.9.2(d) SNP in Apolipoprotein E (ApoE) Gene

ApoE is a lipid transport protein, abundantly found in brain cells. Based on the study of 350 CP and 242 control individuals in a Chinese population, ApoE gene with rs769446 SNP of C allele was detected using MassARRAY platform-based genotyping. Multiple association analysis showed that *ApoE* SNP is associated with CP (Xu et. al., 2014).

1.9.3 Next Generation Sequencing (NGS)

In genome search analysis specifically to identify gene structures and functional characteristics that are implicated in diseases or disorders, several technologies have been developed such as Sanger sequencing, DNA microarray and the newest technology called Next Generation sequencing (NGS). Sanger sequencing is a convential method that sequences one gene or a small group of targeted genes at a time (Brooker, 2012). This is completely opposite to NGS that is capable of simultaneously sequencing thousands of nucleotides in a whole genome side by side at a large-scale

(McCombie et. al., 2019). In accordance to its high yield of sequencing reads, NGS is also wordwidely known as a massively parallel sequencing technology or high-throughput sequencing (Brooker, 2012). DNA microarray or gene chip technology is able to analyze the expression of thousands of genes simultaneously (Wiltgen et. al., 2007). Microarray principle which is using a small glass slide dotted with many different DNA sequences, is able to detect sequences in known genes only (Brooker, 2012).

On the other hand, NGS's sequencing ability is beyond that of microarray since it captures and sequences the genome at a large scale as stated earlier. Thus, this gives the possibility of obtaining any new variants in wider DNA regions besides known genes region such as unknown genes, untranslated regions (UTRs), as well as downstream, upstream and splicing regions. Using NGS, many previous studies had discovered variants such as single nucleotide variants (SNVs), small DNA insertions or deletions (indels), copy number variations (CNVs), stopgains and stoplosses that are present in different DNA regions (Rabbani et. a., 2013; Belkadi et. al., 2015; Yano et. al., 2016). Thus, NGS supersedes both conventional Sanger sequencing and microarray in genome search field. Following that, many different platform technologies have been developed and offered to be applied to run the NGS work. Some platforms that have been widely used are Agilent, Nimblegen, Illumina Nextera and Illumina Truseq whereby each of them has different specifications, performances and principles (Chilakumari et. al., 2014). These platforms are applicable for both techniques in NGS which are whole genome sequencing (WGS) (refer subsection 1.9.3(a)) and whole exome sequencing (WES) (refer 1.9.3(b)).

1.9.3(a) Whole Genome Sequencing (WGS)

WGS sequences the whole genome, is applicable for detection of variants in non-coding DNA regions too and is considered as more advance compared to WES (Belkadi et. al., 2015). However, it is more costly compared to WES. WGS is a common technology that is used in Genome Wide Association Studies (GWAS) that determine the association of specific genetic variations with particular diseases (Bush et. al., 2012). GWAS principle is scanning the genomes of many different individuals to detect any genetic markers that can be used to predict the presence of a disease in a particular studied population (Tam et. al., 2019). Previously, many studies have been reported showing that identification of related genetic markers helps to understand the genes contribution to a particular disease which later led to the development of preventions and treatment strategies.

1.9.3(b) Whole Exome Sequencing (WES)

WES technology provides the sequencing coverage of > 95% in which 85% are mendelian-causal mutations and the rest are SNPs that cause the diseases (Rabbani et. al., 2013). It is expected that this WES analysis would sufficiently cover the most possibilities where and how mutations may occur. On 3rd November 2015, the Human Gene Mutation Database (*http://www.hgmd.cf.ac.uk/ac/index.php*) documented 174,999 pathogenic mutations that cause human diseases. Of these, 97,626 (55.8%) are missense/nonsense, 15,990 (9.1%) are splicing abnormalities and 39,537 (22.6%) are small (< 20 bases) insertion/deletions. These types of mutations (87.5%) are expected to be covered by WES. There are 3297 mutations (1.8%) affecting regulatory arrangements and located in non-coding regions that may be missed by this method.

Nevertheless, most of this type of mutations could be located in 5' UTR, 3'UTR or near to exon-intron boundaries which will be targeted by WES at high sensitivity through sequence captured by hybridization techniques using the exome-capture kit such as Nextera Rapid Capture Expanded Exome Kit (Illumina Technology, USA) that has been applied in this current study. Other than that, WES also has been acknowledged as a cost-effective method for facilitating the identification of genetic variants in neurological disorders that overlap with CP such as brain malformation, cerebellar ataxia, intellectual disability, autism and epilepsy (Bilgüvar, et. al., 2010; Wang et. al., 2011; De Ligt, et. al., 2012; Sanders et. al., 2012).

1.9.3(c) WES detected De Novo Mutations in 3 Known Ataxic Genes

The NGS of exome sequencing study was conducted by Parolin et. al., 2015 for 57 ataxic known genes involving 10 cases whereby four cases had showed the finding of four de novo mutations from three genes of voltage-gate potassium channel activity *Kv3.3* (*KCNC3*) (n=1), *inositol triphosphate receptor* (*ITPR1*) (n=2) and *ß-III spectrin* (*SPTBN2*) (n=1). The *KCNC3* with p.T428I mutation was found in an ataxic CP child (cases 1) of a non-consanguineous family who was born from uneventful pregnancies and birth, and normal brain imaging. A functional study of p.T428I mutation showed a severe negative loss-of-function, whereby 50% potassium channel activity was reduced in the p.T428I mutant compared to wild type. The *KCNC3* mutation was previously shown to be associated with spinocerebellar ataxia type 13 (Figueroa et. al., 2011). Two de novo mutations were shown in *ITPR1* in two different cases of two different consanguineous families. The mutations were P.N602D and P.S1487D of *ITPR1* in case 2 and case 3 respectively in which both brain imaging were normal

(Parolin et. al., 2015). The *ITPR1* encodes for intracellular receptor for inositol 1,4,5-triphosphate, it is also associated with spinocerebellar ataxia type 15 (Van de Leemput et. al., 2007) and autosomal dominant cerebellar ataxia (Huang et. al., 2012). The third gene with one de novo mutation (P.R480W) found in this Parolin's study was *SPTBN2*. The mutation was discovered in the fourth child of a non-consanguineous family from a normal pregnancy but was born through emergency caesarean due to foetal distress. The encoded protein-spectrin beta non-erythrocytic 2 or also known as beta-III spectrin, stabilizes the glutamate transporter at the plasma membrane surface and directly regulates the glutamate signalling pathway. *STPBN2* was also shown to be associated with spinocerebellar ataxic type 14 (Lise et. al., 2012).

1.10 Bioinformatics Analysis

Bioinformatics is a computational analytical program completed with computational languages, tool and software to analyse WES raw data. Bioinformatics analysis involves five analytical steps which are raw quality assessment, pre-processing, alignment, recalibration and variant analyses (detection, filtering and annotation) (Bao et. al., 2014). Raw quality assessment followed by pre-processing are the two steps to inspect for problematic sequencing raw data and prepare the trimmed sequence data for the next alignment procedure. Third analytical step is sequence alignment where the filtered good quality sequences of read are aligned to a reference human genome (hg19) using the Burrows-Wheeler Aligner (BWA-MEM) aligner, producing the BAM format files output. Throughout the alignment and recalibration analytical steps, few bioinformatics tools are involved such as SAMtools and picard tools. The SAMtools is to produce sorted bam file, which is a binary format of the alignment file

whereas the picard tool then is to remove duplicates from the PCR amplification. Variant analysis step is the last analytical procedure which the goal is to filter the sequenced data against multiple genomic databases and annotating them according to few prediction tool programs that sort the variant according to protein changing. For example, Sorting Intolerant From Tolerant (SIFT) tool that gives the variants score in range of intolerant to tolerant (Ng et. al., 2001; Kaminker et. al., 2007), Polymorphism Phenotyping (PolyPhen) tool that predicts human disease-mutation whether or not it is deleterious (Ramensky et. al., 2002), Genomic Evolutionary Rate Profiling Rejected Substitution (GERP++_RS) that uses maximum likelihood evolutionary rate estimation for locus specific scoring (Davydov et. al., 2010). Another prediction tool is Combined Annotation-Dependent Depletion (CADD) (Kircher et. al., 2014) which is to measure deleteriousness of the variants that correlates with the pathogenicity and molecular function. The principle and procedures of bioinformatics analysis are explained in detail in Chapter 2.

1.11 Rationales of The Study

CP cases with unknown causes are found to be associated with genetic alterations which either directly cause CP or contribute to CP susceptibility. This is supported by several studies that showed multiple genetic factors contribute to the CP causals (Verkek et. al., 2009; Moreno-de-luca et. al., 2012; McMichael et. al., 2015). Previously, genetic association studies (case-control approach) have failed to produce strong, replicable results when applied to complex, multi-factorial, and highly heterogeneous groups of disorders, such as CP (Moreno-De-Luca et al., 2012). Thus, this current study adopted the WES approach to explore possible variants that are

implicated in studied CP cases. Several studies had shown the capability of WES in discovering variants in neurological disorders that overlap with CP features (Bilguvar et. al., 2010; Wang et. al., 2011; De Ligt et. al., 2012; Sanders et. al., 2012). The contribution of genomic abnormalities for CP cases has not been researched widely in Malaysia, hence the identification of specific causes of the disorder would be very helpful. This may provide a better understanding of the disorder, accurate assessment of recurrence risk, and early intervention. Our preliminary study provides insights into the genetic cause of CP in selected Malaysian patients, and also provides a basis for fundamental studies into molecular pathologies of CP through the identified mutations. It may eventually encourage further research into the development of molecular targeted drug and therapeutic interventions for CP.

1.12 General Objective

This is an exploratory study whereby no hypothesis of specific loci was proposed since the main objective was to explore the genomic of CP cases. This preliminary study focused on idiopathic CP patients in Kelantan with defined genetic underlying factors using WES technique.

1.12.1 Specific Objectives

(a) To screen the CP families which fulfil the inclusion and exclusion criteria.

(b) To determine the pathogenicity of discovered variants in subjects.

(c) To investigate the inheritance of discovered variants.

A brief description of this study workflow is illustrated in Figure 1.7 below.

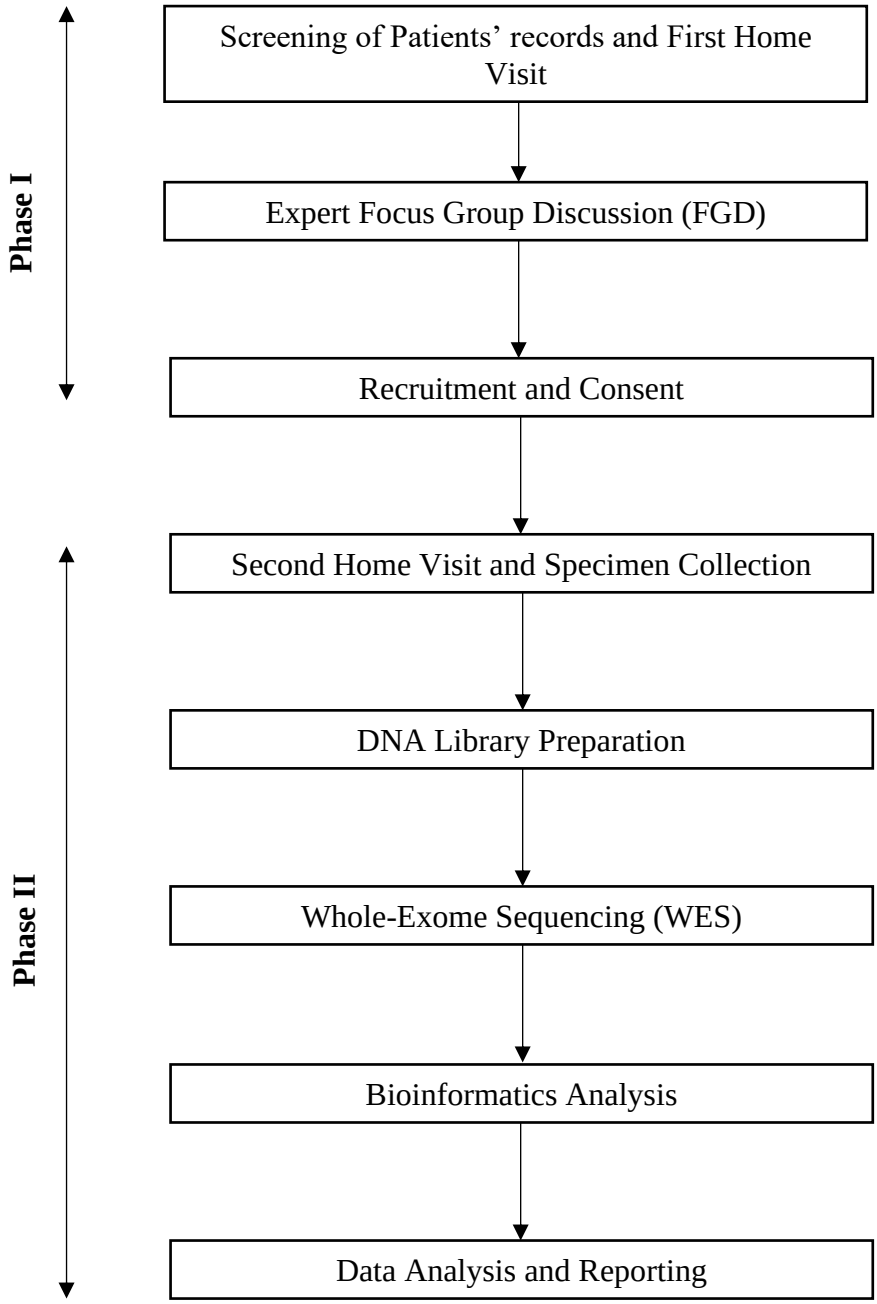


Figure 1.7: Flowchart showing the study dissected into two phases.

CHAPTER TWO

MATERIALS AND METHOD

2.1 Materials and Chemicals

2.1.1 Chemicals and Reagents

All chemicals and reagents used in this study are listed in Table 2.1.

Table 2.1 List of chemicals and reagents.

Name and Brand	Supplier
Absolute Ethanol (90%)	Elite Advance Material, Malaysia
Agarose Powder	Sigma Aldrich, Merck, Germany
DNA Marker (100bp)	Promega Corporation, USA
Elute Target Buffer 2 (ET2)	Illumina Technology, USA
Elution Buffer (AE)	Gene All Technology, Seoul Korea
Enrichment Amplicon Mix (NEM) PCR	Illumina Technology, USA
Enrichment Elution Buffer 1 (EE1)	Illumina Technology, USA
Enrichment Hybridization Buffer (EHB)	Illumina Technology, USA
Enrichment Hybridization buffer (EHB)	Illumina Technology, USA
Enrichment Wash solution (EWS)	Illumina Technology, USA
Ethidium Bromide (EtBr)	Sigma Aldrich, Merck, Germany
Expanded Exome Oligos (EEX)	Illumina Technology, USA
Index 1 (i7) Adapters	Illumina Technology, USA
Index 2 (i5) Adapters	Illumina Technology, USA
Library Amplicon Mix (NLM)	Illumina Technology, USA
Loading Dye	Thermo Fisher, USA
Lysis Buffer (BL)	Gene All Technology, Seoul Korea
Primer Cocktail (PPC)	Illumina Technology, USA
Proteinase K solution	Gene All Technology, Seoul Korea

Table 2.1: Continued.

Name and Brand	Supplier
Resuspension Buffer (RSB)	Illumina Technology, USA
Sample Purification Beads (SPB)	Illumina Technology, USA
Stop Tagment Buffer (ST)	Illumina Technology, USA
Streptavidin Magnetic Beads (SMB)	Illumina Technology, USA
Tagment DNA Buffer (TD)	Illumina Technology, USA
Tris-Borate-EDTA (TBE) 5X	Sigma Aldrich, Merck, Germany
Washing Buffer Solution B (BW)	Gene All Technology, Seoul Korea
Washing Buffer Solution T (TW)	Gene All Technology, Seoul Korea

2.1.2 Consumables

All consumables used in the study are listed in the table below.

Table 2.2: List of consumables

Name and Brand	Supplier
BD Alcohol (70%) Swab	Beckton & Dickinson, USA
EDTA Tube, BD Vacutainer	Franklin Lakes NJ, USA
Electrophoresis Gel Casting Tray with Well Slots Comb	Labnet, Malaysia
Microtubes (1.5ml)	Eppendorf, Germany
Min Elute Column	Gene All Biotechnology, Seoul Korea
Needle	Terumo, USA
PCR tubes	Eppendorf, Germany
Pipette tips	Axygen , USA

2.1.3 Kits

The kits that had been used in the study are listed in Table 2.3 below.

Table 2.3: List of the commercial kits

Name and Brand	Supplier
Gene All DNA Extraction kit	Gene All Biotechnology, Seoul, Korea
Nextera Rapid Capture Expanded Exome kit	Illumina Technology, USA
NextSeq 500 High Output v2 Kit (300 cycles)	Illumina Technology, USA
NextSeq Phix Control Kit	Illumina Technology, USA

2.1.4 Laboratory Apparatus

All apparatuses that had been used for laboratory works are listed in Table 2.4.

Table 2.4: List of laboratory apparatuses.

Laboratory Apparatus	Supplier
Waterbath (Memmert)	Camlab, UK
Bioanalyzer	Agilent, USA
Centrifuge	Eppendorf 5810R
Electrophesis Gel Imaging System	Bio-Rad, USA
Electrophoresis Tank with Cable	Labnet International, Woodbridge
UV Transilluminator A	Labnet International, Woodbridge
Gel Imager	Uvitec, UK
Magnetic Stand	Illumina, USA
Microheating System Microheater	Illumina, USA
Microplate Reader	Thermo Scientific, USA
Microwave	ELBA, Malaysia
Plate Shaker	Thomas Scientific, USA
Qubit DNA Reader	Thermo Fisher Scientific, USA
Mixture Short Spinner	Gaia Science, Malaysia
Thermal Cycler	Bio-Rad, USA
Vacuum Concentrator	Thermo Fisher Scientific, USA
Vortex Mixer	Thermo Fisher Scientific, USA

2.2 Overall Study Design

This was a cross-sectional study design with no hypothesis needed. It was an exploratory study whereby the genomics of idiopathic CP was determined and explained by scanning the whole exome of subjects recruited as shown in the flowchart (Figure 1.7).

2.3 Subjects and Subjects' Criteria

Families in Kelantan were selected primarily through patients' records available at Hospital Universiti Sains Malaysia (Hospital USM), Hospital Raja Perempuan Zainab II (HRPZ II) and Yayasan Orang Kurang Upaya Kelantan (YOKUK). All affected members and at least 2 unaffected members (preferably the parents) per selected family were recruited for this study based on the following inclusion and exclusion criteria. Ethical approval was sought from the Human Research Ethics Committee, USM (approval code: USM/JEPeM/14090310) and the Medical Research Ethics Committee, Ministry of Health Malaysia (MREC) (approval code: NMRR-15- 2273-28140).

2.3.1 Inclusion Criteria

1. Fulfil clinical diagnostic criteria for CP.

AND

2. More than one family member (from the same parental lines) are affected with similar clinical features/progression.

OR

Born from consanguineous parents

OR

Either one or both parents are more than 40 years old at conception

OR

Syndromic features with brain malformations, coagulopathies and prematurity.

2.3.2 Exclusion Criteria

1. Intra or post-partum events that may have led to hypoxia due to mechanical factors.

OR

2. Other syndromic features with defined chromosomal abnormalities.

OR

3. Evident infection during pregnancy.

OR

4. Parental history of identifiable genetic disorder.

OR

5. Family history of identifiable genetic disorder.

2.4 Subjects Selection and Recruitment

2.4.1 Screening of Patients Records

Screening was done retrospectively through YOKUK CP database which then generated a list of 330 CP patients all over Kelantan. This list of CP patients was then screened against the Hospital USM and HRPZ II medical records to produce a list of patients that fulfil the inclusion and exclusion criteria.

2.4.2 Family Data Collection (First Home Visit)

After the attending pediatric neurologist informed the shortlisted families, the first home visit was done to collect the family data including the three generational pedigree charts and patients' physical examination data using the data collection sheet (Appendix A). The three generation pedigree charts are to observe for any history of CP in the previous two generations which are parent and grandparent levels. During the physical examination, a pediatrician examined the CP children for the body area that were affected, the type of motor dysfunction (spastic or non- spastic), muscle tone, functional abilities, mobility, any deformities and any comorbidities. The pediatrician also gathered the family history to see for consanguinity and any previous history of identifiable genetic disorder.

2.4.3 Expert Focus Group

A focus-group discussion (FGD) was conducted, involving pediatric neurologists, molecular biologists and members of the USM-CPRC research team. Eight families were identified by the FGD then as they fulfilled inclusion criteria and exclusion criteria.

2.5 Sample Size and Subject List

Consent was taken from the parents prior to enrolment into the study. Since this study was focusing on CP individuals with suspected underlying genetic factors due to no causal event occurred, thus CP affected children and both unaffected parents were recruited from each family. At the beginning of the study, only six out eight shortlisted families were consented for inclusion into the study. Meanwhile the other two families were not willing to commit for the next study phases due to their personal reasons. Thus, this made up a study provisional total number of 25 subjects as recorded in Table 2.5.

Table 2.5: Provisional list of 25 identified subjects in the beginning of the phase I study.

	Fam 1	Fam 2	Fam 3	Fam 4	Fam 5	Fam 6
Father	D1	D2	D3	D4	D5	D6
Mother	M1	M2	M3	M4	M5	M6
Child 1	B1-1	G2-1	G3-1	B4-T1	B5-1	B6-1
Child 2	NA	B2-2	B3-2	B4-T2	G5-2	B6-2
Child 3	B1-3	NA	G3-3	NA	NA	NA
No. of Subjects	4	4	5	4	4	4
					TOTAL	25

Alphabetical code:

D: Father, M: Mother, B: Boy, G: Girl, T: Twin and NA: Not Available Alphanumeric code:

Fam 1: Family 1, Fam 2: Family 2, Fam 3: Family 3, Fam 4: Family 4, Fam 5: Family 5 and Fam 6: Family 6.

Whereas, -1, -2 and -3: Sibling order for recruited CP children per family Numerical code:

1 to 6 (next to B, D, G and M): Family Identification code (ID)

However, the final total number of sample that were analyzed were 20 samples only (listed in result chapter) after two times sample reduction for reasons as explained in subsection 2.9.1 and subsection 2.9.6. The sample size in this study is comparable to those of other studies reported. A previous study that recruited only four affected individuals in three independent kindred succeeded to identify the cause of Mendelian disorder using exome sequencing (Ng et. al., 2010b). Small sample size were also used to investigate the Kabuki Syndrome (10 individuals) (Ng et al., 2010a) and Hereditary Spastic Paraparesis (three individuals) (Erlich et al., 2011) using WES. Different from case- control association studies that require significant number of cases and controls to generate sufficiently powered analysis, these studies and the current study investigated each family independently and determined the genetic cause in each family. It is comparable with determining disease-causing mutation of beta-globin gene in a case of beta-thalassemia whereby decision on whether or not the mutation causes the disease will rely on the effect of the mutation on protein production, not on the difference in frequency between case and control.

2.6 Specimen Collection (Second Home Visit)

The CP children and their parents eligible for Phase II study were approached for enrolment through their attending pediatric neurologists. Home visits were also done to facilitate the families with bedridden CP children. The pediatrician collected 3 ml of blood from each subject into ethylenediaminetetraacetic acid (EDTA) Tube, BD Vacutainer. Collected blood samples were stored at -20 ºC for genomic DNA (gDNA) extraction.

2.7 DNA Extraction and Quantification

DNA extraction was carried out using Gene All kit (Gene All Biotechnology, Seoul, Korea). Each DNA extraction was started with lysis process by pipetting 20 µl proteinase K solution (20 mg/ml) into a 1.5 ml tube, followed by transferring the 200 µl blood sample into that tube. The mixture was mixed thoroughly, subsequently 200 µl lysis buffer (BL) was added and again followed by vortex-spinning. The mixture was then incubated at 56 ºC for 1 hr 15 min. After incubation, short spin was done to collect any liquid drop inside the lid. Next, 200 µl absolute ethanol (EtOH) was added into the mixture, followed by vortex and short-spin. The lysed sample was then transferred into Min Elute Column using pipette. Centrifugation was done at 19-20 ºC, ≥8000 rpm (≥6000 x g) for 1 min and the collection tube attached at the bottom of the Min Elute Column was replaced with a new one to avoid any contamination with the sample. The washing process began at this stage, when 600 µl of washing buffer (BW) was added to the Min Elute Column followed by centrifugation ≥8000 rpm (≥6000 x g) at 19-20 ºC for 1 min. The collection tube was again replaced.

Washing buffer solution T (TW), 700 µl was added prior to 19-20 ºC centrifugation at ≥8000rpm (≥6000 x g) for 1min. The pass- through was discarded and the collection tube was reinserted, followed by centrifugation at full-speed (13000 x g) for 1 min to remove the residual wash buffer. The Min Elute Column was then placed in a new 1.5 ml tube. Next, the elution step started by adding 30 µl elution buffer (AE) into the Min Elute Column followed by sample incubation at room temperature for 1 hr. This was then followed by centrifugation at full-speed (13000 x g) for 1min. Subsequently, second elution was done by adding another 10 µl AE buffer and the same incubation routine was applied before the full-speed centrifugation was done. All these steps

produced the gDNAs with minimum concentration of 60ng/µl in at least 30 µl volume. Using the microplate reader (Termoscientific, USA), the gDNAs in the collected solution were then measured for their concentrations (A260:A280) and purity values (A260:A230) within the range of optical density (OD) of 1.7 – 2.2 and 2.0 – 2.2 respectively. The quantification results are available in section 3.4 of the Results chapter.

2.8 Gel Electrophoresis

Agarose gel electrophoresis was done to check for DNA integrity. Per each run, the 2% agarose gel was prepared by dissolving 0.8 g of agarose powder in 40 ml 1X TBE in conical flask for 2 min microwave heating. The gel electrophoresis was run in 500 ml 10X Tris-borate-EDTA (TBE) buffer at 90 V for 35 min. Pre-staining using ethidium bromide was applied. The samples were prepared for the gel loading as followed. Each sample well was loaded with 1 µl loading dye and 2 µl gDNA sample, whereas marker well consisted of mixture of 1 µl loading dye, 1 µl Ladder (100bp) and 1µl TBE.

2.9 DNA Library Preparation

Prior to sequencing, the DNA library preparation was carried out at Malaysian Genome Institution (MGI) (Bangi Selangor) used the Nextera Rapid Capture Expanded Exome kit (Illumina Technology) which was an enrichment-based library preparation method. This meant that the method was capturing some off-target regions apart from the exons, which included the untranslated regions (UTRs) and microRNAs (miRNAs). The total covered targeted regions made up to 62 Mb of the genome size. This kit consisted of

transposome enzyme activity and it worked by the Tagmentation principle as described in subsection 2.9.3. The DNA library preparation steps were done simultaneously for all samples so as to reduce any biases and human errors. The whole carried out steps were carried out following the manual book provided by Illumina Technology.

2.9.1 gDNA samples – WES slots Fitting

In this study of six families at the beginning, each family represented by both parents and two CP children, except for Family 3 that had the most subjects of five including both unaffected parents and three CP children (one CP son and two CP daughters), which made up to provisionally 25 total subjects (Table 2.5). However, there were only two sequencing kits provided in this study with a total of only 24 sample slots. Therefore, due to limited sequencing slots, Family 3 of originally five subjects was reduced to only four, including both unaffected parents, one CP son and only one CP daughter with higher DNA concentration. Thus, all six families at this stage were currently represented by four subjects respectively, resulting in total 24 samples to be proceeded for the next procedures that fit to the provided 24 sequencing slots.

2.9.2 Preparation Steps: Index Determination, Quantification and Normalization of gDNA Samples

Prior to Tagmentation step (refer subsection 2.9.3), a few preparation procedures were carried out to prepare the gDNAs for Tagmentation. As explained in section 2.9.1, at this stage only 24 gDNA samples were proceeded, those gDNAs were later organized during the index determination step. In this step, each gDNA sample was assigned with a unique index adapter using Illumina Experiment Manager (IEM) software version 4.0

by following the manual provided in that software. Index determination step was done to distinguish each gDNA as the index acted as unique identifier (tagger) for each gDNA since all gDNA samples that passed the DNA quality test later would be pooled together into two separate pools of DNA library. Indexing was also for the purpose of WES, since multiple gDNA samples with no target gene were processed parallel. The sample index determination details are available in Appendix C.

Next were DNA quantification and normalization, in minimizing the gDNA samples input variability into the Tagmentation step. Illumina Technology strongly recommends a two-step method of gDNA normalization. The gDNAs were quantified using a fluorometric-based method of Qubit. The gDNAs were then diluted and normalized in Elution Buffer (EB) to 10 ng/µl. The first normalized gDNAs were again re-quantified using the same fluorometric-based method. The steps were repeated for the second normalization to a final volume of 10 µl at 5 ng/µl, resulting in 50 ng total of the DNA, as recommended by Illumina Technology.

2.9.3 Genomic DNA Tagmentation

Tagmentation process is a process where the fragmentation and the adapter ligation steps occur simultaneously. While fragmenting the DNA, the transposome will also automatically tag the DNA with the adapter sequences. Up to this stage, all the 24 gDNA samples were still processed simultaneously. In the 96-well midi Nextera Library Tagment (NLT) plate, 10 µl normalized gDNA, 25 µl TD and 15 µl TDE1 were mixed together in each labelled well for each gDNA sample. All of the items were stored at -25 ºC to -15 ºC and thawed on ice before use. The gDNA concentration at this early step was 5 ng/µl in 10 µl EB making up to 50 ng of gDNA per sample. The plate

containing 24 wells of mixture was shaken at 1800 rpm for 1 min, followed by centrifugation at 280 x g for 1 min. The plate was placed on the 58 ºC microheating system with the lid closed, for 10 min. Each well was then added with 15 µl of ST to stop the tagmenting process, followed by shaking of the mixtures at 1800 rpm and centrifugation at 280 x g for 1 min respectively before being incubated for 4 min at room temperature.

2.9.4 Tagmented DNA Clean Up

The tagmented gDNAs were then cleaned-up using Sample Purification Beads (SPB). The clean-up step was done to purify the tagmented DNAs from the Nextera transposomes. The SPB functioned to capture the tagmented gDNA and remove the Nextera Transposome that might still bind to gDNA ends that could interfere with the downstream processes. For this purpose, 65 µl SPB was added to each well, and the plate was again shaken at 1800 rpm for 1 min followed by incubation at room temperature for 8 min and 280 x g for 1 min centrifugation accordingly. The plate was placed on a magnetic stand for 2-5 min until the liquid was clear, indicating that the beads were pulled out from the SPB. The supernatant was discarded from each well, and the tagmented gDNA in each well was washed 2 times by adding 200 µl freshly prepared 80% EtOH to each well. The plate was left on the magnetic stand for 30 s incubation at room temperature, followed by discarding the supernatant from each well of plate later. The pipetting was continued to remove the residual EtOH in each well from the washing step. The plate was air-dried for the 10 min on the magnetic stand. After that, the plate was removed from the magnetic stand and 22.5 µl resuspension buffer (RSB) was added to each well, followed by plate shaking for1 min at 1800 rpm and incubation at room temperature for 2 min. The plate was then centrifuged at 280 x

g for 1 min. Again, the plate was placed on the magnetic stand for 2-5 min until the liquid in each well was clear. The clear liquid (supernatant) now contained the gDNA. From each well, 20 µl supernatant (gDNA input) was transferred into the polymerase chain reaction (PCR) tube respectively. All together were 24 PCR tubes containing 20 µl gDNA sample per tube.

2.9.5 Tagmented DNA Amplification

The purified tagmented gDNAs were then amplified using a 10-cycle PCR. In this step the 2 groups of Index 1 [i7] and Index 2 [i5] together with the sequencing primer were applied to each gDNA during a cluster amplification. Prior to amplification, the Illumina Nextera Index adapters (i5 and i7) were thawed at room temperatures for 20 min and the Library Amplification Mix (NLM) was thawed on ice. The indexes were arranged accordingly and placed on the Truseq Index Plate Fixture to avoid index-swapping. Next, 5 µl of each i7 and i5 was added into the PCR tube of gDNA accordingly, followed by addition of 20 µl of NLM making up to 50 µl mix in total per PCR tube for the amplification (Table 2.6). Each gDNA now has been flanked by 2 groups of Index adapters, (i7) and index 2 (i5). The PCR mixes were shaken at 1200 rpm and centrifuged at 280 x g for 1 min respectively. The 10-cycle PCR was programmed on the thermal cycler as in Figure 2.1.

Table 2.6: The 50 µl PCR mix per tube. The [gDNA] = 2.5 ng/µl in 20 µl RSB with a total of 50 ng gDNA.

Item	Volume (µl)
gDNA	20
i5	5
i7	5
NLM	20
Total	50

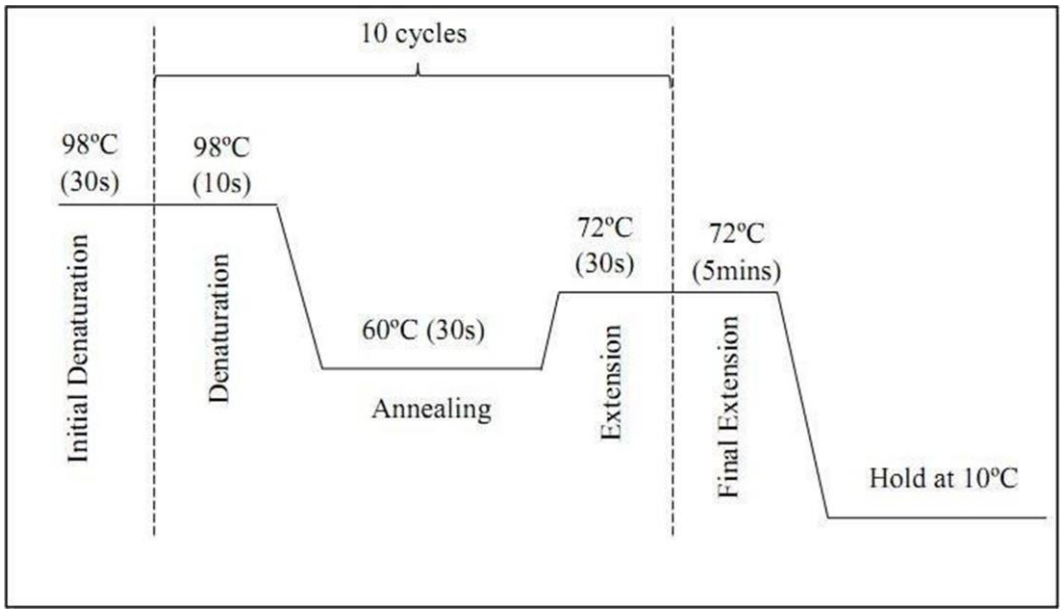


Figure 2.1: PCR Profile according to NLM AMP program on the thermal cycler recommended by Illumina, with option of preheat lid at 100 ºC.

Since the DNA library preparation was a cascading procedure that took more than a day for more than 10 samples, several safe stopping points were available throughout the work. The first safe stopping point was after completing the amplification. The plate was left on the thermal cycle overnight or be stored at 2 ºC – 8 ºC for up to 2 days.

2.9.6 First DNA Quality Checking (QC)

After completing the amplification step, we found that 4 out of 24 samples did not pass the amplification Quality Checking (QC). These 4 gDNA samples had volumes of ≤ 45 µl which were less than the required volume of 50 µl according to Nextera Illumina recommendations for WES. This indicated that these 4 gDNA samples contained < 50 ng gDNA input while the optimum input for WES is 50 ng as recommended by Illumina technology. Therefore starting from this step, from 24 samples, these abovementioned 4 samples that failed the QC were excluded, resulting in only 20 gDNA samples that were proceeded with the next steps and subjected to WES analyses.

2.9.7 Amplified DNA Clean Up

This step is to purify the DNA libraries. Similar to tagmented DNA clean-up step, the amplified DNA clean-up was also done using the SPB while the plate was placed on the magnetic stand. The qualified 20 PCR products from previous step (subsection 2.9.6) were centrifuged at 280 x g for 1 min before being transferred to the 20 corresponding wells of Nextera Library Clean-up (NLC) plate. Into each well of 50 µl PCR product, 90 µl SPB was added and the plate was shaken at 1800 rpm for a minute before being incubated at room temperature for 10 min. The plate was then centrifuged at 280 x g for 1 min before being put on the magnetic stand for 2-5 min, the supernatant

was discarded from each well. Continued with adding 200 µl freshly prepared 80% EtOH into each well, followed by a 30 s incubation on magnetic stand and lastly the supernatant was discarded. These 3 steps, were repeated. Using 20 µl pipette, the residual 80% EtOH was removed from each well, and the plate was air-dried on the magnetic stand for 10 min. The RSB of 27 µl was added into each well before the plate was shaken at 1800 rpm for 1 min and incubated at room temperature for 2 min followed by 1 min of 280 x g centrifugation. The plate was then placed on the magnetic stand for 2 min until the liquid in each well was clear. These clear supernatants in all 20 different wells were our gDNA inputs. From each well, 25 µl of the supernatant was transferred into 20 PCR tube respectively for the next probes hybridization step.

2.9.8 First Probes Hybridization and capturing

This step randomly combined the uniquely-indexed 20 gDNA libraries into two separate pools in which there were 10 gDNA samples per pool using 1.5 ml tubes. Each pool was concentrated using a vacuum concentrator with no heat and a medium drying rate to achieve the volume range of 40 µl $\leq$ X $\leq$ 50 µl according to Illumina technology recommendation. Each pool now contained a mixture of uniquely- indexed 10 gDNAs at 45.1 ng/µl gDNA input in pool 1 (PO1) and 40.2 ng/µl in pool 2 (PO2) respectively. Next, the following items in the order listed in Table 2.7 were added into each of the two PCR tubes.

Table 2.7: PCR mixture tube prepared for the subsequent Hybridization process.

Item	Volume (µl)
DNA Libraries Pool	40
EHB	50
EEX	10
Total	100

The PCR mixture was re-suspended for 10 times followed by centrifugation at 280 x g for 1 min. The PCR tubes containing 100 µl mixture/supernatant were then placed on the programmed thermal cycler and the Nextera Rapid Capture Hybridization (NRC HYB) program was set up to 100 ºC with the preheat lid option and setting of 95 ºC for 10 min. There were 18 cycles started at 94 ºC then decreasing 2 ºC per cycle, each cycle was a 1 min run. The reaction was held at 58 ºC for at least 90 min or up to a maximum of 24 hr. Followed by hybridized probes capturing step in which the targeted regions of the DNA were bound with the captured probes. The capturing step used Streptavidin Magnetic Beads (SMB) to capture probes hybridized to the targeted regions of interest. The two-heated washing principle was used to remove nonspecific binding from the beads producing the enriched libraries. Next, the enriched libraries were eluted from the beads for the second hybridization later. The PCR tubes consisting of probe hybridized mixtures were centrifuged at 280 x g for 1 min.

2.9.9 Second DNA Quality Checking (QC)

Again, the QC was done which was currently to check the DNA quality after first probes hybridization step by excluding the pool that lost >15% of the 100 µl volume (Table 2.7). Greater loss of more than 15% of the actual volume would disqualify the gDNA input from proceeding to the next steps. After passing the QC, 100 µl of PO1 and PO2 was transferred to a well of NEW1 plate, 25 µl SMB was then added into each well. The plate was shaken at 1200 rpm for 5 min, centrifuged at 280 x g for 1 min and then placed on the magnetic stand and was left for 2- 5 min until the liquid was clear. The supernatant was discarded from each well. The plate was left on the magnetic stand for the 2X washing procedure. Next, 200 µl EWS was added into each well, the plate was then shaken at 1800 rpm for 4min and continued with re-suspending the bead pellet in

EWS. The plate was placed on the 50 ºC microheating system with the lid closed for 30 min and placed again on the magnetic stand for 2-5 min until the liquid was clear. The supernatant from each well was again discarded. Those steps starting from EWS addition were repeated for the second time washing. Finishing the second time washing, the plate was removed from the magnetic stand for the elution process. The elution premix for each library pool was prepared in the 1.7 ml microcentrifuge tube by mixing 57 µl of enrichment elution buffer 1 (EE1) with 3 µl of two-normality sodium hydroxide (2N NaOH or HP3). The 60 µl premix was mixed thoroughly before 23 µl of it was added into each well containing PO1 and PO2 respectively.

The plate was then shaken at 1800 rpm for 2 min, incubated at room temperature for 2 min and again being centrifuged at 280 x g for 1 min. The plate was placed on the magnetic stand again for 2min until the liquid was clear. From each well, 21 µl supernatant was transferred to the corresponding PCR tube. Into each PCR tube containing PO1 and PO2, 4 µl of elution target buffer 2 (ET2) was added and the 25 µl mixture per tube were shaken at 1200 rpm for 1 min and centrifuged at 280 x g for 1 min. For the purpose of postponing the procedures till the next day, here was the second Safe Stopping Point whereby the 25 µl supernatant of eluted hybridized probes (PO1 and PO2) then were stored at -25 ºC to -15 ºC for up to 7 days.

2.9.10 Second Probe Hybridization and Capture

This step bound targeted regions of the enriched gDNAs (DNA libraries) in PO1 and PO2 with captured probes for the second time in ensuring high specificity of the captured regions. Second capturing process used the SMB to capture the probes hybridized to the targeted regions of interest. Again, two-heated washing removed the non-specific binding from the beads and the enriched DNA libraries were then eluted from the beads. The same reagents of Enrichment Hybridization Buffer (EHB) and Expanded Exome Oligos (EEX) were used for both first and second hybridization steps. However, for the second hybridization RSB was used re-suspend the enriched DNA libraries.

The second hybridization was done to each 25 µl libraries supernatant of PO1 and PO2 by adding the reagents as in Table 2.8. After adding the reagents, each PCR tube of PO1 and PO2 now consisted of 100 µl supernatant. Each tube was shaken at 1200 rpm for 1 min and followed by centrifugation at 280 x g for also 1 min. It was next placed on the preprogrammed thermal cycler for NRC HYB run process as in the first hybridization step, then followed by capturing step. Both second hybridization and capturing procedures were exactly the same steps as in subsection 2.9.8. However, a different plate was used to transfer the captured libraries supernatant. For the second hybridization step, the well-plate used at this stagewas NEW2 instead of NEW1. After completing the all procedures, each well now consisted of eluted 25µl supernatant of PO1 and PO2 respectively.

Table 2.8: The list of reagents to be added into each PCR tube containing 25µl of PO1 and PO2 respectively.

Item	Volume (µl)
RSB	15
EHB	50
EEX	10
Total	75

2.9.11 Captured Library Clean Up

All cleaning protocols were similar to those described in section 2.9.4 except for some reagent volumes. To clean the captured library, less reagent volumes were applied for example, 45 µl SPB and 25 µl RSB. After finishing all the cleaning protocols, 25 µl of supernatant was transferred from each well (PO1 and PO2) into the PCR tubes. Here was the third Safe Stopping Point as the supernatant could be stored for up to a week at -25 ºC to – 15 ºC.

2.9.12 Enriched Library Amplification

Into each PCR tube consisting 25 µl of PO1 and PO2, 5 µl PPC and 20 µl Enrichment Amplification Mix (NEM) were added accordingly, resulting in 50 µl of mixture/supernatant per tube. Next followed by amplification steps with the same PCR profile as illustrated in Figure 2.1 of section 2.9.5. Here, was the fourth Safe Stopping Point by storing the processed samples at 2 ºC to 8 ºC for up to 2 days.

2.9.13 Amplified Enriched Library Clean Up

The clean-up protocols for the amplified enriched library were the same as described in section 2.9.7 except for the plate used. In this section, the supernatant had been transferred from PCR tubes into the NEC2 midi well-plate. Completing the clean-up protocols, 30µl of supernatant from each well was transferred to the PCR tubes and ready for exome sequencing run.

2.9.14 Enriched Library Quantification

Before being subjected to sequencing, the libraries were checked for quality and size using the Qubit Quantification and Bioanalyzer. The enriched libraries of PO1 and PO2 were checked for quality using Qubit Quantification to ensure optimum cluster densities on the flow cell. The concentration of each pool library was then calculated using the formula below:

$$\frac{\text{(Concentration in ng/µl)}}{\text{(660 g/mol x Average Library Size)}} \times 10^6 = \text{Concentration in nM}$$

The optimum average library size was 400bp as recommended by Illumina Technology for Nextera Expanded Exome Kit that had been used in this study. The PO1 and PO2 libraries were next sent for library size check using Bioanalyzer. This was done to obtain the actual insert (gDNA library) size after deducting 120bp adapter size from transposome activity earlier that flanked our gDNA input.

2.10 Whole Exome Sequencing

The gDNA libraries were sent to UKM Medical Molecular Biology Institute (UMBI) (Cheras, Selangor) for sequencing. For this purpose, Illumina system was employed. The libraries of PO1 and PO2 were subjected to the NextSeq 500 sequencer (High Throughput NGS machine) for generation of paired-end sequencing data.

The sequencing coverage per gDNA was calculated using the formula:

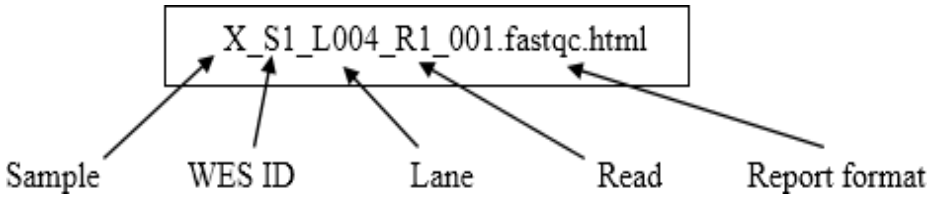
$$\frac{\text{gDNA data size}}{\text{62 Mb}} \times 0.65 \times 0.75 = \text{coverage (X)}$$

62 Mb = Nextera kit standard Genome size covered (Nextera targeted regions)

0.65 = Nextera Kit Capturing Efficiency

0.75 = NextSeq Pass filtered range

Using the Illumina system, each base pair was sequenced one cycle at a time. For example, 100 cycles produced 100 bp reads. The read length (L) is equivalent to the number of bases sequenced. Details of sequencing coverage and read length are described in the Results chapter (section 3.7.1). The WES generated pair-end sequencing read (raw data) were stored in a compressed zip format file (fastq.gz). The Illumina sequencer (NextSeq) also provided the FASTQC inspected raw data reports in fastqc.html files format. The read files showed that one sample had four lanes ({1..4}). Each lane had both forward (R1) and reverse (R2) reads resulting in eight Fastqc file reports per sample. An example of the report file format is shown below.



2.11 Bioinformatics Analysis

WES raw data were interpreted and analyzed using computational codes/languages and tools to discover the variants in multiple genes. Bioinformatics analysis consists of three main processes which are pre-processing, variant discovery and callset refinement. A brief worldwide standard bioinformatics analyses pipeline is illustrated in a workflow in Figure 2.2. Our studied case was CP with underlying genetic factors which indicate that this potentially genetic CP is a germline variant disorder. Hence, we applied the Genome Analysis ToolKit (GATK) package version 4.0 command (available at *https://software.broadinstitute.org/gatk/*), which specifically started at base recalibration step until variant annotation step (Figure 2.2). The earlier steps of raw read analyses, reads alignment and duplicate marking were performed using the worldwide standard pre-processing command with Picard tool package version 2.15.0, (available at *https://broadinstitute.github.io/picard/*).

WES analyses were carried out in collaboration study with Perdana University, Kuala Lumpur. Related raw data, folders and file locations were stored in their server. Since our operating system (OS) was a Microsoft Windows, PuTTY terminal (version 0.70, Simon, Tatham, UK) was used. This terminal is compatible with various versions of Microsoft Windows. The PuTTY terminal acts as a typing interface to access to server, which is termed as Secure Shell (SSH) communication in bioinformatics. Besides that, PuTTY terminal interface was also used for scripting (typing) and executing the bioinformatics instructions for our variant analyses work. In bioinformatics, a scripted computational code instruction that is later executed to analyze the sequence raw data is termed as “command”.

In our study, we also used the Editpad Lite version 7.3.8.0 (Just Great Software) to draft the commands before officially scripting and executing them on the terminal. The command scripting is subject to change according to the file locations, WES sample names/ID and input files that are set by the data owner. A set of multiple bioinformatics commands or also known as bioinformatics coding for this study analyses are available in Appendix D.

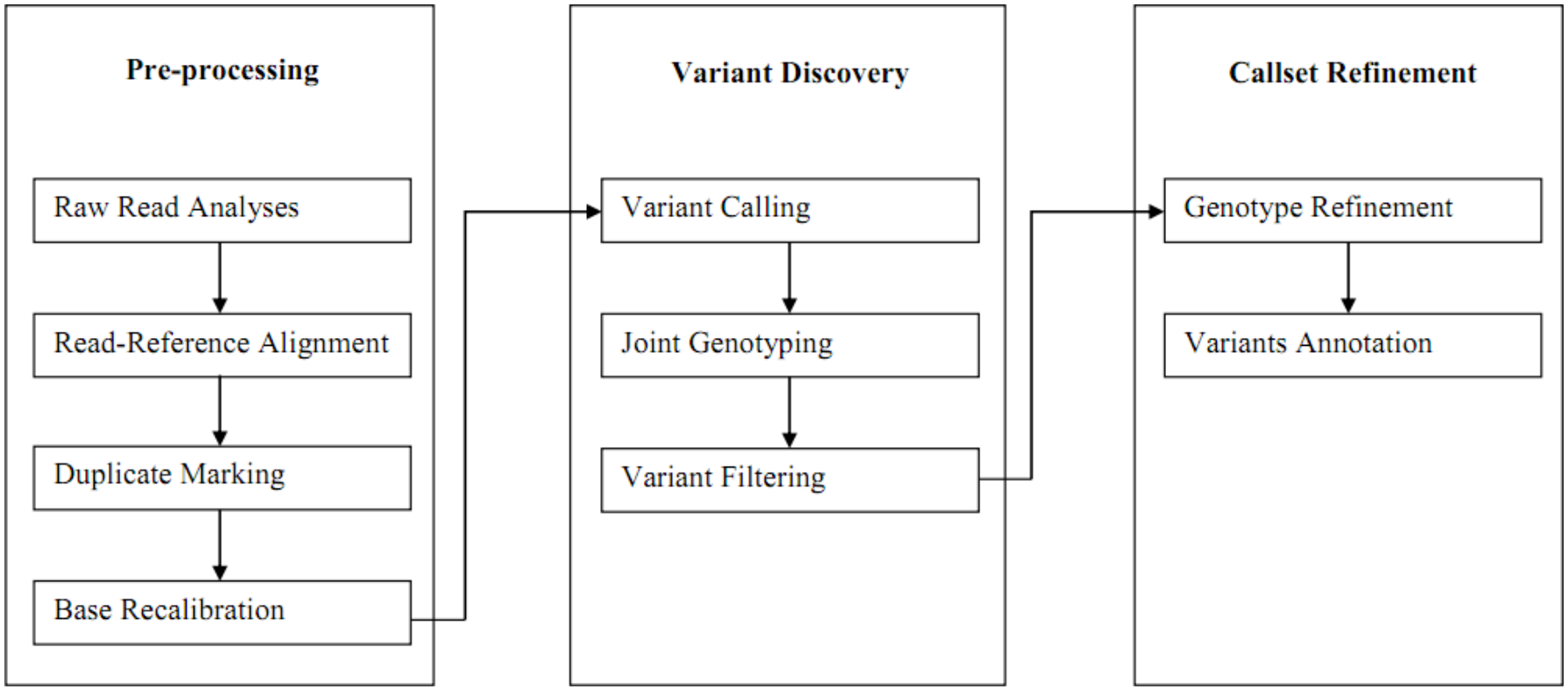


Figure 2.2: The illustration of the standard bioinformatics pipeline for Germline Variants in Whole Exome Sequencing. (*https://software.broadinstitute.org/gatk/)*

2.11.1 Raw Read Analyses

The pre-processing process was carried out to prepare the sequencing raw data, FastQC files (section 2.10) prior to further advance analyses steps. The first step in pre-processing process is raw read analysis which is to check for quality of raw sequencing data. The raw read analyses procedures are simplified in Figure 2.3.

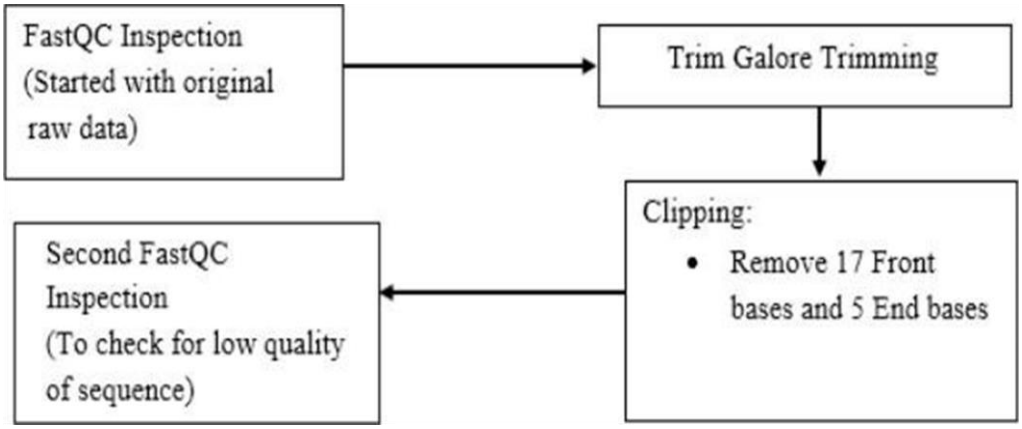


Figure 2.3: The workflow of raw reads preparation prior to further bioinformatics analyses steps.

The current study had two groups of sequencing raw reads which were “batch 1 fastq.gz” and “batch 2 fastq.gz” representing PO1 and PO2, respectively. Starting at this point of raw read analyses, the bioinformatics commands/arguments for the steps of FastQC, trimming and other further steps were scripted and executed on the terminal interface while accessing to the server via SSH. The FastQC inspection was carried out by scripting FastQC tool argument in order to access to fastqc.html files and display the raw sequencing data reports. This was followed by checking for problematic raw reads that were indicated with red circle alert symbol (X). In the current study, original raw data from WES were in two file formats which were the compressed format fastq.gz file (non- viewable) and fastqc.html (viewable report) file. For the computational command purpose, we assigned batch 1 as “array” and batch 2 as “array2” (refer Appendix D.1).

All of the sample reads for 20 subjects had similar base percentage graph that were reported in (.html) file with two biases of base percentage. Therefore, the Trim Galore trimming was carried out to remove unnecessary sequences (biases) and leaving only the gDNA sample sequence. The FastQC tool was again scripted in the command for second FastQC step to access to the report (.html) file after trimming, this is to check for residual low-quality sequences after all the trimming process was completely done. All the scripted commands for the raw read analysis steps are described in Appendix D.1.1, Appendix D.1.2, Appendix D.1.3 and Appendix D.1.4.

2.11.2 Read-Reference Alignment

This is the second step in pre-processing process which is also known as an assembly (align and merge) step. This step is to map the filtered read sequences input to the reference database. The input read batches of the current study were in compressed fastq.gz files (section 2.11.1) at this phase and currently termed as "{element}.gz" for batch 1 reads and "{element2}.gz" for batch 2 reads in the scripted command (refer Appendix D.2). Prior to the assembly step, the human genome database (hg19) or also known as Genome Reference Consortium Human genome build 37 (GRCh37) was downloaded to the server in the indexed hg19 format from the University of California Santa Cruz (UCSC) Genome Browser, a worldwide downloadable browser. The indexed hg19 is a human genome database that is available in Fasta file format (.fa) with companion files of dictionary file format (.dict) and index file format (.fai). The Fasta file is a text-based nucleotide sequence format.

As mentioned earlier in section 2.11, this study applied the GATK bioinformatics pipeline for our germline variants in CP. Thus, GATK pipeline required the indexed hg19 downloaded in the file formats as mentioned above to allow for efficient random access to the reference bases throughout the analysis steps. Next, the {element}.gz file and {element2.gz} file were manifested to unmapped Binary Alignment Map (uBAM) files, in order to re-format the compressed (.gz) files into the intermediate output of compatible files format (.bam) using FastqToSam tool of Picard bioinformatics program (*https://broadinstitute.github.io/picard/*). This was followed by checking read changed quality to produce the clean (.bam) files with forward and reverse pair-reads using the SamToFastq and MarkIlluminaAdapter picard tools with few default parameters.

According to the standard guideline that is available at the website of *https://javadoc.io/static/org.broadinstitute/gatk/4.1.4.1/picard/sam/SamToFastq.html*, few default parameters are used such as Clipping_Attribute, Clipping_Action and Interleave as in scripted command in order to extract read sequences and qualities from the input SAM/BAM file and make a record into the original fastq (.gz) compressed (non-viewable) file. Throughout the analyses of our data, there was also temporary directory (TMP_DIR) command argument in related steps to temporarily store its resources (intermediate output) before processing in the next steps. The next steps were merging and aligning the clean (.bam) files to reference read of hg19 (.fa). These steps were carried out using Burrows-Wheeler Aligner (BWA-MEM) software. The BWA software aligned the {element}, {element2} reads of (.bam) or (.gz) files to various type of genomic variations in hg19 database including Single Nucleotide Polymorphism (SNP), Single Nucleotide Variant (SNV), Insertion and Deletion using MergeBamAlignment picard tool (version 2.15.0). All the commands for the above bioinformatics step were recorded as in Appendix D.2. The intermediate output for this read-reference alignment step were compressed (non-viewable) BAM files of {element}_L00${i}_merged.bam and {element2}_L00${i}_merged.bam for both read batches respectively.

2.11.3 Duplicate Marking

This was part of BAM file cleaning phases whereby the input for this step were the previous {element}_L00${i}_merged.bam and {element2}_L00${i}_merged.bam files containing the reads that had been aligned to a hg19 reference genome. This step was carried out to mark any possible duplicated reads resulting from the sequencing step due to possible uneven amplification of DNA libraries. The duplicate marking was

done using MarkDuplicates tool from Picard program (*http://.picard.sourceforge.net*). The standard principle for duplicate marking is determining the duplicated reads by their 5' mapping coordinates and their genome orientation. Within this step, Picard tool also generated the metrics file (.txt) to distinguish the duplicated reads and index file (.bai), a companion file that acted as an external table of contents which allowed any necessary programs to jump directly to specific parts of the (.bam) files without reading through all the sequences. The intermediate outputs for this analysis step were {element}_L1- 4.bam, {element}_L1- 4.txt and {element}_L1-4.bai for batch 1 reads and {element2)_L1- 4.bam, {element2}_L1-4.txt and {element2}_L1-4.bai for batch 2 reads. The scripted commands for Duplicate Marking are available in Appendix D.3.

2.11.4 Base Recalibration

Besides duplicate marking step, this Base Quality Scores Recalibration (BQSR) was also part of BAM file cleaning phases whereby it also ended the raw data pre-processing process as illustrated earlier in Figure 2.2. This BQSR step is a commonly used GATK quality control program to detect systematic errors that are produced by the sequencing machine which increases the base read calling scores reported from sequencing. This was carried out through four tasks accordingly which were analyzing covariation pattern in sequence dataset, second analyzing on the remaining covariation, generating before and after recalibration plots and applying the recalibrated scores to the sequence data. The {element}_L1- 4.bam for batch 1 and {element2)_L1-4.bam for batch 2 from the duplicate marking step were the current input for BQSR step. The first analysis on covariation pattern was to check for similar pattern of sequence data with the reference sequences, followed by adjusting the quality scores accordingly. This analysis was done using the BaseRecalibrator GATK tool. The process was to build up a model of

covariation based on sequence dataset and a set of known variants database (dbSNP and Mills_and_1000G_gold_standard.indels) and the downloadable list of targeted region (62Mb) of Nextera Rapid Capture Expanded Exome kit in Browser Extensible Data (.bed) file format which is available at *support.illumin.com/sequencing/sequencing_kits/Nextera-rapid-capture-exome-kit/downloads.html*. The raw outputs for this step were {element}_L1-4_recal data.table file and{element2}_L1-4_recal_data.table file for batch1 and batch 2 reads, respectively. This covariation pattern analyzing was repeated for the second checking on both intermediate outputs and generated the next intermediate outputs of {element}_L1- 4_post_recal_data.table file and {element2}_L1-4_post_recal_data.table file, respectively.

This was followed by the third task which was generating before and after recalibration plots to compare between the outputs of first recalibration and second recalibration process using the AnalyzeCovariates GATK tool, in order to see whether or not there was any bias between these two plots. The BQSR recalibrated scores model then, had been applied in the next fourth task. In this task the recalibrations scores in the {element}_L1-4_recal_data.table file and {element2}_L1-4_recal_data.table file were applied to input data of ({element}_L1- 4.bam and {element2)_L1-4.bam) using the PrintReads GATK tool which then produced the intermediate output files of {element}_L1-4_recal_reads.bam and {element2}_L1-4_recal_reads.bam respectively. All the scripted commands executed for this step are available in Appendix D.4.

2.11.5 Variant Calling

This is the first step of variant discovery process in Bioinformatics pipeline. The step was carried out to simultaneously call various possible variants that are present in each sample of {element}_L1-4_recal_reads.bam and {element2}_L1-4_recal_reads.bam inputs, since the current study had two sequencing read batches of 10 samples per batch. Firstly, the HaplotypeCaller GATK tool analyzed multiple individual samples at a time against the reference hg19 and targeted regions of Nextera Rapid Capture Expanded Exome kit to call for variants. This was carried out using the Genome Variant Calling Format (GVCF) mode with genotyping_mode DISCOVERY tool argument to produce the intermediate file format (g.vcf) containing alternate alleles per sample. To run this step, few GATK standard parameters were included such as Standard Emit Confidence 10 (stand_emit_conf.10), Standard Call Confidence (stand_call_conf.30), genotyping_mode DISCOVERY and variant_index_type LINEAR. The stand_emit_conf.10 (minimum variant confidence) and stand_call_conf.30 are the thresholds that determine whether a particular site is a low or high quality variant.

The emit confidence allows the variant to be emitted (presented) in the (.vcf) files. The call confidence then defines the variant quality. For example, if a particular site of variant quality is 10 (QUAL=10), it will be presented in the (.vcf) output files but will be defined/filtered as low quality due to the QUAL that is lesser than the call confidence (QUAL=30). Up to this extent, all the 20 samples of both batch 1 and batch 2 with 10 samples respectively were analyzed resulting into multiple separate individual intermediate (g.vcf) files of genotype likelihood data. The outputs were {element}_L1-4_raw_snps_indels.g.vcf and {element2}_L1- 4_raw_snps_indels.g.vcf. The commands are available in Appendix D.5.1.

2.11.6 Joint Genotyping

The next step in Variant Discovery process is a Joint Genotyping step that was executed to combine all the individual sample (g.vcf) files into one (.vcf) file. The inputs for this step were {element}_L1- 4_raw_snps_indels.g.vcf that contained 10 Sample_L1-4_raw_snps_indels.g.vcf files of batch 1 and {element2}_L1-4_raw_snps_indels.g.vcf that contained another 10 sample_L1-4_raw_snps_indels.g.vcf files for batch 2 sequence. The GenotypeGVCFs GATK tool was used to combine all 20 sample_L1-4_raw_snps_indels.g.vcf files which later resulting in one CombinedBatch1_2AfterJoinGeno.vcf call set file. This was currently an intermediate output in the middle of Variant Discovery process. The commands for this joint genotyping step can be obtained in Appendix D.5.2.

2.11.7 Variant Filtering

Variant filtering is the last step in Variant Discovery process (Figure 2.2). The variants in raw call set which is in (.vcf) file format are filtered using the Variant Quality Score Recalibration (VQSR) model, a machine learning system that identifies the variant profiles and reduces the amount of false-call positive. The VQSR model separates the "good" scored variants from "bad" scored variants using the VariantRecalibrator GATK tool. Variant filtering involves four tasks which are building the SNP recalibration model, applying the desired level of recalibration to the SNPs in the call set, building the Indel recalibration model and applying the desired level of recalibration to the Indels in the call set. In this current study, variant filtering was carried out to filter the raw call set of CombinedBatch1_2AfterJoinGeno.vcf files that was generated from previous joint genotyping step (section 2.11.6). Started with first task of building the SNP

recalibration model, few databases arguments were scripted in the execution command. The VQSR used the gold-standard validated variants from resources such as Haplotype Mapping (HapMap) database (*http://www.hapmap.org*), Omni database (omnidb.org/en/), SNP (dbSNP) database (http://www.ncbi.nlm.nih.gov/SNP/) and 1000 Genome (1000G) project database (doi:10.1038/nature15393) together with hg19 as reference input. Besides that, to build up the VQSR SNP recalibration model, the GATK "truth and training set" for the recalibration model was implemented for each resource as in the following Table 2.9.

Table 2.9: Description of each database as resource for building SNP VQSR recalibration model as recommended by GATK pipeline package (Version 4.0).

Database	Argument in Command	Description	Phred Scale
HapMap	known=false, training=true truth=true	The dataset was used as the training resource SNVs in this dataset were all the true variants	Prior=15.0
Omni	known=false, training=true truth=true	The dataset was used as the training resource SNVs in this dataset were all the true variants	Prior=12.0
1000G	known=false, training=true truth=false	The dataset was used as the training resource SNVs in this dataset could be both true or false- positive variants	Prior=10.0
dbSNP	known=true, training=false truth=false	The dataset was used as the known sites resource SNVs in this dataset could be both true or false- positive variants	Prior=2.0

These resources in (.vcf) files format are also available in GATK bundle files (*https://software.broadinstitute.org/gatk/documentation/article?id=11050*). Involved in this SNP VQSR recalibration model are few default variant annotation parameters for germline WES sample. There are Quality by Depth (QD), Sequencing Depth/Coverage (DP), Fisher Strand (FS), Strand Odds Ratio (SOR), Mapping Quality (MQ), Mapping Quality Rank Sum test (MQRankSum), Read Position Rank Sum test (ReadPosRankSum) to evaluate the likelihood of true positive calls. For the current study VQSR threshold parameter (tranches) of 100%, 99% and 90% arguments were included to obtain the threshold of VQSR Log Odds (VQSRLOD), a threshold of true variant score versus false variant score. Followed by the second task, this SNP VQSR recalibration model was then applied to the study dataset, CombinedBatch1_2AfterJoinGeno.vcf file using the SNP mode by ApplyRecalibration execution tool with SNP mode.

During this process, the VQSR searched for variants in the CombinedBatch1_2AfterJoinGeno.vcf input file against all these resources to trace the SNP variants that overlapped with the validated SNP variants across the resource datasets. At this stage, all the SNPs were already annotated with their recalibrated quality scores whereby their “PASS” or “FILTER” outcome depended on whether or not they were included in the tranches set up earlier. The intermediate output for this step was recalibrated_snps_raw_indels.vcf file which was later used as input for the next step for variant recalibration model. The above two steps of building the model and applying it to SNPs in dataset were repeated as the third and fourth tasks of variant filtering process in order to filter Indels in dataset. The involved databases to build the Indels recalibration model are listed in Table 2.10.

Table 2.10: Description of each database as a resource for building an Indels VQSR recalibration model as recommended by GATK pipeline (Version 4.0).

Database	Argument in Command	Description	Phred Scale
mills	known=false, training=true truth=true	The dataset was used as the training resource SNVs in this dataset were all the true variants	Prior=12.0
dbSNP	known=true, training=false truth=false	The dataset was used as the known sites resource SNVs in this dataset could be both true or false-positive variants	Prior=2.0

In this third task of building the Indels recalibration model, the same parameters as in SNP recalibration model building process were used. Besides that, an additional parameter of Maximum Gaussians (maxGaussians) was also used. The Gaussians is the probability distribution for the statistical model, whereby for the VQSR model it is used for variants positive call model with smaller sample size. The default distribution value of maxGaussians is eight. However, in this study GATK recommended to lower to four in order to increase the number of positive variants call since the number of samples study was less than 30 (WES sample size according to GATK requirement). This was done to avoid any error reported when executing the command. Similar to second task earlier, this current fourth task was carried out to apply this Indels recalibration model to filter the Indels in dataset. This was carried out by using the Indel mode argument in scripted commands. The full command for Variant Filtering was available in Appendix D.6. The intermediate output for this step was reported in recalibrated_variants.vcf file format.

2.11.8 Genotype Refinement

This is the first step of the third main bioinformatics pipeline of Call set Refinement as illustrated in Figure 2.2 earlier. This step involves three tasks which are deriving posterior probabilities of genotypes, filtering low quality genotypes and annotating possible de novo mutations. The first task was carried out to derive posterior probabilities of genotypes using CalculateGenotypePosteriors tool. This was to improve the accuracy of genotype calls in our recalibrated_variants.vcf file by analyzing against an additional database as provided in GATK pipeline, 1000G_phase3_v4_20130502 database. Besides that, in fulfilling this task, the pedigree (.ped) file format was

generated using the EditpadLite version 7.3.8.0 (Just Great Software) for all families in this study. This (.ped) file was also another additional supporting generated data for the analyses since the current study was not represented by only affected children but also included the unaffected parents. This required (.ped) file described the familial relationship between samples. This task then, generated an intermediate output of recalibratedVariants.postCGP.vcf file which was later being processed in the second task called filtering low quality genotypes. In this second task, the VariantFiltration GATK tool was used to filter genotype calls that were uncertain whether or not carrying the variants. Below than 20 score of Genotype Quality (GQ<20) in recalibratedVariants.postCGP.vcf dataset will be filtered as "lowGQ" indicating that these particular genotypes were not suitable for the next downstream analysis steps. This was followed by the third task which was annotating possible de novo mutations in the recalibratedVariants.postCGP.vcf dataset using the VariantAnnotator GATK tool, as per scripted in the command available in Appendix D.7. This produced the recalibratedVariants.postCGP.Gfiltered.deNovos.vcf output file that also contained a list of affected children samples with possible de novo mutations.

2.11.9 Variants Annotation

Variants annotation is the final step for the Bioinformatics pipeline which was executed using the Annotate Variations (ANNOVAR) software tool version 2018 (frequently updated version) to annotate the data of chromosome (chr), start position, end position, reference nucleotide and observed nucleotides in (.vcf) dataset of the study, resulting in ANNOVAR table of annotated CP SNPs and Indels. The ANNOVAR annotation was carried out through four operations which were gene-based annotation (g), region-based

annotation (f) and filter based annotation (f) databases that provided by ANNOVAR as in Table 2.11.

Table 2.11: The list of default databases provided in the ANNOVAR software which were used in variant annotation step.

Build Version	Table Name (protocol)	Description	Operation
hg19	1000g2015aug_all	The 1000G team fixed bug in chrX frequency calculation.	f
hg19	avsift	Whole exome SIFT scores for non- synonymous variants	f
hg19	cg69	Allele frequency in 69 human subjects sequenced by Complete Genomics organization.	f
hg19	clinvar_20140929	CLINVAR database with Variant Clinical Significance (Unknown, histocompatibility, untested, pathogenic, non-pathogenic, probable-non-pathogenic, probable-pathogenic,drug-response) and Variant disease name.	f
hg19	clinvar_20170130	CLINVAR version 20170130 with separate columns (CLINSIG CLNDBN CLNACC CLNDSDB CLNDSDBID)	f
hg19	cosmic70	COSMIC database version 68 on Whole Genome Sequencing (WGS) data	f
hg19	cytoBand	Specific genomic regions (chromosome band) of genes	r
hg19	ensGene	FASTA format sequences for all annotated transcripts in ENSEMBL gene database	g
hg19	esp6500siv2_all	Exome Sequencing Project (ESP) with 6500 exomes, including the indel calls and the chrY calls.	f

Table 2.11: Continued.

Build Version	Table Name (protocol)	Description	Operation
hg19	exac03	ExAC 6500 exome allele frequency data for all (ALL), African (AFR), Admixed American (AMR), East Asian (EAS), Finnish (FIN), Non-finnish European (NFE), South Asian (SAS) and other (OTH).	f
hg19	exac03nontcga	ExAC on non-TCGA samples	f
hg19	hrcr1	40 million variants from 32K samples in haplotype reference consortium	f
hg19	icgc21	International Cancer Genome Consortium version 21	f
hg19	kaviar_2015092 3	Compilation of SNVs, Indels, complex variants in human	f
hg19	kgXref	Known gene cross-reference, which is opened by default below the knownGene table in ANNOVAR annotation.	f
hg19	knownGene	FASTA format sequences for all annotated transcripts in USCS known gene database	g
hg19	ljb26_all	ANNOVAR generic database	f
hg19	nci60	NCI-60 human tumor cell line panel exome sequencing allele frequency data.	f
hg19	refGene	sequences for all annotated transcripts in RefSeq gene database.	g
hg19	snp137	The variants with reference SNP cluster ID (rs number) (version GRCh37)	f
hg19	snp138	The variants with reference SNP cluster ID (rs number) (version GRCh38)	f

In this step, the applied databases were termed as “protocol” (Table 2.11), meanwhile the term “symbol” was called as “string” for command scripting purpose. The dot (.) string that defines “not available string” (nastring) was used in scripted command to annotate any scores and/or data that are not available in database. The full command for this step is available in Appendix D.8. The ANNOVAR annotation was displayed on the terminal in table (.annovar) format. By generating the final output from ANNOVAR annotation, this ended the bioinformatics analysis phase. This bioinformatics analysis final output was then interpreted to generate the main study results.

2.12 Data Interpretation

The ANNOVAR output was then translated into an excel file (.xlsx) format consisting of a big data annotated variants list. However, this output was not the official study data since this big data consisted of variants include the SNPs and mutations. For this preliminary study of CP related genes, the limit was set to focus on the mutations only and exclude the SNPs since SNPs were known to be easily polymorphic and abundantly present (>1%) in the population. The variants were also checked for their domains and protein sequences through online database, ensemble.org. The selected variants were then determined for their pathogenicity according to American College Medical Genetics standard guidelines (ACMG) (Richards et. al., 2015).

CHAPTER THREE

RESULTS

3.1 Subject Details

Instead of provisional 25 subjects (Table 2.5) at the beginning of the study, the main study output was generated from 20 final subjects only (Table 3.1) due to two-time sample reductions. The chronologies of two-time sample reductions are as explained in detail in subsection 2.9.1 and subsection 2.9.6. These final list of 20 study subjects are comprised of Family 1, Family 2, Family 4 and Family 5 with four subjects per family, whereby each of these four families are represented by both unaffected parents and two CP children. However, Family 3 with three subjects only, i.e. both unaffected parents and a CP daughter. Family 6 contributes only one CP son as a subjects with no parents' WES data, due to fail the DNA quality test (refer subsection 2.9.6).

3.1.1 Identification Code for Final Study Subjects

These 20 study subjects consist of five unaffected mothers, five unaffected fathers, three CP daughters and seven CP sons were randomly grouped into two different pools for two separate sequencing batches (Table 3.1). Pool 1 (PO1) was assigned for batch 1 and Pool 2 (PO2) was for batch 2. In order to distinguish these 20 final study subjects from the provisionally 25 recruited subjects (Table 2.5), their IDs are currently improvised by assigning each subject with different unique WES ID (S#) (Table 3.1).

Example of the subject/sample's improvised ID format is as follows:

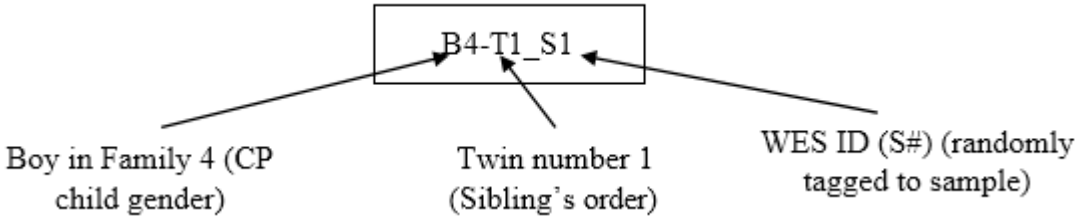


Table 3.1: The final list of 20 subjects with WES ID (S#) at the end.

WES Pool (PO)	Fam 1	Fam 2	Fam 3	Fam 4	Fam 5	Fam 6	Number of WES sample per pool
PO1	D1_S3	D2_S7	D3_S9				10
	M1_S2	M2_S6	G3-3_S10				
	B1-1_S1	G2-1_S8					
	B1-3_S4	B2-2_S5					
PO2			M3_S6	D4_S4	D5_S9		10
				M4_S3	M5_S8		
				B4-T1_S1	B5-1_S7		
				B4-T2_S2	G5-2_S10	B6-1_S5	
						Total	20

(Note: Fam 1 to 6= Family 1 to Family 6; D= Father; M= Mother; B= Boy; G= Girl; T= Twin; 1 to 6 (after B, D, G and M) = Family ID; 1 to 3 (after "-") = Sibling's order; S# (#: 1 to 10) after "_" = reads ID per WES pool).

3.2 Family Pedigree

A three-generation pedigree was constructed for each family in the study.

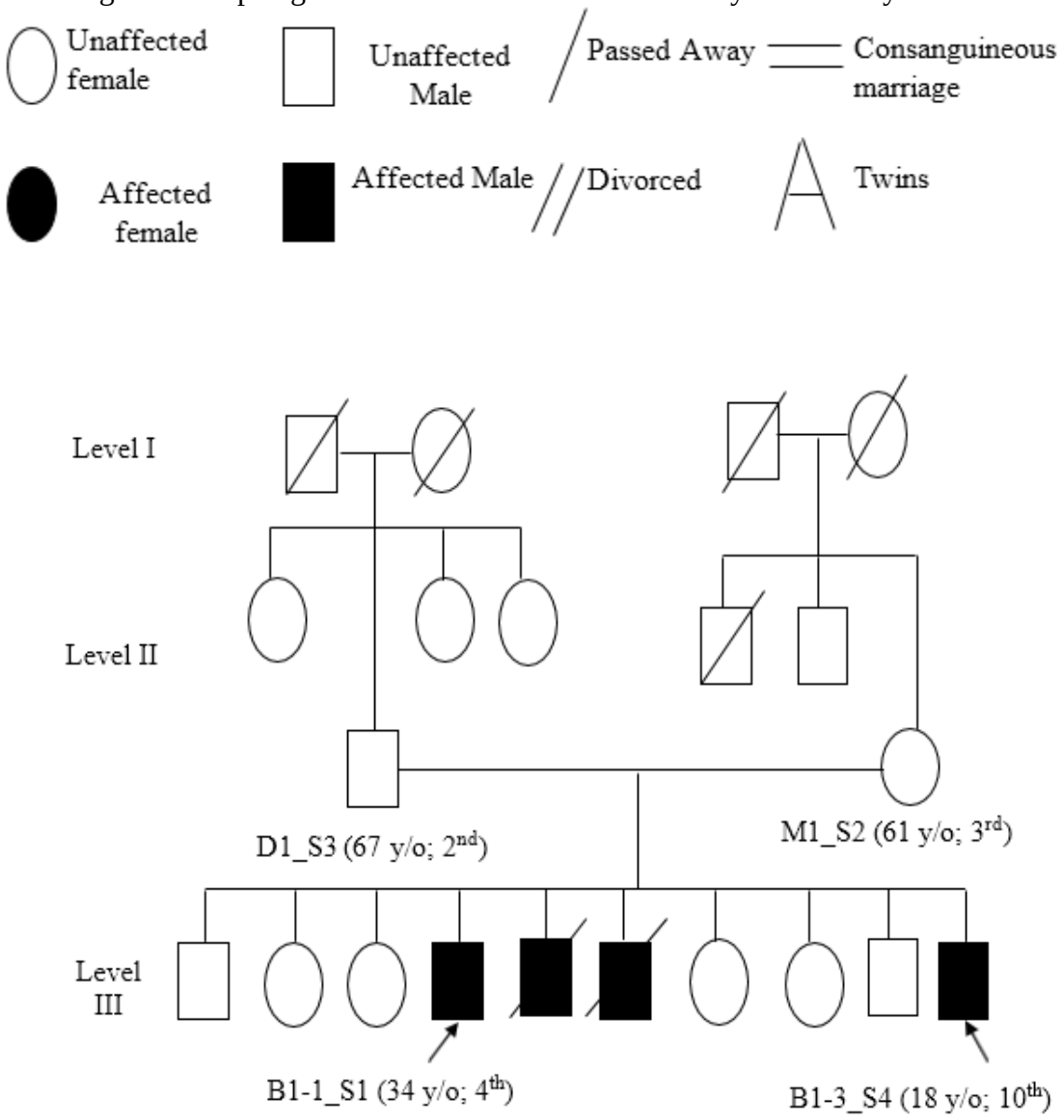


Figure 3.1: The three-generation pedigree of Family 1. (y/o: years old, 1st to 10th: sibling's order).

Figure 3.1 shows the three-generation pedigree of Family 1 with 10 children of both unaffected parents in a non-consanguineous marriage. Out of these 10 siblings there are four male siblings with CP in which only two of them are surviving. They are 34 years old CP son and 18 years old CP son who were born via spontaneous vagina delivery (SVD) with no antenatal, intrapartum and postpartum events. The other two CP sons were the fifth and sixth children who passed away at the age of 31 years old and 11 years old respectively.

Both unaffected parents (D1_S3 and M1_S2) and these two CP sons of spastic diplegia, B1-1_S1 and spastic quadriplegia, B1-3_S4 were included in the final list of 20 subjects. These two CP sons are bedridden with no means of mobilizing on their own and fully dependent on caregivers. None of the family members in the previous generation was affected by CP nor having other identifiable genetic disorders. However, there is another ataxic CP grandson (a son of the third female child) which is in the fourth generation, but he is not included in this third-generation pedigree. This however suggests possible genetic factor for familial CP in Family 1.

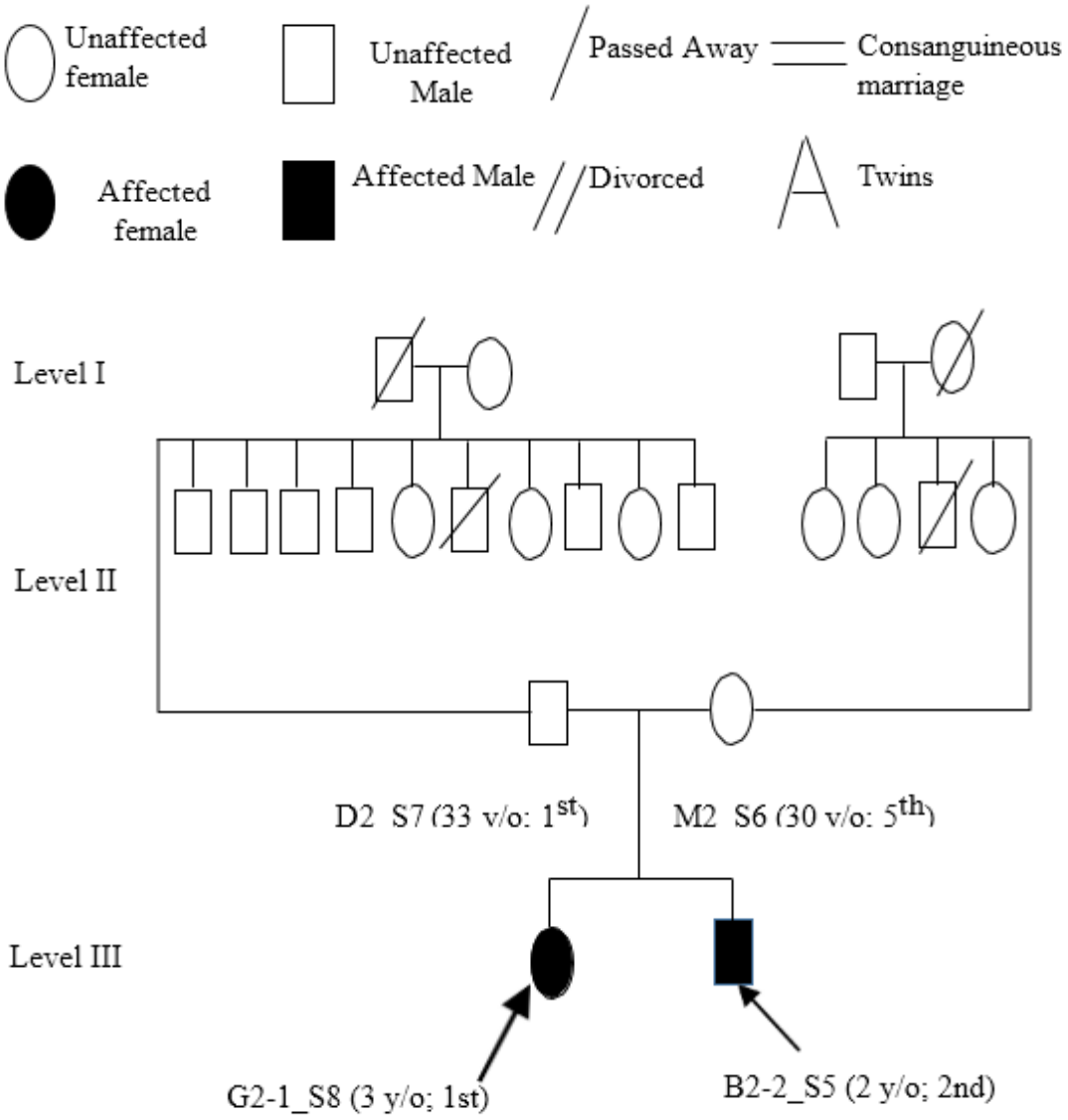


Figure 3.2: The three-generation pedigree of Family 2 (y/o: years old, 1st to 5th: sibling's order).

Figure 3.2 shows three-generation pedigree of non-consanguineous Family 2 with two children. Both siblings are spastic diplegia CP of 3 years old female and 2 years old male. These CP children were born via SVD with no risk factor incidents. They are also mobilizing using wheelchair with help from caregivers, do not need the trunk support device and able to grasp objects. Both unaffected parents (D2_S7 and M2_S6) and both CP children (G2-1_S8 and B2-2_S5) are also included in the final list of 20 study subjects (Table 3.1). There was no CP history nor other identifiable genetic disorders in the previous generations. However, since both children in Family 2 are CP affected despite of uneventful delivery and no risk factors, this indicates that this neurological condition is possibly considered as familial CP.

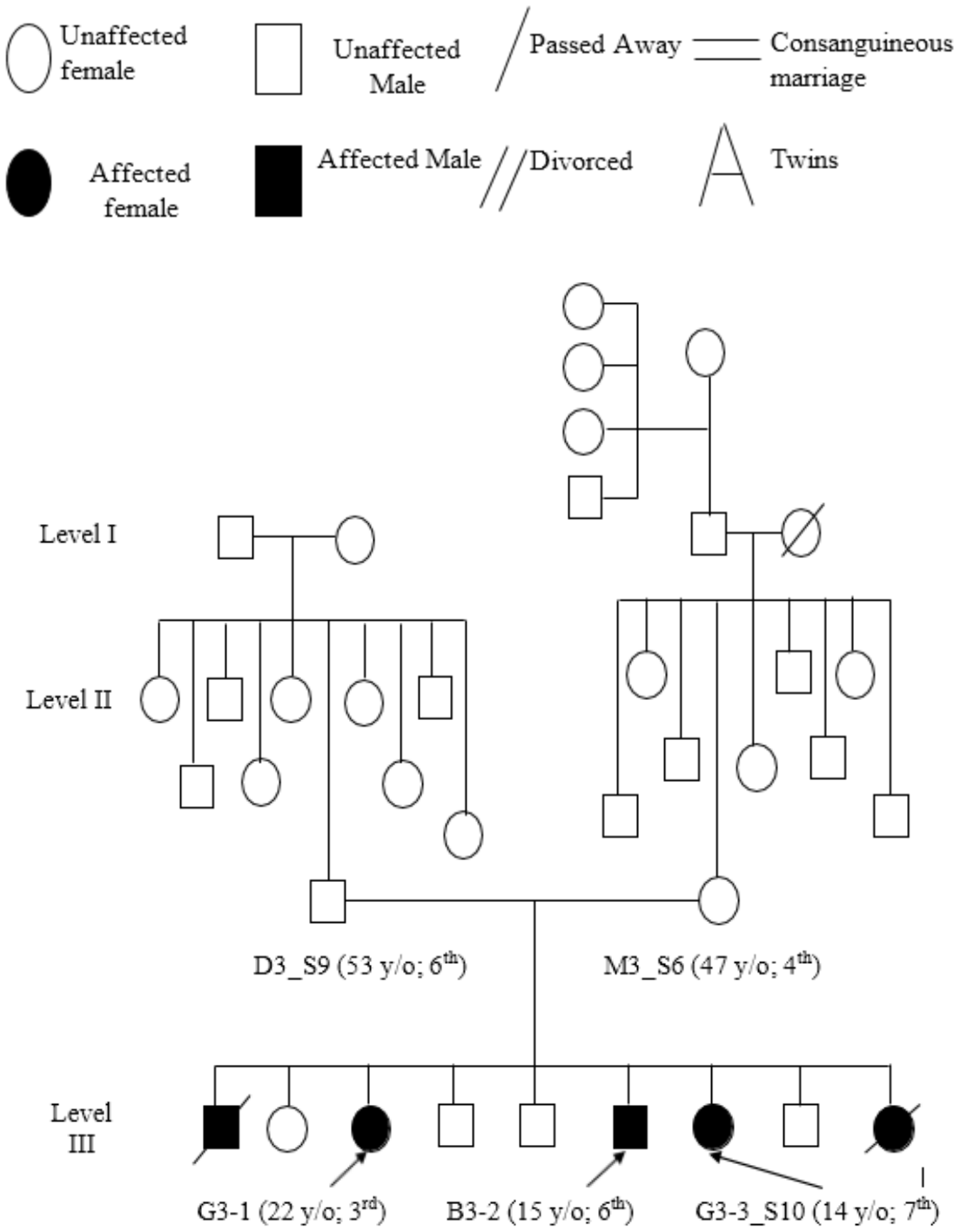


Figure 3.3: The three-generation pedigree of Family 3 (y/o: years old, 1st to 7th: sibling's order).

The Family 3 three-generation pedigree shows a pair of normal parents in a non-consanguineous marriage with nine children. Five out these nine siblings are CP affected, however only three CP children are still surviving. Meanwhile, the other two CP children who were first son and youngest daughter died at 6 months old and 7 years old respectively. All of these spastic CP children who were born through SVD with no mechanical factors are bedridden and fully caregivers-dependent. Only both parents (D3_S9 and M3_S6) and one CP daughter (G3-3_S10) are included in the final list of 20 subjects (Table 3.1). Regardless of no risk factors and no CP history in the past generations, having multiple CP children from a same parental line suggests that this Family 3 is also a familial genetic CP.

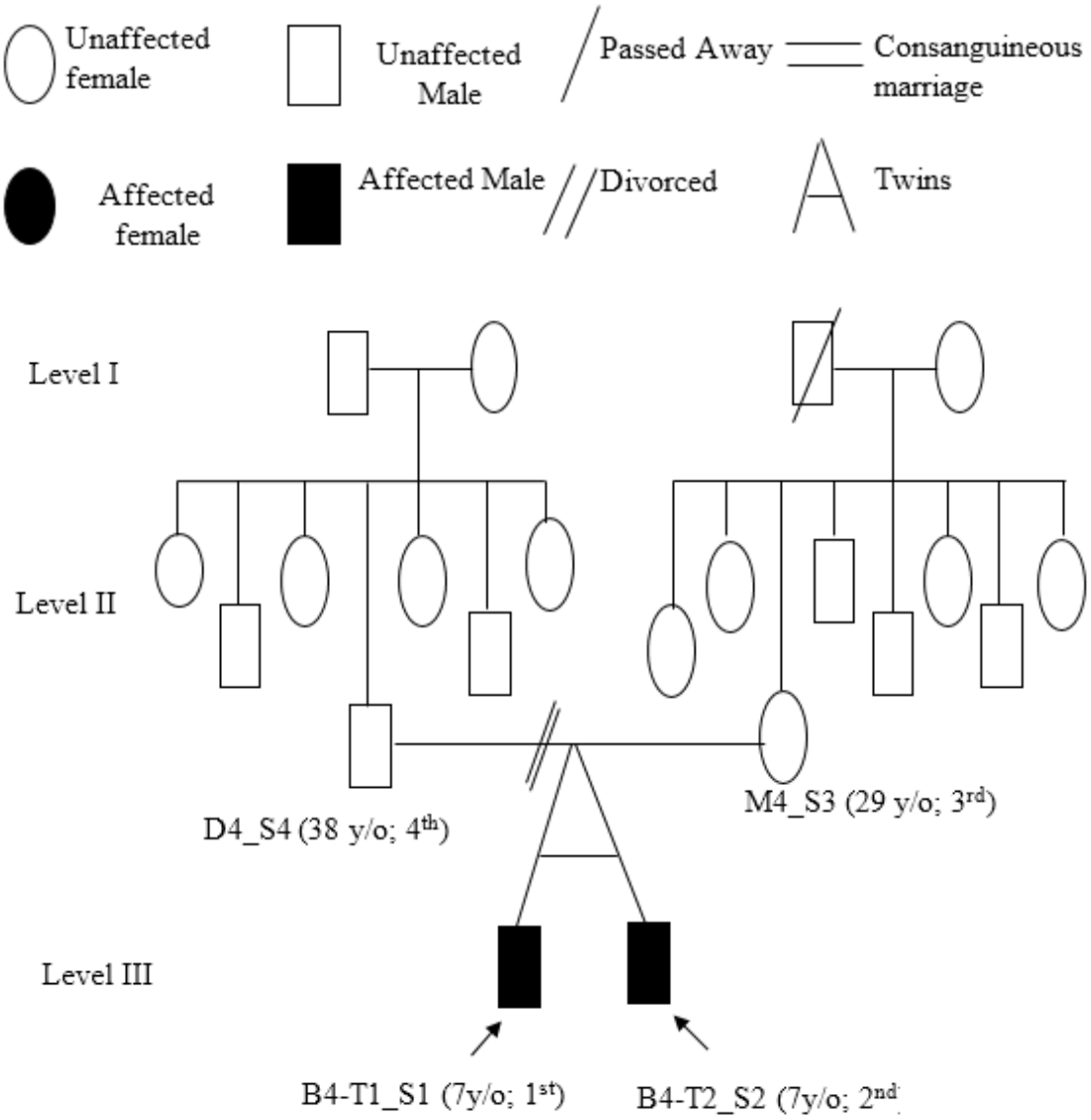


Figure 3.4: The three-generation pedigree of Family 4 (y/o: years old, 1st to 4th: sibling's order).

Figure 3.4 shows three-generation pedigree of Family 4 with a pair of monochorionic diamniotic (monozygotic) seven years old male twins of B4-T1_S1 and B4-T2_S2. The twins together with their parents (D4_S4 and M4_S3) are all included in the 20 final study subjects that generated the main study output. The twins are classified with different CP type which are spastic diplegia B4-T1_S1 and spastic left hemiplegia B4-T1_S2. The spastic diplegia B4-T1_S1 is mobilizing using wheelchair and able to propel it himself meanwhile the spastic hemiplegia B4-T2_S2 is walking with hemiplegic gait. Their divorced parents were a non-consanguineous pair who also had no CP history from past generations of each family line. The twins were born via emergency caesarian at gestational age of 30 weeks, whereby prematurity is one of the subject inclusion criteria for this current CP genetic study. They also fulfilled another two inclusion criteria which are CP diagnosed confirmation and more than one family members are CP affected.

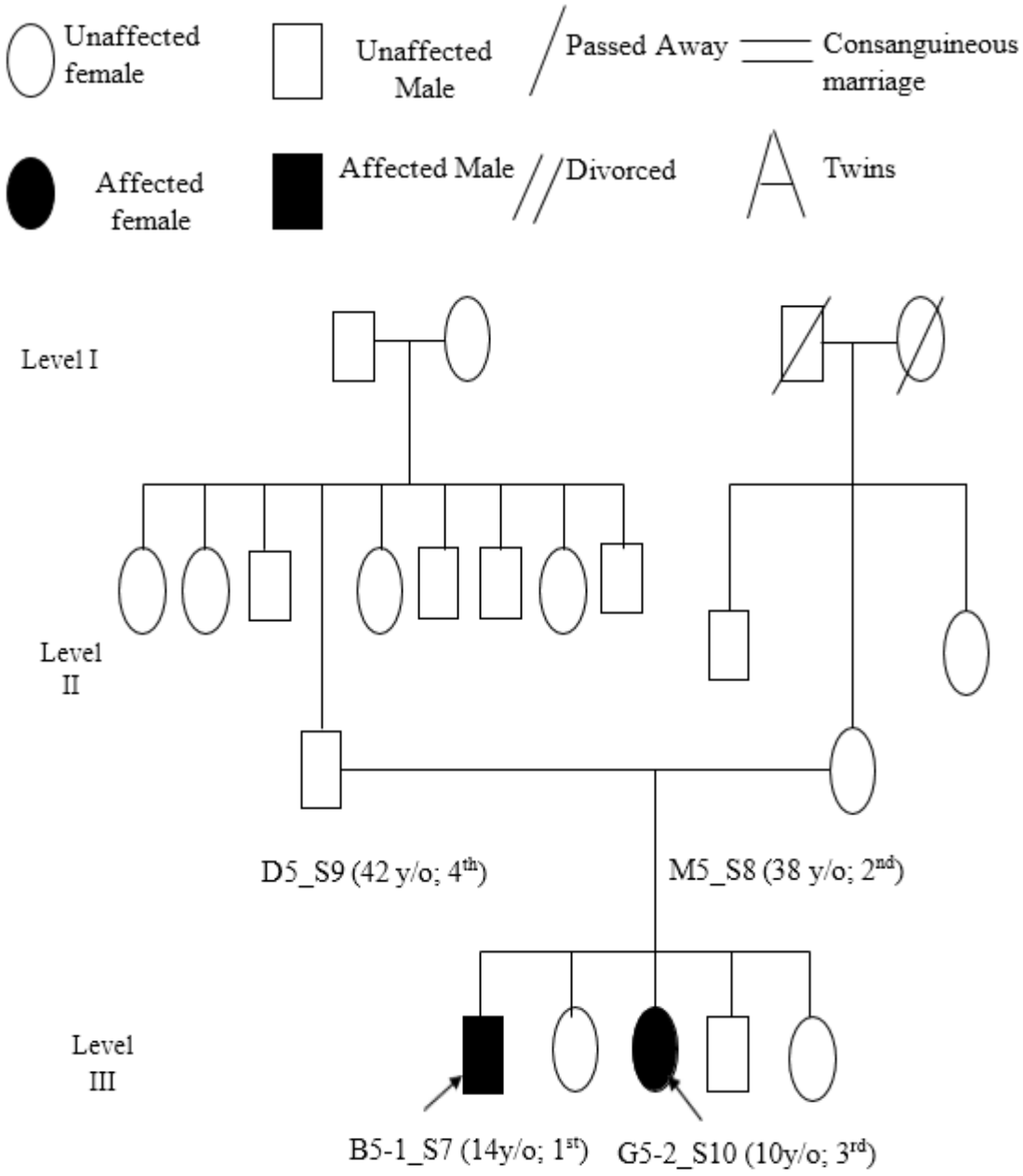


Figure 3.5: The three-generation pedigree of Family 5 (y/o: years old, 1st to 4th: sibling's order).

Family 5 three-generation pedigree in figure 3.5 shows that two out of five children are CP affected who were born to non-consanguineous unaffected parents. Both unaffected parents (D5_S9 and M5_S8), a spastic diplegia son (B5-1_S7) ages 14 years old and a spastic diplegia (G5-2_S10) daughter ages 10 years old are the subjects for this current study. These two CP children have the same CP type however show different physical manifestations. B5-1_S7 child is able to walk but with abnormal gait and has functional hands, meanwhile G5-2_S10 child is mobilizing using wheelchair and able to propel wheelchair by herself. They were both born via SVD with no causal event occurs. However, B5-1_S7 was born as a premature infant at 33 weeks. Even though there was no CP history in the previous generation, this family is considered as familial genetic CP since they fulfilled more than three inclusion criteria which are CP diagnosed children, having more than one CP children in a family and born premature.

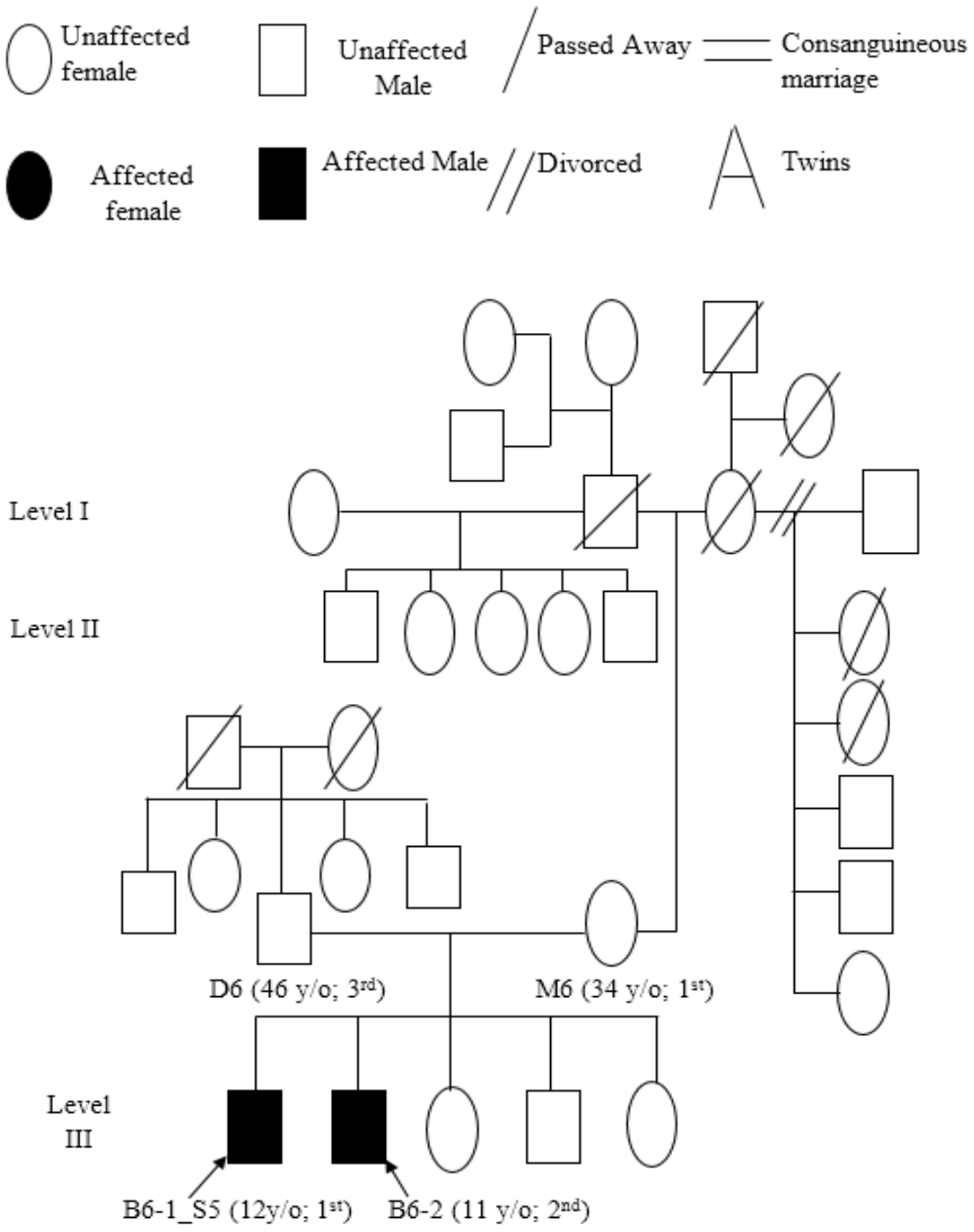


Figure 3.6: The three-generation pedigree of Family 6 (y/o: years old, 1st to 3rd: sibling's order).

Figure 3.6 shows Family 6 three-generation pedigree of non-consanguineous unaffected parents with two spastic quadriplegia CP sons (12 years old and 11 years old) that were born via SVD and another three unaffected children. Both CP sons are bedridden and fully caregivers-dependent with no means of mobilizing themselves. Besides Family 3, this Family 6 also had some WES excluded subjects/samples which were both parents' (D6 and M6) and second CP son (B6-2) because of did not pass the DNA quality test. Hence, from this family only one subject (B6-1_S5) is included in the 20 final study subjects to generate study output (Table 3.1). The pedigree shows that there was no previous CP history in the past generations.

3.3 Clinical Features of CP Subjects

The clinical data of CP subjects were recorded from during home-visit for physical examination and from hospital medical records especially for the diagnosis phase (≤5 years old) data. These clinical data are summarized in Table 3.2. In total, there are 10 CP subjects ranging from 2 years old to 34 years old in this study in which seven are male CP subjects and three are female CP subjects.

Table 3.2: The summary of clinical features (DD, LD, and muscle conditions) of all 10 CP affected subjects.

Case	Gender	Age (y/o)	DD	LD	Muscle Tone	Deep Tendon Reflex	Muscle Wasting	Muscle Weakness
Family 1								
B1-1_S1	♂	34	/	/	↑	↑	/	/
B1-3_S4	♂	18	/	/	↑	↑	/	/
Family 2								
G2-1_S8	♀	3	/	NA	↑	↑	/	/
B2-2_S5	♂	2	/	NA	↑	↑	/	/
Family 3								
G3-3_S10	♀	18	/	/	↑	↑	/	/
Family 4								
B4-T1_S1	♂	7	/	Mild	↑	↑	/	/
B4-T2_S2	♂	7	/	/	↑	↑	/	/
Family 5								
B5-1_S7	♂	14	/	X	↑	↑	/	/
G5-2_S10	♀	10	/	X	↑	↑	/	/
Family 6								
B6-1_S5	♂	12	/	/	↑	↑	/	/

(Note: DD= Developmental Delay; MR= Mental Retardation; LD= Learning Disability; ♀= Female; ♂= Male; ↑= hyper; ↓= hypo; /= Yes (problem present); X= No (problem absent); NA= Not applicable for children < 5 years old).

Table 3.2 shows the clinical data for each CP subjects in this study that are tabulated according to family group. Medical records for these CP subjects show that they were observed for their nervous system function by examining for delay in their developmental milestones. The developmental delay assessment is done during the phase of ≤ 5 years old. Based on hospital medical records, it is shown that all of these CP subjects have developmental delay problem that are reflected by few inabilities such as unable to roll over, unable to sit even with support, delay and/or unable to walk and hand fisting. They were also assessed for the higher mental function (learning disability assessment). Both CP sons in Family 1 (diplegic B1-1_S1 and quadriplegic B1-3_S4), CP daughter (quadriplegic G3-3_S10) in Family 3 and single subject of Family 6 (quadriplegic B6- 1_S5) have learning disability.

In Family 4, both CP twins have learning disability, but it is milder in B4-T1_S1 compared to B4-T2_S2. Both diplegic children in Family 5 (B5-1_S7 and G5-2_S10) have no learning disability. Meanwhile, both CP children in Family 2 who are 3 years old G2- 1_S8 daughter and 2 years old B2-2_S5 son do not have any records for learning disability assessment (NA) as in Table 3.2. This is because the assessment is not applicable for patients aged < 5 years old. From these 10 CP subjects, only four (B4-T1_S1, B4-T2_S2, B5-1_S7 and G5-2_S10) who are able to talk, however only B5-1_S7 and B4-T1_S1 who are attending normal school. Besides obtaining the clinical data from hospital medical records, physical examination data were collected during the home visits to check for their current physical conditions. The physical examination was done for muscle conditions (Table 3.2), motor dysfunction types and motor function that includes sensory problem, cerebellar dysfunction, cranial nerve abnormalities and (Table 3.3). As for muscle tone as shown in Table 3.2, all CP subjects

have hypertonic muscle tone. Hypertonia causes them to have increased muscle tension while hypotonia caused them to have decreased muscle tension. Both muscle tone conditions are noticeable although the patient is at rest without doing any movements. As recorded in Table 3.2, these 10 CP subjects are showing hyperreflex of deep tendon. Hyperreflex tendon is a condition of an increasing speed in their reflex or also known as excessive response. All of these CP subjects also have both muscle wasting (deteriorated muscle) and muscle weakness problem which contribute to their restricted movement.

Table 3.3: The summary of clinical features (motor function, type of plegia, type of motor dysfunction and mobility) of all 10 CP affected subjects.

Case	Gender	Age (y/o)	Cerebellar Dysfunction	Sensory Problem	Cranial Nerve Abnormalities	Topographic Distributions	Flaccid /Spastic	Mobility (GMFCS)
Family 1								
B1-1_S1	♂	34	X	X	X	Diplegia	Spastic	Bedridden (Level V)
B1-3_S4	♂	18	X	X	X	Quadriplegia	Spastic	Bedridden (Level V)
Family 2								
G2-1_S8	♀	3	X	X	X	Diplegia	Spastic	Stroller (Level IV)
B2-2_S5	♂	2	X	X	X	Diplegia	Spastic	Stroller (Level IV)
Family 3								
G3-3_S10	♀	18	X	X	X	Quadriplegia	Spastic	Bedridden (Level V)
Family 4								
B4-T1_S1	♂	7	X	X	X	Diplegia	Spastic	Wheelchair (Level III)
B4-T2_S2	♂	7	X	X	X	Hemiplegia	Spastic	Hemiplegic gait (Level II)
Family 5								
B5-1_S7	♂	14	X	X	X	Diplegia	Spastic	Abnormal gait (Level II)
G5-2_S10	♀	10	X	X	X	Diplegia	Spastic	Wheelchair (Level III)
Family 6								
B6-1_S5	♂	12	X	X	X	Quadriplegia	Spastic	Bedridden (Level V)

(Note: ♀= Female; ♂= Male; / = Yes (problem present); and X= No (problem absent))

Table 3.3 shows clinical data for motor function (cerebellar dysfunction, sensory problem and cranial nerve abnormalities), topographic distribution, type of motor dysfunction (flaccid/spastic) and mobility. All of these 10 spastic CP subjects have no cerebellar dysfunction observed. They also show no sensory involvement and cranial nerve abnormalities as recorded in Table 3.3. There are four bedridden CP subjects with impaired truncal and fully caretaker-dependent, which are both CP sons in Family 1 (B1_1_S1 and B1-1_S1), G3-3_S10 of Family 3 and 6-1_S5 of Family 6. Oher four CP subjects that mobilize using wheelchair and/or stroller are both children in Family 2 (G2- 1_S8 and B2-2_S5), B4-T1_S1 of Family 4 and G5-2_S10 of Family 5. Among these wheelchair-CP subjects, only B4-T1_S1 and G5-2_S10 are able to propel the wheelchair on their own. There are also two CP children, B5-1_S7 (Family 5) and B4-T2_S2 (Family 4) who walk with abnormal gait and functional hands.

Table 3.4: The summary of other medical records of all CP affected subjects.

Case	Gender	Age (y/o)	Other Medical Records
Family 1			
B1-1_S1	♂	34	Spontaneous vagina delivery (SVD), no rolling over, motor age delayed at 2 to 3 months old, Spastic diplegia, Pest cavus foot, hallux valgus foot, hand fisting, unable to grasp object, contracture, scoliosis, clonus, hips dislocation (right anterior & left posterior), cannot talk, has eye contact and hearing.
B1-3_S4	♂	18	SVD, no rolling over, born to both parents aged >40 y/o, spastic quadriplegia, pes cavus foot, hallux valgus foot, bedridden, hips dislocation (right & left both posterior), cannot talk, has eye contact and hearing.
Family 2			
G2-1_S8	♀	3	Uneventful of full term (FT) SVD (2.75kg), Magnetic Resonance Imaging (MRI) brain, inborn errors of metabolism (IEM) are normal, spastic diplegia.
B2-2_S5	♂	2	SVD, unable to sit at the age of 13/12, spastic diplegia.
Family 3			
G3-3_S10	♀	18	FT SVD (2.85kg), unable to sit within first years of life, spastic quadriplegia, eczema, bedridden but able to move hands and slide body, has eye contact

(Note: ♀= Female; ♂= Male)

Table 3.4: Continued.

Case	Gender	Age (y/o)	Other Medical Records
Family 4			
B4-T1_S1	♂	7	Premature at 30 week (emergency caesarean; 1.32kg), ventilated for 1 day, spastic diplegia, mobilize using wheelchair, lower limb wasting, can communicate, attending normal school, mild specific learning.
B4-T2_S2	♂	7	Premature at 30 week (emergency caesarean; 0.67kg), ventilated for 40 days, spastic left hemiplegia, upper limb contractures, lower limb wasting, able to communicate, unable to recognize ABC, aggressive (behavioral issue).
Family 5			
B5-1_S7	♂	14	Born SVD, premature at 33 week, able to stand at 14 month old, normal MRI, spastic diplegia, walk with abnormal gait, lower limb wasting, attending normal school, hyperpigment and has rash on skin.
G5-2_S10	♀	10	Born SVD, spastic diplegia, cannot walk, able to sit, uses hands to move around, lower limb wasting, able to communicate.
Family 6			
B6-1_S5	♂	12	SVD (2.4kg), postdate 13 days, spastic quadriplegia, head lag, no rolling over, no head control, mild dystonia, microcephaly, cannot communicate, eczema, good eye contact.

(Note: ♀: Female, ♂: Male)

Table 3.4 show other medical records that are obtained from hospital medical records. There are eight CP subjects who were born via SVD, whilst the twins of Family 4 were born via emergency caesarean. Three were premature babies, they are the twins who were born at 30 weeks of gestational age and also B5_1_S7 (Family 5) who was born at 33 weeks. Overall, few insights could be obtained from these clinical data in Table 3.2, Table 3.3 and Table 3.4. Firstly, in this current study, all of these 10 CP subjects are classified under the same CP physiology which is spastic but with different topographic distribution such as six are diplegia (B1-1_S1, B2-1_S5, B4-T1_S1, B5-1_S5, G2-1_S8 and G5- 2_S10), three are quadriplegia (B1-3_S4, G3-3_S10 and B6-1_S5) and one is hemiplegia (B4-T2_S2) subjects.

This phenotype difference also happens among CP siblings from the same parental line. For example, Family 1 has two bedridden CP sons with spastic diplegia and spastic quadriplegia and Family 4 with a pair of monozygotic twins with spastic diplegia and spastic hemiplegia. Apart from that, as recorded in Table 3.2, it is obvious that individuals with the same CP phenotype do not necessarily share the same clinical features or comorbidities. They also show different gross motor function. Despite are classified under the same CP type, these CP individuals show different skill or ability. For example, in particular to diplegic CP as the most CP type that is present in this current study, diplegic CP subjects show different ability whereby some of them are able to utilize the hands such as to grasp an object and to propel the wheelchair but some are poor in hands functionality. Some diplegic CP subjects are also able to walk with abnormal gait and some are bedridden. All of the bedridden patients in this current study show poor head control. Thus, in order to mobilize them, it requires the ambulatory aid device that has truncal support function. In this current study, CP

subjects are ranging from 2 years old to 34 years old, this shows a possible normal life span for CP individuals. However, it still is depending on whether or not they are associated with any medical conditions as well. Lastly, besides not having facial dysmorphic features, all of these CP subjects who were born from uneventful pregnancies and deliveries also have no other known genetic disorders. This indicates that their CP conditions are not their secondary medical condition but a possible result of genetic underlying factors.

3.4 Concentration and Quality of Extracted DNA Sample

The DNA extracted from the blood samples of all subjects had good concentration of ≥ 60 ng/µl as recommended by Illumina technology for WES. The concentration ranged from 61.243 ng/µl to 213.475 ng/µl in 30 µl volume (four readings taken per sample). All extracted DNA samples achieved the optimum optical density (OD) of A260/280 within the range of 1.7– 2.2 and OD of A260/230 within the range of 2.0 - 2.2. This indicated that, all of the DNA samples were free from any contaminations such as protein, salt or residual alcohol. The list of DNA concentration ([DNA]) can be obtained in Appendix B. The DNA samples were also of good integrity as indicated by a thick band with no smear in each well (Figure 3.7). This Agarose gel electrophoresis image also double confirmed that no contamination occurred in extracted DNAs based on only a single band that was present per well.

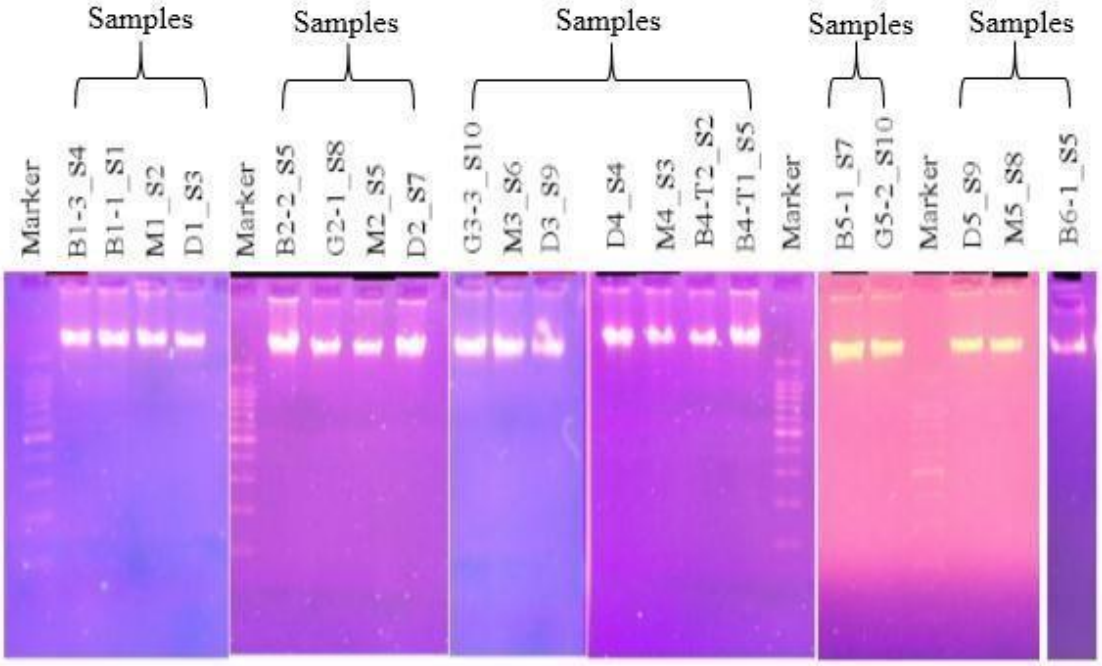


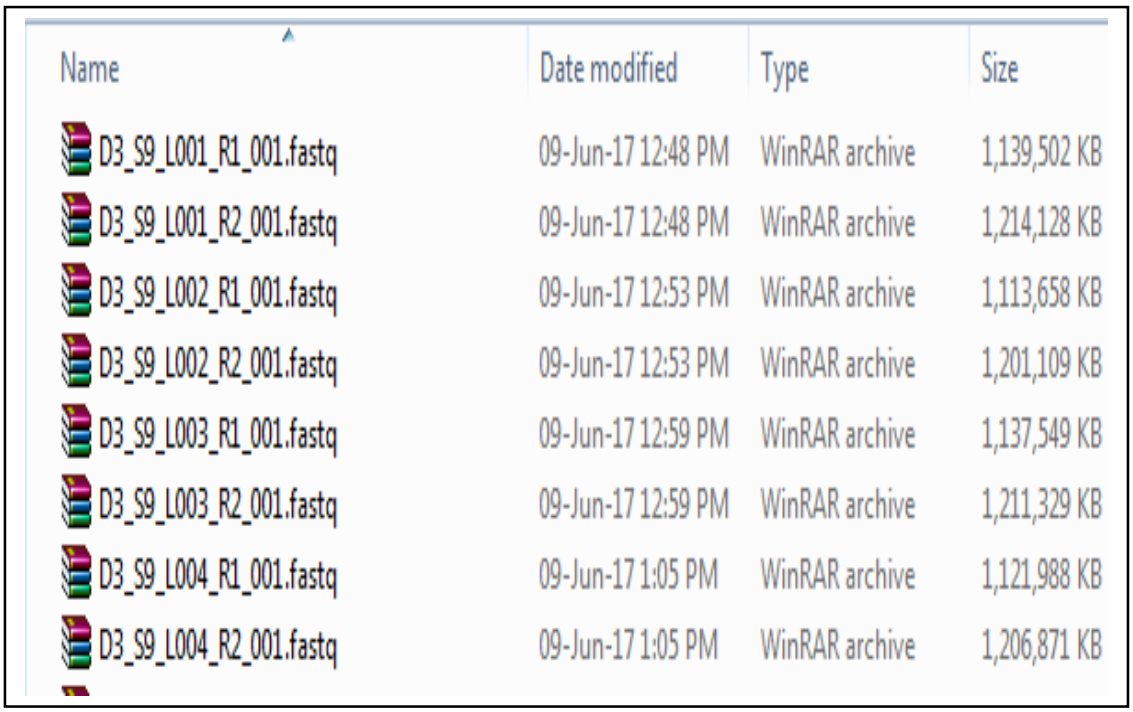
Figure 3.7: The gel image of all 20 gDNA samples. (Marker: mixture of 1 µl Ladder 100 bp, 1 µl loading dye and 1 µl TBE. Samples: mixture of 1µl loading dye and 2µl gDNAs).

3.5 DNA Libraries

The final Qubit values for libraries in pool 1 (PO1) was 19.0 ng/µl of 71.97 nM and libraries in pool 2 pool (PO2) was 19.6 ng/µl of 74.24 nM pool concentration. The average library size of PO1 was 347 bp and PO2 was 325 bp respectively. The actual insert sizes after deducting the size of adapters (120 bp) that flanked the insert (input gDNA) were 227 bp and 203 bp for PO1 (batch 1) and PO2 (batch 2) respectively.

3.6 Whole Exome Sequencing (WES) Data

Each sequencing batch contained 10 gDNA samples which yielded 96 Gb sequencing data per batch. This indicated that each gDNA data size was 9.6 Gb (96 Gb/10 gDNA samples per sequencing batch). Using the formula in section 2.10, the sequencing coverage was 75X, which means the whole exomes were sequenced for an average of 75 times. The reported sequencing read length was 2 x 150 bp. The Illumina Technology applied the Phred Quality Score of Q30 for sequencing reads. It is logarithmically related to the base calling error probabilities of an incorrect base call of 1 in 1000 times with 99.9% accuracy. According to Illumina, the study read length of 2 x 150bp had Q30 of minimum 75% bases is considered a standard score (*https://www.illumina.com/systems/sequencingplatforms/nextseq/specifications.html*) . For the current study, the sequencing raw data was therefore reliable. An example of sequencing output file generated from WES sequencing is shown in Figure 3.8.



Name	Date modified	Type	Size
D3_S9_L001_R1_001.fastq	09-Jun-17 12:48 PM	WinRAR archive	1,139,502 KB
D3_S9_L001_R2_001.fastq	09-Jun-17 12:48 PM	WinRAR archive	1,214,128 KB
D3_S9_L002_R1_001.fastq	09-Jun-17 12:53 PM	WinRAR archive	1,113,658 KB
D3_S9_L002_R2_001.fastq	09-Jun-17 12:53 PM	WinRAR archive	1,201,109 KB
D3_S9_L003_R1_001.fastq	09-Jun-17 12:59 PM	WinRAR archive	1,137,549 KB
D3_S9_L003_R2_001.fastq	09-Jun-17 12:59 PM	WinRAR archive	1,211,329 KB
D3_S9_L004_R1_001.fastq	09-Jun-17 1:05 PM	WinRAR archive	1,121,988 KB
D3_S9_L004_R2_001.fastq	09-Jun-17 1:05 PM	WinRAR archive	1,206,871 KB

Figure 3.8: The format for the sequencing reads output files

Figure 3.8 shows the example of WES raw data for one sample that was reported into eight fastq.gz files, a compressed zip file format. As explained in section 2.10, each sample of pair-end sequencing had four lanes and each lanes had two reported files making up to eight reported sequencing reads files in total. Each sequencing batch has 80 sequencing reads, thus there are 160 sequencing reads in total from two batches in this current study.

3.7 Bioinformatics

3.7.1 Assigned Sample Codes

As explained in section 2.11.1, for computational command purpose the sequencing reads for batch 1 and batch 2 were assigned as “array” and “array2” respectively. Then, a group of 10 samples in “array” was termed as {element} and another group of 10 samples in “array2” was termed as {element2} (Table 3.5). These “array”, “array2”, {element} and {element2} terms are standardized as computational languages/codes to be used throughout bioinformatics processes for this current study. This is to avoid any analysis error due to failure to distinguish sample read groups since this analysis was done simultaneously for both sample groups with multiple sequencing reads. Basically, Table 3.5 and Table 3.1 earlier show similar content but with different purposes. Table 3.1 describes 20 study samples that were grouped into two sequencing batches to produce raw reads data, whereas Table 3.5 describes the produced raw reads batches that were coded with computational languages for bioinformatics analysis to generate variants output.

Table 3.5: Assigned sample codes for further analyses using the bioinformatics command.

Sequencing reads	Assigned Bioinformatics Code	List of Element
Batch 1	"array" and {element}	B2-2_S5, G2-1_S8, B1-1_S1, M2_S6, G3-3_S10, D2_S7, D1_S3, B1-3_S4, D3_S9 and M1_S2.
Batch 2	"array 2" and {element 2}	D4_S4, M4_S3, B4-T2_S2, B4-T1_S1, D5_S9, G5-2_S10, B5-1_S7, M5_S8, M3_S6 and B6-1_S5.

3.7.2 Fastqc Inspection

Each sequencing raw data was inspected for any low quality sequences by accessing to fastqc.html report as described in section 2.11.1. Fastqc inspection was done on raw read files and received the problem alert (X in red circle symbol) which showed that the sample had biases of base percentage as in Figure 3.9. There were two biases of sequence (fluctuate peaks) located at the front and at the end of read. These biases occurred in all reads of the 20 samples due to tagmentation adapters by transposome activity of Nextera kit that was used during the gDNA library preparation procedure, whereby the adapters flanked each of gDNA as explained in subsection 2.9.3. Therefore, to remove these biases the trimming process is required. Trimming was done by scripting the computational codes to give a command of cutting off the biases base sequence from sequencing read sequences, leaving only the gDNA input sequences to be read and analyze further in the next steps of bioinformatics.

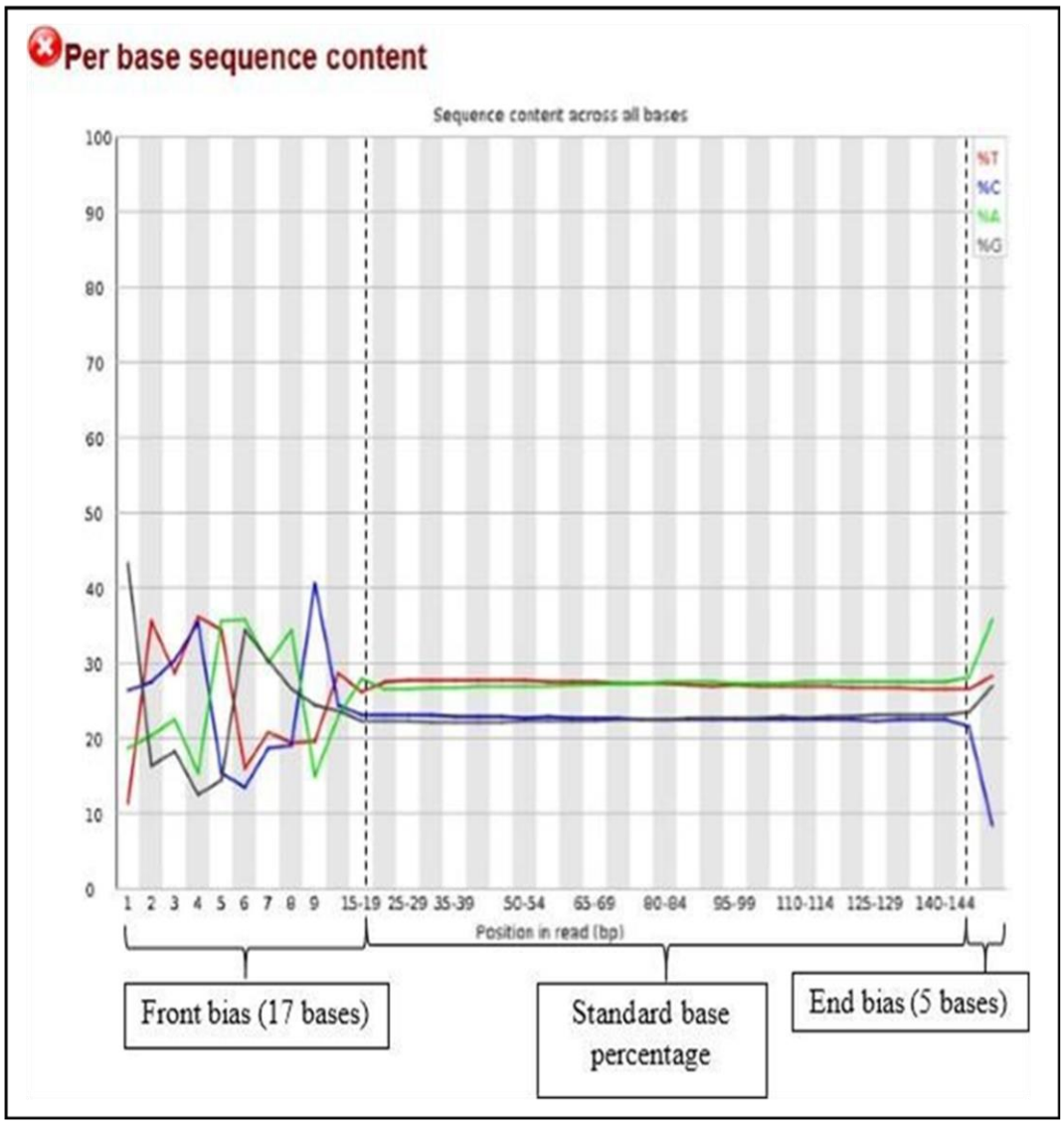


Figure 3.9: Example of fastqc.html report files of raw read sequence obtained from WES showing the base percentage in sequencing reads.

After trimming, we re-inspected the fastqc.html file again and received the good alert (√ in green circle), indicating that we managed to obtain the standard base percentage plot which was the parallel lines plot as in Figure 3.10. These line plots that run parallel to each other indicated that the proportion of each of the four bases remain relatively constant over the length of the read with %A=%T and %G=%C, which was acceptable for the next assembly process.

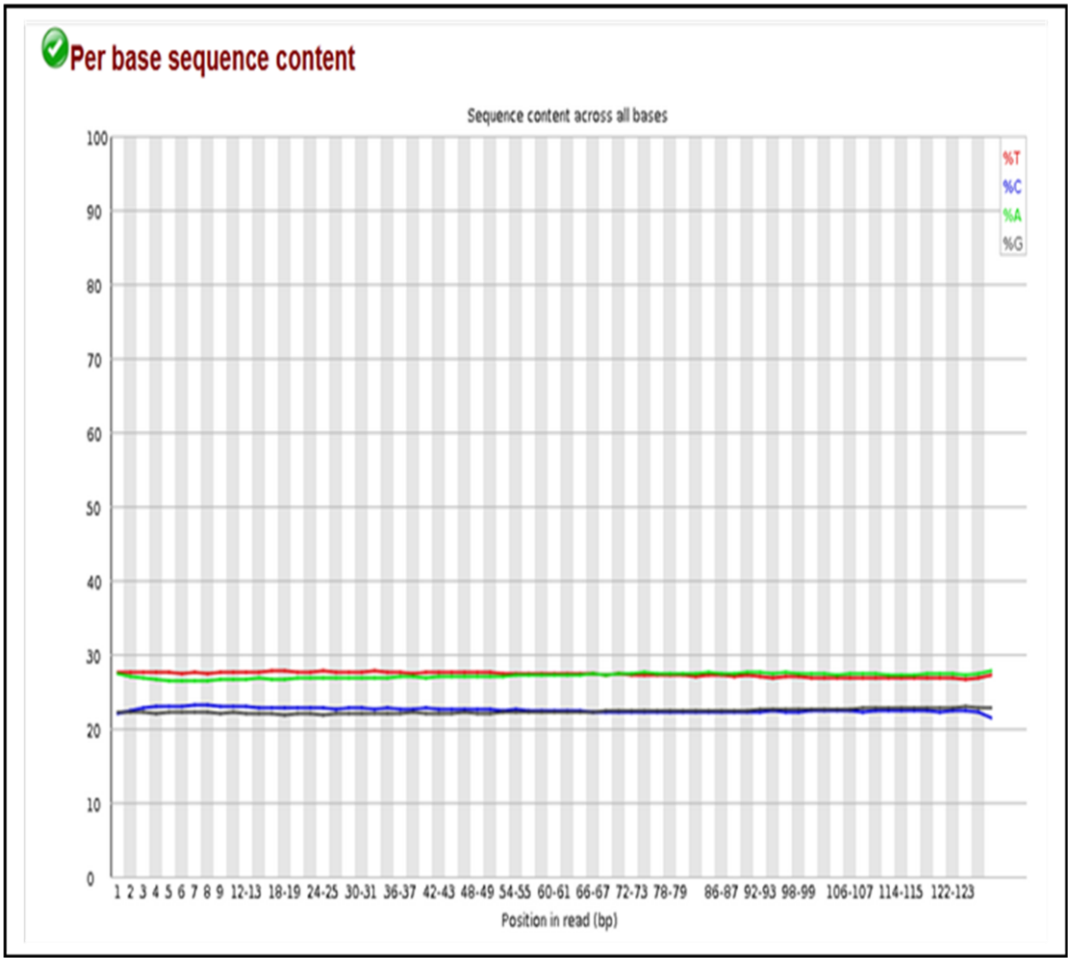


Figure 3.10: The optimum base percentage after trimming (clipping) the fluctuating peaks.

3.7.3 Pedigree File (PED)

The PED file was a generated tabular text file containing 1 row per person/sample and 6 mandatory columns that described the familial relationship between samples in alphanumerical and numerical codes respectively, to be used in the last process of bioinformatics, i.e genotype refinement (Table 3.6). The family ID in PED file was totally different from the family ID of the whole study (Figure 3.11), since the PED file has its own standard code.

Table 3.6: The generated PED file as supporting data for genotype refinement step.

Family ID	ID	Father	Mother	Gender	Phenotype
Fam01	B1-1_S1	D1_S3	M1_S2	1	2
Fam01	D1_S3	0	0	1	1
Fam01	M1_S2	0	0	2	1
Fam02	B1-1_S1	D1_S3	M1_S2	1	2
Fam02	D1_S3	0	0	1	1
Fam02	M1_S2	0	0	2	1
Fam03	G2-1_S8	D2_S7	M2_S6	2	2
Fam03	D2_S7	0	0	1	1
Fam03	M2_S6	0	0	2	1
Fam04	B2-2_S5	D2_S7	M2_S6	1	2
Fam04	D2_S7	0	0	1	1
Fam04	M2_S6	0	0	2	1
Fam05	G3-3_S10	D3_S9	M3_S6	2	2
Fam05	D3_S9	0	0	1	1
Fam05	M3_S6	0	0	2	1
Fam06	B4-T1_S1	D4_S4	M4_S3	1	2
Fam06	D4_S4	0	0	1	1
Fam06	M4_S3	0	0	2	1
Fam07	B4-T2_S2	D4_S4	M4_S3	1	2
Fam07	D4_S4	0	0	1	1
Fam07	M4_S3	0	0	2	1
Fam08	B5-1_S7	D5_S9	M5_S8	1	2
Fam08	D5_S9	0	0	1	1
Fam08	M5_S8	0	0	2	1
Fam09	G5-2_S10	D5_S9	M5_S8	2	2
Fam09	D5_S9	0	0	1	1
Fam09	M5_S8	0	0	2	1
Fam10	B6-1_S5	0	0	1	2

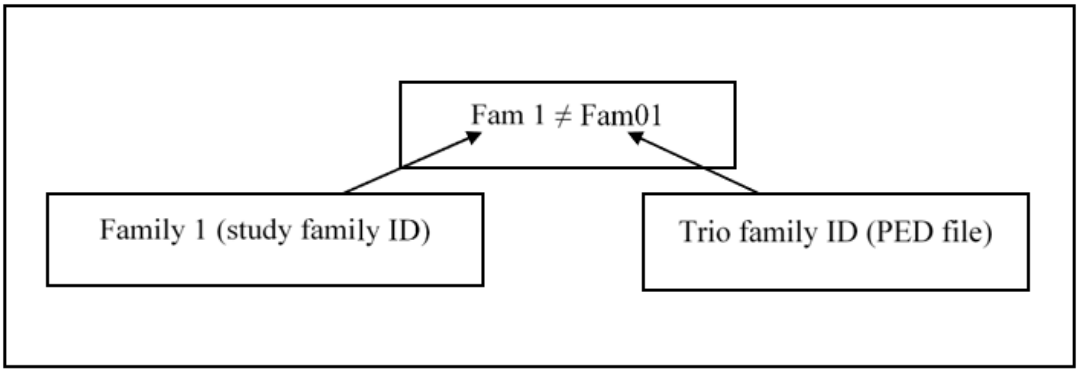


Figure 3.11: Difference in alphanumerical code of family ID used for PED file and the whole study respectively. (Details are in Table 3.7).

In PED file in Table 3.6, each family was designated in tabulated trio relationship consisting of mother, father and one child only. Thus, if a family has two CP subjects, this family was then stratified into two groups of trio separately with two different "Family ID". By referring to Table 3.6, there were 4 families (Family 1, Family 2, Family 4 and Family 5) that have two CP subjects thus resulting in two trios of PED Family ID respectively, as described in Table 3.7.

Table 3.7: The PED file interpretation.

Column	Variables	Description
1	Family ID	By referring to alphanumerical code in Figure 3.11: Fam01 and Fam02= two trio groups ID for Family 1 Fam03 and Fam04= two trio groups for Family 2 Fam05= a trio group for Family 3 Fam06 and Fam07= two trio groups for Family 4 Fam08 and Fam09= two trio groups for Family 5 Fam10= a trio group for Family 6
2	ID	Sample/subject ID (each row for each analyzed subject of affected and unaffected subjects).
3	Father	Orientated to column 2: If the subject (column 2) has the analyzed paternal data, thus the father ID shall be inserted. 0= absent of paternal data
4	Mother	Orientated to column 2: If the subject (column 2) has the analyzed maternal data, thus the mother ID shall be inserted. 0= absent of parents' data
5	Gender	1= Male, 2= Female
6	Phenotype	1= Unaffected, 2=Affected

3.7.4 Annovar Annotated Variants

After completing the bioinformatics, a big data of 172428 (~172k) annotated variants which consists of SNPs and mutations was obtained after being filtered against multiples databases. These annotated variants were reported in excel files with 111 columns of parameters, variants scores, databases that were applied during the bioinformatics analyses.

3.8 Interpreted Data

As explained in section 2.12, this annotated ~172k variants are not the main study output. Since this current study produced a big variants data within limited time available, thus at this preliminary study stage we decided to focus the interpreation on the mutation only. In order to obtain only mutations that are related to CP in this study, further data interpretation was done for this big data by multiple screening/selection processes according to few criteria such as variants located on exonic regions, variants that are present in CP subjects only, variants that are present in all subjects including unaffected parents, null variant types (frameshift, stopgain and stoploss) and non-synonymous variants based on pathogenicity scores in variant effect prediction tool databases (Figure 3.12).

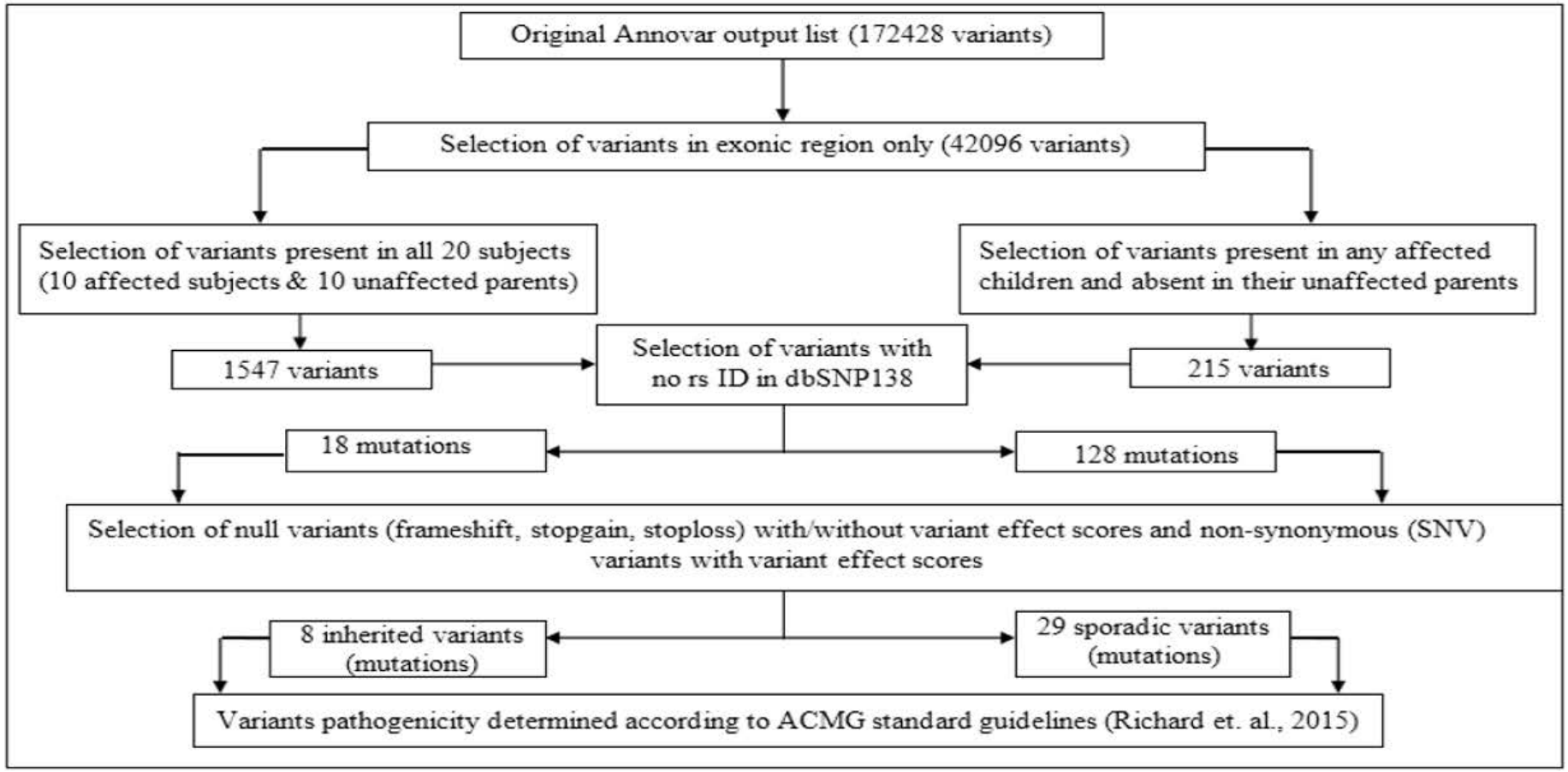


Figure 3.12: Workflow of variants data analysis. (ACMG: American College Medical Genetics variant pathogenicity standard guidelines, dbSNP138: SNPs database, rs ID: indicates the variant is SNP not mutation and variant effect scores: Polyphen2_HDIV, Polyphen2_HVAR, SIFT, CADD and GERP++_RS).

As stated earlier (section 2.11.9), few widely used variant effect prediction tools such as SIFT, Polyphen2_HDIV, Polyphen2_HVAR, CADD_phred and GERP++_RS were applied during Annovar annotation in this current study. The variant effect prediction tool interpretations are available in Table 3.8. The main output generated from this analysis consists of eight inherited variants (Table 3.9) and 29 sporadic variants (Table 3.10), which are defined as mutations since all of them have no rs ID number indicating that they are not classified as SNPs. Both inherited and sporadic variants pathogenicity were determined according to multiple criteria available in worldwide standard variants pathogenicity guideline established by ACMG (Richard et. al., 2015).

Table 3.8: Variant effect prediction tools interpretation.

Prediction Tool	Function	Score Range	Deleterious Score
SIFT	To predict affected protein function due to amino acid substitution in non-SNV and missense mutation.	1 to 0	< 0.05
Polyphen2_HDIV	To predict human damaging mutations by assuming differences between human proteins and their closely related mammalian homologs as non-damaging	0 to 1	> 0.05
Polyphen2_HVAR	To predict human disease-causing mutations by assuming common human nonsynonymous SNPs as non-damaging	0 to 1	> 0.05
CADD_phred	To predict deleteriousness of single nucleotide variants and indels by integrating multiple annotations including variant conservation and functional information.	0 to 35	> 15
GERP++_RS	To predict the locus conservation due to any amino acid changing.	-12.0 to 6.17	>0.047

Multiple prediction tools are functioning to predict a variant deleteriousness such as whether or not a particular variant have a damaging impact due to protein changing. According to Table 3.8, the SIFT score range is 0 to 1 whereby a variant that has score < 0.05 is predicted to be deleterious meanwhile a variant with score of > 0.05 is predicted to be tolerated or benign variant. The Polyphen2 (HDIV & HVAR) ranging from 1 to 0, whereby a variant with Polyphen 2 score of > 0.05 is considered deleterious or also termed as "damaging" in Polyphen2 program, whereas a variant with < 0.05 score in Polyphen is considered benign. CADD_phred score range is 0 to 35, a variant with > 15 score is considered deleterious in protein function. Meanwhile, the GERP++_RS score range is -12.3 to 6.17, whereby a variant with > 0.047 is considered deleterious or also specifically termed as "less conserved" in this prediction tool. These variant effects determine the molecular functionality score of each variant of all 37 variants discovered in this current study as tabulated in Table 3.9 and Table 3.10. Both inherited (section 3.8.1) and sporadic (section 3.8.2) variant pathogenicities were determined according to multiple criteria available in worldwide standard variants pathogenicity guideline established by ACMG (Richard et. al., 2015).

3.8.1 Inherited Variants

Table 3.9: List of 8 inherited variants with some details from Annovar output and ACMG classification.

No	Mutated Gene	Position (hg19)	Mutation Type	Mutation	SIFT	HDIV	HVAR	CADD_phred	GERP++_RS	ACMG
1	*CTBP2*	chr10_126683179_126683210_CAAGTAGGGGTCATAAAATATGACGCTGAATC_-	Frameshift Deletion	exon5:c.2228_2259del:p.G73fs (deletion of 32 nucleotide)	.	.	.	.	.	Likely Pathogenic
2	*CTDSP2*	chr12_58220811_58220811_G_T	Non SNV	exon4:c.C322A:p.L108I	0	1	0.999	21.4	0.817	Uncertain Significance
3	*CTDSP2*	chr12_58220831_58220831_C_G	Non SNV	exon4:c.G302C:p.R101T	0	0.68	0.473	22.7	4.92	Uncertain Significance
4	*ANKRD 36*	chr2_97818261_97818262_GC_-	Frameshift Deletion	exon14:c.1183_1184del:p.A395fs	.	.	.	.	.	Likely Pathogenic

(Notes: Genes of CTBP2 (C-Terminal Binding Protein 2), CTDSP2 (Carboxy-Terminal Domain Small Phosphatase 2) and ANKRD36 (Ankyrin Repeat Domain 36); Position= Mutation position in chromosome (chr) in hg19 database; ins= insertion; del= deletion; SIFT (< 0.05) = Deleterious, SIFT (> 0.05) = Benign; Polyphen2_ (HDIV and HVAR) (> 0.05) = Deleterious; CADD_phred (> 15) = Deleterious; GERP++_RS (> 0.047) = Deleterious; "." = Indicates that the variant does not have score for variant effect in prediction tool programs; and ACMG = ACMG classification).

Table 3.9: Continued.

No	Mutated Gene	Position (hg19)	Mutation Type	Mutation	SIFT	HDIV	HVAR	CADD _phred	GERP++ _RS	ACMG
5	*ANKRD 36*	chr2_97818264_97818264_-_TT	Frameshift Insertion	exon14:c.1186_1187insTT:p.V396fs	.	.	.	.	.	Likely Pathogenic
6	*KRTAP 19-6*	chr21_31913982_31913982_G_-	Frameshift Deletion	exon1:c.171delC:p.F57fs	.	.	.	.	.	Likely Pathogenic
7	*FAM104 B*	chrX_55172685_55172685_-_T	Frameshift Insertion	exon3:c.182_183insA:p.S61fs	.	.	.	.	.	Likely Pathogenic
8	*FAM104 B*	chrX_55172689_55172689_G_-	Frameshift Deletion	exon3:c.179delC:p.A60fs	.	.	.	.	.	Likely Pathogenic

(Notes: Genes of *ANKRD36* (Ankyrin Repeat Domain 36), *KRTAP19-6* (Keratin Associated Protein 19-6) and *FAM104B* (Family With Sequence Similarity 104 Member B); Position= Mutation position in chromosome (chr) in hg19 database; ins= insertion; del= deletion; SIFT (< 0.05) = Deleterious, SIFT (> 0.05) = Benign; Polyphen (HDIV and HVAR) (> 0.05) = Deleterious; CADD_phred (> 15) = Deleterious; GERP++_RS (> 0.047) = Deleterious; "." = Indicates that the variant does not have score for variant effect in prediction tool programs; and ACMG = ACMG classification).

Table 3.9 shows eight mutations that are categorized as inherited variants since they are present not only in all CP affected children but also in all unaffected parents. They comprise of two non-SNV variants, four frameshift deletion and two frameshift insertion mutations of total five genes (Table 3.9). In the Annovar output, six out of eight variants which are c.2228_2259del:p.G743fs, c.1183_1184del:p.A395fs, c.1186_1187insTT:p.V396fs, c.171delC:p.F57fs, c.182_183insA:p.S61fs and c.179delC:p.A60fs are denoted as "." in molecular functionality prediction tool columns of SIFT, Polyphen 2_HDIV, Polyphen2_HVAR, CADD_phred and GERP++_RS indicating that these variants have no score available in these molecular functionality prediction tool database programs. However, according to ACMG classification these six variants are all likely pathogenic. Only two variants have the scores in prediction tool programs which both are non-SNV of CTDPS2 gene. The mutations are exon4:c.C322A:p.L108I and exon4:c.G302C:p.R101T.

In particular to these two mutations of CTDSP2 gene, exon4:c.C322A:p.L108I mutation has 0 score for SIFT, 1 score for Polyphen2_HDIV, 0.999 score for polyphen2_HVAR whereby these three scores show that this mutation is deleterious. It also shows a pathogenic score (21.4) for CADD_phred and less conserved score (0.817) for GERP++_RS. Another CTDSP2 mutation, exon4:c.G302C:p.R101T shows the same deleterious prediction but with slightly different score values such as 0 for SIFT, 0.68 for Polyphen2_HDIV, 0.473 for polyphen2_HVAR, pathogenic score (22.7) for CADD_phred and less conserved score (4.92) for GERP++. These prediction tool scores show a possibility of protein function changing due to these variants. However, the ACMG classified these two variants as of uncertain significance.

3.8.2 Sporadic Variants

From the data interpretation, we discovered 29 sporadic variants of total 25 genes which are as following, Mucin 16, Cell Surface Associated (MUC16) (two variants; n=2), Human Leukocyte Antigen-Major Histocompatibility Complex Class II Beta Chain 1 (HLA- DRB1) (four variants; n=4), whereas one variant (n=1) for each gene of ANKRD36, Family with Sequence Similarity 163 Member A (FAM163A), Tetratricopeptide repeat domain protein 13 (TTC13), Bestrophin 3 (BEST3), Glyoxylate Reductase 1 Homolog (GLYR1), Fragile X Mental Retardation 1 Autosomal Homolog 2 (FXR2), Dynein Axonemal Heavy Chain 17 (DNAH17), Microtubule Associated Protein RP/EB Family Member 3 Gene (MAPRE3), Collagen Type VI Alpha 6 Chain Gene (COL6A6), Asteroid Homolog 1 (ASTE1), Collagen Type VII Alpha 1 Chain Gene (COL7A1), Human Leukocyte Antigen- Major Histocompatibility Complex Class II Beta Chain 5 (HLA-DRB5), Potassium Voltage-Gated Channel Subfamily Q Member 3 Gene (KCNQ3), Cysteine And Histidine Rich 1 Gene (CYHR1), Potassium Two Pore Domain Channel Subfamily K Member 18 (KCNK18), Centrosomal Protein 164 (CEP164), Adhesion G Protein-Coupled Receptor (GPR97), Cyclin Dependent Kinase Like 2 (CDKL2), RNA Binding Motif Protein X- Linked (RBMX), Lysosomal Trafficking Regulator (LYST), Anoctamin 5 (ANO5), Dachsous Cadherin-Related 2 (DCHS2) and Delta Like Canonical Notch Ligand 1 (DLL1) (Table 3.10).

Table 3.10: The list of 29 Likely Pathogenic sporadic variants with their few extracted details from Annovar output.

No	Mutated gene	Position (hg19)	Mutation type	Mutation	SIFT	Polyphen2_HDIV	Polyphen2_HVAR	CADD_phred	GERP++_RS
1	*MUC16*	chr19_8999498_8999501_GCTT_-	Frameshift Deletion	exon56:c.40674_40677del: p.K13558fs	.	.	.	.	.
2	*MUC16*	chr19_8999502_8999502_-_CCGA	Frameshift Insertion	exon56:c.40672_40673insT CGG:p.K13558fs	.	.	.	.	.
3	*KCNK18*	chr10_118969015_118969015_-_T	Frameshift Insertion	exon3:c.361dupT:p.G120fs	.	.	.	.	.
4	*MAPRE3*	chr2_27248517_27248517_C_-	Frameshift Deletion	exon5:c.536delC:p.A179fs	.	.	.	.	.
5	*FAM163A*	chr1_179783220_179783220_C-	Frameshift Deletion	exon5:c.400delC:p.P134fs	.	.	.	.	.
6	*GLYR1*	chr16_4862229_4862229_C_-	Frameshift Deletion	exon13:c.1140delG:p.G380fs	.	.	.	.	.
7	*KCNQ3*	chr8_133150233_133150233_T_-	Frameshift Deletion	exon12:c.1599delA:p.K533 fs	.	.	.	.	.
8	*GPR97*	chr16_57717983_57717983_-_G	Frameshift Insertion	exon9:c.1022dupG:p.341f s	.	.	.	.	.

(Note: Position= Mutation position in chromosome (chr) in hg19 database; ins= insertion; del= deletion; dup= duplication; SIFT (< 0.05) = Deleterious, SIFT (> 0.05) = Benign; Polyphen (HDIV and HVAR) (> 0.05) = Deleterious; CADD_phred (> 15) = Deleterious; GERP++_RS (> 0.047) = Deleterious; and "." = Indicates that the variant does not have score for variant effect in prediction tool programs).

Table 3.10: Continued.

No	Mutated gene	Position (hg19)	Mutation type	Mutation	SIFT	Polyphen2_HDIV	Polyphen2_HVAR	CADD_phred	GERP++_RS
9	*DNAH17*	chr17_76506467_76506467_T_-	Frameshift Deletion	exon27:c.4235delA:p.K1412fs	.	.	.	.	.
10	*ANO5*	chr11_22296174_22296174_C_A	Stopgain	exon20:c.C2295A:p.Y765X	1	.	.	41	2.37
11	*FXR2*	chr17_7517826_7517826_C_-	Frameshift Deletion	exon1:c.25delG:p.D9f s	.	.	.	.	.
12	*COL7A1*	chr3_48613856_48613856_C_-	Frameshift Deletion	exon69:c.5749delG:p.E1917fs	.	.	.	.	.
13	*CYHR1*	chr8_145689659_145689659_C-	Frameshift Deletion	exon3:c.430delG:p.A144fs	.	.	.	.	.
14	*HLA- DRB1*	chr6_32552137_32552137_-_A	Frameshift Insertion	exon2:c.118_119insT:p.P40fs	.	.	.	.	.
15	*HLA- DRB1*	chr6_32552141_32552141_G_-	Frameshift Deletion	exon2:c.115delC:p.Q39fs	.	.	.	.	.

(Note: Position= Mutation position in chromosome (chr) in hg19 database; ins= insertion; del= deletion; dup= duplication; SIFT (< 0.05) = Deleterious, SIFT (> 0.05) = Benign; Polyphen (HDIV and HVAR) (> 0.05) = Deleterious; CADD_phred (> 15) = Deleterious; GERP++_RS (> 0.047) = Deleterious; and "." = Indicates that the variant does not have score for variant effect in prediction tool programs).

Table 3.10: Continued.

No	Mutated gene	Position (hg19)	Mutation type	Mutation	SIFT	Polyphen2_HDIV	Polyphen2_HVAR	CADD_phred	GERP++_RS
23	*HLA- DRB5*	chr6_32489880_32489881_TG-	Frameshift Deletion	exon2:c.171_172del:p.H57f s	.	.	.	.	.
24	*ASTE1*	chr3_130733047_130733047_T_-	Frameshift Deletion	exon7:c.1969delA:p.R657f s	.	.	.	.	.
25	*TTC13*	chr1_231064801_231064801_G_-	Frameshift Deletion	exon12:c.1339delC:p.L447fs	.	.	.	.	.
26	*LYST*	chr1_235883963_235883963_G_T	Stopgain	exon40:c.C9558A:p.Y3186X	1	.	.	51	3.74
27	*ANKRD36*	chr2_97830031_97830031_C_ T	Stopgain	exon19:c.C1441T:p.Q481X	1	.	.	26.2	-0.435
28	*DLL1*	chr6_170597334_170597334_G_T	Stopgain	exon4:c.C663A:p.C221X	1	.	.	40	4.25
29	*CEP164*	chr11_117222648_117222648_-_A	Frameshift Insertion	exon5:c.337dupA:p.K113fs	.	.	.	.	.

(Note: Position= Mutation position in chromosome (chr) in hg19 database; ins= insertion; del= deletion; dup= duplication; SIFT (< 0.05) = Deleterious, SIFT (> 0.05) = Benign; Polyphen (HDIV and HVAR) (> 0.05) = Deleterious; CADD_phred (> 15) = Deleterious; GERP++_RS (> 0.047) = Deleterious; and "." = Indicates that the variant does not have score for variant effect in prediction tool programs).

Table 3.10: Continued.

No	Mutated gene	Position (hg19)	Mutation type	Mutation	SIFT	Polyphen2 _HDIV	Polyphen2 _HVAR	CADD _phred	GERP++ _RS
23	*HLA- DRB5*	chr6_32489880_32 489881_TG-	Frameshift Deletion	exon2:c.171_172del:p .H57f s	.	.	.	.	.
24	*ASTE1*	chr3_130733047_1 30733047_T_-	Frameshift Deletion	exon7:c.1969delA:p.R 657f s	.	.	.	.	.
25	*TTC13*	chr1_231064801_2 31064801_G_-	Frameshift Deletion	exon12:c.1339delC:p. L447fs	.	.	.	.	.
26	*LYST*	chr1_235883963_2 35883963_G_T	Stopgain	exon40:c.C9558A:p.Y 3186X	1	.	.	51	3.74
27	*ANKRD36*	chr2_97830031_97 830031_C_ T	Stopgain	exon19:c.C1441T:p.Q 481X	1	.	.	26.2	-0.435
28	*DLL1*	chr6_170597334_ 170597334_G_T	Stopgain	exon4:c.C663A:p.C22 1X	1	.	.	40	4.25
29	*CEP164*	chr11_117222648_ 117222648_-_A	Frameshift Insertion	exon5:c.337dupA:p.K 113fs	.	.	.	.	.

(Note: Position= Mutation position in chromosome (chr) in hg19 database; ins= insertion; del= deletion; dup= duplication; SIFT (< 0.05) = Deleterious, SIFT (> 0.05) = Benign; Polyphen (HDIV and HVAR) (> 0.05) = Deleterious; CADD_phred (> 15) = Deleterious; GERP++_RS (> 0.047) = Deleterious; and "." = Indicates that the variant does not have score for variant effect in prediction tool programs).

Table 3.10 shows only a total of five stopgain mutations that have the molecular functionality scores in the prediction tool database programs. The first stopgain mutation is exon20:c.C2295A:p.Y765X (ANO5 gene) with benign/tolerated SIFT score (1), highly pathogenic (41) CADD_phred score, less conserved (2.37) GERPT++_RS score with no scores (.) available in Polyphen2 databases. Next is stopgain mutation of exon9:c.C1796A:p.S599X (DCHS2 gene) with tolerated SIFT score (0.84), highly pathogenic (41) CADD_phred score, less conserved (4.35) GERPT++_RS score with no scores (.) available in Polyphen2 databases. Thirdly, exon40:c.C9558A:p.Y3186X (LYST gene) with SIFT score 1 (tolerated), highly pathogenic (51) CADD_phred score, less conserved (3.74) GERPT++_RS score with no polyphen2 scores.

This if followed by exon19:c.C1441T:p.Q481X stopgain mutation of ANKRD36 with tolerated SIFT score (1), pathogenic (26) CADD_phred score, very less conserved (-0.435) GERPT++_RS score with no Polyphen2 database scores. The fifth stopgain is exon4:c.C663A:p.C221X in DLL1 gene with SIFT score 1 (tolerated), highly pathogenic (40) CADD_phred score, less-conserved (4.25) GERPT++_RS score with no polyphen2 scores. These scores of proteins changing prediction of mentioned stopgain mutations sometime show a contra prediction score pattern, for example predicted to be tolerated according to SIFT but predicted to be highly pathogenic according to other prediction tools. However, all of these sporadic variants including the mutations with no prediction (.) score in databases were classified as likely pathogenic according to ACMG standard guideline (Richard et. al., 2015). Other additional prediction tools such as Mutation Taster, Mutation Assessor and Functional Analysis through Hidden Markov Models (FATHMM)

were also applied yet these frameshift variants for both inherited and sporadic categories are still denoted as "." indicating that they have no deleteriousness prediction scores for these aditional prediction tool programs. Therefore, since this is a preliminary study, other studies with independent analysis on each of the platform are recommended for both inherited and sporadic variants that have no score (.) for the molecular functionality prediction tool program (platform).

3.8.2(a) Sporadic Variants As De Novo Mutations in Family

After comparing the sporadic variants in each family, these 29 sporadic variants (Figure 3.12) are only present in the CP children and are totally absent in unaffected parental samples. They are then considered as de novo mutations (Oliveira et. al., 2001; Richard et. al., 2015). However, there is no de novo mutation in the CP child of Family 3. Besides that, the variants in CP child (B6-1_S5) of Family 6 could not be determined as de novo mutations through parent-child variants comparison due to no parental variant data, since the parental samples did not pass the DNA quality test for WES as explained earlier. Overall, only eight CP subjects have de novo mutations. Unlike the inherited variants, these mutations are not shared among the CP subjects nor among the same CP plegia subtype (Table 3.11). For example, the mutated COL7A1 gene of exon69:c.5749delG:p.E1917fs mutation is only present (/) in B2-2_S5 (diplegia) and B4- T1_S1 (diplegia), while the other four diplegic and other plegic CP subjects do not have

(NA) this mutation. Even for the monozygotic twins in Family 4 with different plegia type, they do not share the same sporadic variants between themselves. The profile of 29 de novo mutations is available in the following Table 3.11.

Table 3.11: Profile of 29 de novo mutations in CP subjects in the siblings.

Variant (de novo)	Family 1		Family 2		Family 3	Family 4		Family 5	
	B1-1_S1 (dip)	B1-3_S4 (quad)	G2-1_S8 (dip)	B2-2_S5 (dip)	G3-3_S10 (quad)	B4-T1_S1 (dip)	B4-T2_S2 (hemi)	B5-1_S7 (dip)	G5-2_S10 (dip)
MUC16 exon56:c.40674_40677del:p.K1355 8fs	NA	NA	/	NA	NA	NA	NA	/	/
MUC16 exon56:c.40672_40673insTCGG:p. K13558fs	NA	NA	/	NA	NA	NA	NA	/	/
KCNK18 exon3:c.361dupT:p.G120fs	NA	NA	NA	NA	NA	NA	NA	/	/
MAPRE3 exon5:c.536delC:p.A179fs	/	NA	NA	NA	NA	/	NA	NA	/
FAM163A exon5:c.400delC:p.P134fs	NA	NA	NA	NA	NA	NA	/	NA	/
GLYR1 exon13:c.1140delG:p.G380fs	NA	NA	NA	NA	NA	/	NA	NA	/
KCNQ3 exon12:c.1599delA:p.K533fs	NA	NA	NA	NA	NA	/	NA	NA	/
GPR97 exon9:c.1022dupG:p.R341fs	NA	NA	NA	/	NA	NA	NA	/	NA
DNAH17 exon27:c.4235delA:p.K1412fs	NA	NA	NA	NA	NA	NA	NA	/	NA
ANO5 exon20:c.C2295A:p.Y765X	/	NA	NA	NA	NA	NA	NA	NA	/
FXR2 exon1:c.25delG:p.D9fs	NA	NA	NA	NA	NA	NA	NA	/	NA

(CP subtypes are dip= diplegia, quad= quadriplegia, hemi= hemiplegia; / = De novo variant present; and NA= De novo variant absent).

Table 3.11: Continued.

Variant (de novo)	Family 1		Family 2		Family 3	Family 4		Family 5	
	B1-1_S1 (dip)	B1-3_S4 (quad)	G2-1_S8 (dip)	B2-2_S5 (dip)	G3-3_S10 (quad)	B4-T1_S1 (dip)	B4-T2_S2 (hemi)	B5-1_S7 (dip)	G5-2_S10 (dip)
COL7A1 exon69:c.5749delG:p.E1917fs	NA	NA	NA	/	NA	/	NA	NA	NA
CYHR1 exon3:c.430delG:p.A144fs	NA	NA	/	NA	NA	/	NA	NA	NA
HLA-DRB1 exon2:c.118_119insT:p.P40fs	NA	NA	/	NA	NA	NA	/	NA	NA
HLA-DRB1 exon2:c.115delC:p.Q39fs	NA	NA	/	NA	NA	NA	/	NA	NA
HLA-DRB1 exon2:c.111delG:p.L37fs	NA	NA	/	NA	NA	NA	/	NA	NA
HLA-DRB1 exon2:c.109delC:p.L37fs	NA	NA	/	NA	NA	NA	/	NA	NA
CDKL2 exon3:c.222dupA:p.R75fs	/	NA	NA	NA	NA	NA	/	NA	NA
DCHS2 exon9:c.C1796A:p.S599X	/	NA	NA	NA	NA	/	NA	NA	NA
RBMX exon9:c.905_906insCC:p.P302fs	NA	/	NA	NA	NA	/	NA	NA	NA
BEST3 exon7:c.773delT:p.L258fs	NA	/	NA	NA	NA	NA	/	NA	NA
COL6A6 exon19:c.4620delC:p.G1540fs	NA	/	NA	NA	NA	NA	/	NA	NA

(CP subtypes are dip= diplegia, quad= quadriplegia, hemi= hemiplegia; / = De novo variant present; and NA= De novo variant absent).

Table 3.11: Continued.

Variant (de novo)	Family 1		Family 2		Family 3	Family 4		Family 5	
	B1-1_S1 (dip)	B1-3_S4 (quad)	G2-1_S8 (dip)	B2-2_S5 (dip)	G3-3_S10 (quad)	B4-T1_S1 (dip)	B4-T2_S2 (hemi)	B5-1_S7 (dip)	G5-2_S10 (dip)
HLA-DRB5 exon2:c.171_172del:p.H57fs	NA	/	NA	NA	NA	/	NA	NA	NA
ASTE1 exon7:c.1969delA:p.R657fs	NA	/	/	NA	NA	NA	NA	NA	NA
TTC13 exon12:c.1339delC:p.L447fs	NA	/	/	NA	NA	NA	NA	NA	NA
LYST exon40:c.C9558A:p.Y3186X	NA	/	NA	/	NA	NA	NA	NA	NA
ANKRD36 exon19:c.C1441T:p.Q481X	NA	NA	NA	/	NA	NA	NA	NA	NA
DLL1 :c.C663A:p.C221X	/	NA	NA	NA	NA	NA	NA	NA	NA
CEP164 exon5:c.337dupA:p.K113fs	NA	/	NA	NA	NA	NA	NA	NA	NA

(CP subtypes are dip= diplegia, quad= quadriplegia, hemi= hemiplegia; / = De novo variant present; and NA= De novo variant absent).

CHAPTER FOUR

DISCUSSION

4.1 General Discussion

NGS specifically WES technology is widely used as a diagnostic tool for diseases and disorders, varying from single gene disorders to the more complex genetic disorder of cancers and traits (Majewski et. al., 2011). Bioinformatics analysis helps in discovering the variants in WES data, and in our study, we applied the GATK pipeline, a standard bioinformatics pipeline for the germline variants case. The multi-filtering against the public databases such as dbSNP, 1000 Genome Project and HapMap helped to produce the initial raw data of ~172k annotated variants. For the data interpretation and analyses purpose, this Annovar output was then displayed in an excel file format.

Our Bioinformatics analysis yielded a number of variants found in various regions as the following: exonic (~73k), UTR (~84k), Splice site (93), Upstream (113), Downstream (109), intronic (579), and other known regions (~13k). From this premature raw bioinformatics data, further interpretation was done according to the criteria of variant selection as described in the Figure 3.12. Genetics of CP has not yet been established in Malaysia, thus we decided to prioritize our data filtering on the exonic region only. Exons are DNA regions where multiple specific proteins are encoded. Besides that, there are two types of variants we are currently focusing on. Firstly, the inherited variant type, a situation where a particular variant is observed to be shared by all subjects including affected children and unaffected parents per family. Secondly, the sporadic variant type, a situation where variants are scattered among the CP subjects only. These sporadic variants are present only in the affected children for

the first time in a family but absent in the normal parents, thus they are considered as de novo mutations. This study revealed 37 novel variants from 29 genes that are present in our subjects including both CP children and unaffected parents, whereby each subject is shown to be having multiple mutations. Only the mutated *ANKRD36* has both types of variant which are sporadic (n=1) and inherited (n=2) variants. Meanwhile, the other 28 genes have only one variant (n=1) type whether it is a sporadic or inherited variant type accordingly. There are eight inherited variants from five genes which are *ANKRD36* (n=2), *CTBP2* (n=1), *CTDSP2* (n=2), *FAM104B* (n=2) and *KRTAP19-6* (n=1). All of these CP subjects together with all the unaffected parent subjects are sharing all the same eight inherited variants regardless of their different CP phenotypes. For the record, six of inherited variants are classified as Likely Pathogenic variants whereas the other two inherited variants are classified as Uncertain Significance according to the ACMG classification criteria (Richard et. al., 2015). The details for this inherited variant classification criteria will be highlighted in the subsequent paragraphs per gene and variant respectively.

Meanwhile, there are 29 de novo (sporadic) mutations from 25 genes (*ANKRD36* (n=1), *ANO5* (n=1), *ASTE1* (n=1), *BEST3* (n=1), *CDKL2* (n=1), *CEP164* (n=1), *COL6A6* (n=1), *COL7A1* (n=1), *CYHR1* (n=1), *DCHS2* (n=1), *DLL1* (n=1), *DNAH17* (n=1), *FAM163A* (n=1), *FXR2* (n=1), *GLYR1* (n=1), *GPR97* (n=1), *HLA-DRB1* (n=4), *HLA-DRB5* (n=1), *KCNK18* (n=1), *KCNQ3* (n=1), *LYST* (n=1), *MAPRE3* (n=1), *MUC16*(n=2), *TTC13* (n=1) and *RBMX* (n=1)). These 29 sporadic variants they are all Likely Pathogenic according to ACMG classification guideline as they fulfilled two criteria of “a null variant” and “assumed as de novo but without paternity and maternity

confirmation test". None of the sporadic variants have been uniformly shared by all the affected subjects, either among CP subjects with the same phenotypes nor between the monozygotic twins in Family 4. However, a pair of sibling with different phenotypes in Family 5 (B5-1_S7 and G5-2_S10) share three de novo mutations which are frameshift deletion in *MUC16* of exon56:c.40674_40677del:p.K13558fs, frameshift insertion in *MUC16* of exon56:c.40672_40673insTCGG:p.K13558fs and frameshift insertion in *KCNK18* of exon3:c.361dupT:p.G120fs.

For the purpose of checking the status of all the variants reported, we used the online Leiden-Open Variant Database (LOVD) available at *https://www.lovd.nl* to check whether or not these specific variants had been reported before. This status checking step was done repeatedly as this LOVD database is regularly updated. The LOVD is a reliable database and has been used in many studies to check for the locus specific variant reported. Its system of locus-specific databases (LSDBs) is maintained according to the Human Genome Variation Society (HGVS) recommendations (Fokkema et. al., 2005; Fokkema et. al., 2011). All the eight inherited variants found in our study are previously not reported to be related to CP and are also not available in LOVD, indicating that these variants are new and have never been reported to be associated with any diseases or disorder previously. Similar with the inherited variants, all of sporadic variants are also not found to be associated with CP previously. Almost all of them are also not reported in LOVD except for two variants of g.135956571_g.135956572insG *RBMX* mutated gene and g.32552137_g.32552138insA *HLA-DRB1* mutated gene. With such result summary, we can clearly see that there is no definite pattern of correlation between these 37 variants

and CP phenotypes due to limited variants analysis available in this study. Plus, throughout the literature, none of these 37 variants has been reported to be contributing to the CP pathology. However, we noticed that a few genes belong to the CNS-related genes category (~31.03%), immune system-related genes category (~31.03%), potentially immune-related genes category (~10.34%) and other category genes (~27.59%). We are anticipating that these immune- and CNS-related genes are implicated in our CP subjects. This assumption is supported by two evidences which are the CP definition itself and the association of autoimmune disease with CP. CP as a chronic disability of CNS origin (Grether et. al., 1992) is also discovered to be associated with autoimmune disorder whereby autoimmune plays important role in CP etiology (Nelson et. al., 1998). This literally opens up for more potential studies in the future since all of these variants have never been previously reported to be CP related variants/genes. Since these 37 variants of 29 genes that we discovered in our CP data have never been reported before, they will be discussed in more detail based on specific gene categories instead of dissecting strictly only on the obtained novel variants, by expecting at least some plausible pathway or mechanism of any genes related that might have potential to be associated with CP.

4.2 CNS-related Genes Category

For the CNS-related genes category, there are nine genes that are implicated in our data. Eight of them are *CDKL2, CEP164, FAM104B, FAM163A, FXR2, KCNK18, KCNQ3, MAPRE3* and *RBMX* in which each gene shows one de novo mutation (n=1). Meanwhile another gene is *FAM104B* with two inherited variants (n=2). *CDKL2*, also known as protein kinase P56 *KKIAMRE* is discovered with only one sporadic frameshift variant,

exon3:c.222dupA:p.R75fs (4q21.1) that is present in two subjects i.e B1-1_S1 (diplegia) and B4-T2_S2 (hemiplegia). This gene was found to be expressed in human fetal brain (Taglienti et. al., 1996; Fang et. al., 2018). It was also discovered as a member of CDC2-related serine/threonine protein kinases family (Milanesi et. al., 2005). This *CDKL2* together with the other *CDK* domains are localized at the base and/or tip of primary (non-motile) cilia which are directly involved in regulating primary cilia length and function by influencing the Intraflagellar Transport (IFT) activity (Niwa et. al., 2012; Canning et. al., 2018). Besides *CDKL2, CEP164* is identified as a marker for the primary cilia structure formation and it is localized at the matured centriole (Graser et. al., 2007). *CEP164* expression occurs throughout the brain development (Devlin et. al., 2020). In our study, a variant of *CEP164* of frameshift insertion, exon5:c.337dupA:p.K113fs (11q23.3) is present in B1-3_S4 (quadriplegia). Primary cilia that are found in most eukaryotic cells including neurons, play essential roles in cell signaling, human sensory physiology, and development (Oh et. al., 2012; May-Simera et. al., 2012; Mukhopadhyay et. al., 2014).

It is also significant for cerebellar sonic hedgehog (Shh) dependent expansion (Spassky et. al., 2008; Boddaert, et. al., 2010) that has multiple roles during brain development (Ingham et. al., 1998; Spassky et. al., 2008), this includes the importance of primary cilia during the embryonic development (Wheway et. al., 2018). Based on many studies that showed the significance of primary cilia in brain development, we suspect that any mutations in *CDKL2* and/or in *CEP164*, could affect its role in regulating primary cilia function and structure which can result in primary cilia defect. This defect could be a potential pathomechanism which might lead to CNS defect. This assumption is

supported by the finding of primary cilia dysfunction that has been shown to cause CNS defect (Lee et. al., 2011), which indirectly reminded us of CP as a CNS disability origin. Furthermore, the primary cilium is classified as a cell structure and metabolism organelle, therefore the deficiency in this organelle because of the gene error may contribute to metabolic problem which is also a part of CP comorbidities (Gupta et. al., 2001). Further functional studies of these *CDKL2* and *CEP164*, CNS-related genes in our CP data are suggested since not many studies are found for CP related genes, plus both variants detected in our study are not reported in LOVD, indicating that they are new variants. Apart from CP, this *CDKL2* was also discovered as a novel promoter of breast cancer progression (Li et. al., 2014) while *CEP164* was discovered as nephronophthisis related gene (Chaki et. al., 2012; Slaats et. al., 2014). Both are epithelial-mesenchymal transition (EMT) related genes (Li et. al., 2014; Slaats et. al., 2014).

FAM104B has two inherited variants in our data which are exon3:c.182_183insA:p.S61fs (frameshift deletion) and exon3:c.179delC:p.A60fs (frameshift deletion) at Xp11.21 chromosome band. Both variants are present in all affected children and unaffected parents. These variants are classified as Likely Pathogenic according to ACMG classification standard guideline (Richard et. al., 2015) and have never been reported in LOVD showing they are new variants. However, currently it seems that the variants have been reported but are not related to the development of CP. According to Non-Human Primates Reference Transcriptome Resource (NHPRT project), this gene was reported in NCBI database as ubiquitously expressed, including in the brain as available at

https://www.ncbi.nlm.nih.gov/ieb/research/acembly/av.cgi?db=human&term=FAM104B&submit=Go.

FAM163A with a sporadic frameshift variant of exon5:c.400delC:p.P134fs (1q25.2) is present in B4-T2_S2 (hemiplegia) and G5-2_S10 (diplegia), whereby the variant was not reported previously. This gene is also known as *Cebelin* or neuroblastoma derived 94 secretory gene (*NDSP*) which is located at chromosome band 1q25.2 that encodes for 167AAs protein with a putative signal peptide and previously was shown to be associated with neuroblastoma cases (Vasudevan et. el., 2009). A study by Miwa's team (2018) showed that mouse Cebelin gene is actually an ortholog of the human Cebelin gene. They found that human cebelin amino acid (AA) sequence (167 AAs) with a putative secreted signal sequence (30 AAs) is almost similar to mouse cebelin protein by ~85% AA identity. This gene expression in the cerebellum is influenced by its own encoded cellular protein (167 AAs) (Miwa et. al., 2018). Thus, we anticipate that the mutation in this gene might influence its expression in the cerebellum which later may affect cerebellum role as a motor system central component that is closely related to muscle tone, motor activity coordination and posture (Stoodley et. al., 2009) whereby these three mentioned motor systems are related to CP clinical features. No specific publication on these variants' expression is found. Throughout previous studies that used NGS technology, *FAM104B* was shown to be related with some diseases, for example in the novel X-linked syndrome (Barboza et. al., 2013) and prostate cancer (Gambhira et. al., 2015). However, these NGS studies did not discuss in detail the function and pathophysiology of this gene. This gene function remains unknown (Barboza et. al., 2013; Kulecka et. al., 2017). Similar to the aforementioned studies, our

current preliminary analysis study cannot prove that the FAM104B gene is associated with CP as available information is limited. The *FAM104B* variants may have essential roles in CP diagnostic and/or prognostic. Further functional analysis studies are therefore needed in the future. Fragile X Mental Retardation 1 Autosomal Homolog 2 (*FXR2*) gene was found to be expressed in brain, testis and prenatal tissue (Tamanini, et. al., 1997; Agulhon et. al., 1999). In our data, this CNS-related gene with a de novo mutation at exon1:c.25delG:p.D9fs of chromosome band 17p13.1 is present in B1-3_S4 (quadriplegia) and B5-1_S7 (diplegia). This variant is not reported previously. Throughout the literature, the *FXR2* role was found as a homolog to Fragile X Mental Retardation 1 (*FMR1*) gene, in which its protein will form a protein family with *FMR1* and *FMR1 Autosomal Homolog 1* (*FXR1*) by structure and function. This interaction among these genes is related to Fragile X syndrome pathogenesis (Zhang et.al., 1995).

Molecularly, with 60% identity that is similar to an unknown specific function FMR1 gene, this *FXR2* encodes a 74 kDa FXR2 novel protein which contains two KH domains that has RNA binding capability and is cytoplasm-localized just like *FMR1* (Zhang et. al., 1995). The previous data was only showing the peripheral function of *FXR2* as a homolog that supports the main functional gene, whereas the publication on its main function has been hardly found. Hence the *FXR2* knockout study was done using mouse model that showed the *FXR2* knockout mice was hyperactive, had motor coordination impairment, prepulse inhibition reduction and less sensitive to heat stimulus. These were collectively classified under CNS function (Bontekoe et. al., 2002). Other than that, microdeletion in *FXR2* has been reported to be associated with disorders of dysmorphic features and developmental delay (Kumari et. al., 2019). Therefore, we

suspect that any pathogenic or possibly pathogenic mutations that happen to *FXR2* might impact its role in CNS which later may lead to CP specifically on developmental delay and motor coordination impairment.

In our study, this *KCNK18* shows a de novo mutation of exon3:c.361dupT:p.G120fs at 10q25.3 that is present in both diplegic affected children of Family 5, BS-1_S7 and G5-2_S10. This new variant has not been reported in LOVD. Previously *KCNK18* was found to be associated with migraine with aura disease (Lafrenière et. al., 2010). Its function is to encode for the tandem of pore domains in a weak inward rectifying K+ channel (TWIK)-related spinal cord potassium (K+) channel (TRESK), a member of K+ channel superfamily proteins that contains two pore-forming P domain K+ (K2P). TRESK protein functions as an outward rectifying K+ channel and a background K+ channel that sets the resting membrane potential in spinal cord (Sano et al., 2003; Enyedi et. al., 2010). These K+ channels involve in many cellular processes such as the action potential maintenance (Maljevic et. al., 2013), muscle contraction (Petkov G. V., 2011), hormone secretion (Ashcroft et. al., 2005; Cherki et. al., 2006; MacDonald et. al., 2006; Stojilkovic et. al., 2017), osmotic regulation (Andronic et. al., 2013; Xu et. al., 2016) and ion flow (Waxman et. al., 2014). The background K+ currents stabilize the negative resting membrane potential and counterbalance depolarization (Enyedi et. al., 2010). The hyperpolarization of resting membrane potential affects the individuals' excitability properties (Klein et. al., 2015). This is consistent with other studies that showed CP individuals having hyperexcitable motoneurons (Barolat-Romana et. al., 1980; Futagi et. al., 1985; O'Sullivan et al., 1998; Tekgul et al., 2013). By definition, hyperpolarization is a condition that makes the neuron to less likely transmit an impulse

(Martin et. al., 2008). Further study is suggested to investigate any effects of *KCNK18* mutation on its spinal cord background K+ channel function in setting up the resting membrane potential which might be responsible for hyperpolarization condition resulting in motor neuron impairment in CP. In our CP data, the *KCNQ3* has a sporadic variant of exon12:c.1599delA:p.K533fs, a frameshift mutation at 8q24.22 that is discovered in B4-T1_S1 (diplegia) and G5-2_S10 (diplegia). *KCNQ3* protein is expressed in the brain (Kanaumi et. al., 2008). This gene encodes for a subunit protein which later cooperates with another subunit encoded by *KCNQ2* to make up the M-channel (current) (Wang et.al., 1998). The M-current regulates the neuron excitability and synaptic responsiveness (Brown D.A., 1988; Yamada et. al., 1998), which participates in behavior and cognition (Rutecki et. al., 1992).

Thus, we anticipate that the mutation in KCNQ3 gene might affect its subunit protein function which later contributes to abnormal excitability resulting in brain electrical activity changes that lead to features of behavior and cognition problems in CP patients. This assumption is supported by a study by Klein and team (2015) that showed the neuron excitability in CP patients was found to be abnormal than the normal individuals, as explained earlier in *KCNK18* part. Other than that, the mutation in either KCNQ2 or KCNQ3 genes was shown to cause inherited autosomal epilepsy (Lewis et. al., 1993; Biervert et. al., 1998; Charlier et. al., 1998; Singh et. al., 1998) and familial epilepsy (Miceli et. al., 2015), whereby epilepsy is a CP comorbidity with highly association (Knezević-Pogancev et. al., 2010). Several studies had shown that epilepsy is commonly present in hemiplegia and quadriplegia types of CP (Hadjipanyis et. al., 1997; Kwong et. al., 1998; Odding et. al., 2006). However, this is slightly different to

our data when this *KCNQ3* mutation is found in diplegia subjects instead of hemiplegia and quadriplegia. Therefore, more studies should come later to verify the *KCNQ3* mutation whether or not it contributes to epilepsy in CP and larger sample size is suggested for the purpose of analyzing the association between epilepsy and diplegia types of CP.

MAPRE3, a gene that is previously reported to be associated with pulmonary hypertension (Atwal, 2017). It is expressed in the brain (Su et. al., 2001). Our result shows there is only one sporadic variant of *MAPRE3* which is exon5:c.536delC:p.A179fs (2p23.3) that is present in three diplegic subjects of different families, B1-1_S1, B4-T1_S1 and G5-2_S10. This specific variant is not recorded in LOVD indicating it is a new variant. *MAPRE3* expresses RP/EB protein family that promotes microtubules growth rate, whereby microtubules are involved in the spindle function (Yang et. al., 2017a). Spindle formation requires microtubules network to constantly alter its dynamics and organization as a cell supporting network during the cell division process. Microtubules as a cytoskeleton unit relatively acts together with the actin filaments as the intracellular transport tracks throughout the neuron (Nirschl et. al., 2017) and interacts with motor protein by binding to myosin and kinesin respectively (Woehlke et. al., 1997). In accordance to that, we assume that any possible mutations in *KCNQ3* might affect the RP/EB protein expression which later may lead to malfunction of microtubules. Microtubules deficiency may contribute to poor binding interaction with kinesin and myosin motor proteins, resulting in interrupted function of intracellular transport tracks. This is supported by a study that showed implication of intracellular trafficking defect on an amyotrophic lateral sclerosis (ALS)

neurodegenerative disorder, a disorder which shares some common features with CP on limb paralysis and brain plasticity (Soo et. al., 2011). Besides that, the interrupted binding interaction between motor proteins and microtubules due to kinesin malfunction had shown to be contributing to neurodegenerative diseases called hereditary spastic paraplegia (HSP) (Reid et. al., 2002). HSP is a condition which appears as the most sensitive to the underlying mutations, where it is actually the result of degeneration of the longest motor and sensory axons on the spinal cord (Crosby et. al., 2002). Hence, we suspect that the mutation of the genes which encode for any of myosin, kinesin or microtubules itself might interrupt that binding interaction between them which then lead to dysfunctional motor protein in intracellular trafficking function. In parallel to that, further study is suggested in validating the implication of *MAPRE3* mutation on microtubules growth rate, which might be contributing to CP features such as limb paralysis and brain problem.

Heterogeneous Nuclear Ribonucleoprotein G (*hnRNPG*) or also called as *RBMX* with a sporadic variant of exon9:c.905_906insCC:p.P302fs at Xq26.3 is found in B1- 3_S4 (quadriplegia) and B4-T1_S1 (diplegia). This variant is reported in LOVD database and marked by a clinical team of University of Medical Center, Netherland, however there is no class clinical reported plus it is recorded as “unpublished data”. RBMX is ubiquitously expressed in many organs such as brain, lung, heart, skeletal muscle, liver, pancreas, kidney, small intestine, colon, spleen, thymus, leukocytes, prostate, placenta, ovary, and testis (Lingenfelter et. al., 2001; Nasim et. al., 2003). From a study on RNA-binding motif, Y chromosome gene (*RBMY*), it showed that human genome *RBMX* is a homologue to *RBMY* (Delbridge et. al., 1999). Both genes encode for RNA binding

protein that is involved in splicing (Soulard et al. 1993; Delbridge et al. 1999; Venables et al. 2000). A study by Tsend-Ayush et. al., 2015 showed that zebrafish *rbmx* cDNA (human *RBMX* orthologue) plays an essential role in the brain development whereby its loss of function leads to the Zebrafish brain morphology defect. Thus, any alterations in *RBMX* is suspected to influence RNA binding protein expression, whereby the RNA binding protein is involved in the transcriptional regulation (Omura et. al., 2009) and it plays important roles in the CNS (Sharifnia et. al., 2015). Further study on human RBMX protein functional study is suggested to confirm its association in any possible CNS defects especially in the brain. For additional information, *RBMX* was implicated in other cases such as cancer and viral infection (Elliot et. al., 2019) and also in human systemic lupus erythematosus (SLE) disease (Soulard et al., 2002).

4.3 Immune-related Genes Category

The immune system-related gene category includes nine genes (*COL6A6* (n=1), *COL7A1* (n=1), *CYHR1* (n=1), *DLL1* (n=1), *GPR97* (n=1), *HLA-DRB1* (n=4), *HLA*-DRB5 (n=1), *LYST* (n=1) and *MUC16* (n=2)) that are present in our CP data which are all 13 de novo mutations, indicating that these alteration are present in the CP subjects only. In our study, *COL6A6* at chromosome band 3q22.1 of a de novo frameshift mutation of exon19:c.4620delC:p.G1540fs is present in two subjects of B1-3_S4 (quadriplegia) and B4-T2_S2 (hemiplegia). This variant has never been reported. *COL6A6* was found to be widely expressed in human body such as fetal and adult lung, kidney, liver, spleen, heart, skeletal muscle (Fitzegerald et. al., 2008) and skin (Sabatelli et. al., 2011; Soderhall et al., 2007). *COL6A6* that has 36 exons and a coding region of 6789bp is responsible to encode for a 2262-amino acid (alpha 6 chain) (Fitzegerald et.

al., 2008). Collagen VI (COL6) takes part as a connective tissue component by allowing the appropriate connection between muscle cells and extracellular matrix (ECM) through its binding to collagen IV and perlecan in the basal lamina (Kuo et. al., 1997). Despite its expression could be found abundantly in matrix of tissues and was also discovered to be associated with basement membranes, the precise role of *COL6* with *alpha chain 1* to *6* (*α1* to *α6*) is still not clearly defined (Fitzegerald et. al., 2013). However, its presence in skeletal muscle was proven to be functionally impactful when the defect in its protein expression was pathogenically causing collagen VI-related myopathies, a subgroup of neuromuscular disease which is an autoimmune disorder (Tagliavini et. al., 2014; Hunter et. al., 2015). With its implication in our study data, we anticipate that besides contributing to skeletal muscle impairment, it may also develop a possibility of immune disease as secondary condition in CP individuals.

Next, *COL7A*, with a sporadic variant of exon69:c.5749delG:p.E1917fs at 3p21.31 chromosome band is present in diplegic subjects (B2-2_S5 and B4-T1_S1). This *COL7A1* with the large exon number of 118 exons encodes for a 290 kDa α chain that makes up the collagen VII (COL7) (Christiano et. al., 1994b), making up a major component of the skin basement membrane component called anchoring fibrils (Chung et. al., 2010). *COL7A1* was found to be expressed in the skin (Christiano et. al., 1994b), pituitary gland, pineal gland and choroid plexus (Paulus et. al., 1995). Multiple mutations in *COL7A1* were previously found to be causing dystrophic epidermolysis bullosa (DEB), a disease of a fragile and vulnerable skin condition that occurs when the epidermis is not attached to the layer beneath it called dermis. This is due to the failure function of protein that is expressed by *COL7A1* that is supposedly to act as an adhesive

agent which binds these two mentioned skin components (Christiano et. al., 1994a; Hovnanian et. al., 1997; Massé et. al., 2005; Uitto et. al., 2005). The DEB is an inherited skin disease named Epidermolysis Bullosa (EB), in which multiple *COL7A1* mutations were found to be contributing to both dominant DEB andrecessive DEB types (Varki et. al. 2007). A study showed that interleukins and cytokines were both associated with *COL7* regulation and their expression level were significantly high in DEB patients indicating the potential of autoimmune disease development (Esposito et. al., 2016). Other than that, cytokines level was significantly correlated with the skin lesion (Ludwig et. al., 2011). Despite our *COL7A1* variant of c.5749delG is not reported as one of those *COL7A1* mutations that caused autoimmune disease of DEB, however the aforementioned pathological information earlier gives us a notion that the presence of *COL7A1* mutation in CP may also develop a potential comorbidity of immune disease in CP patients and it also might be a possible contributing gene for skin lesions problem in CP individual. Plus, the *COL7* via animal model study was shown to be involved in autoimmune reaction (Woodley et. al., 2007; Ishii et. al., 2011).

CYHR1 which is also known as Cysteine and Histidine Rich (*CHRP*) gene had never been found in any CP cases before. However, in our study, this protein coding gene with a sporadic frameshift deletion of exon2:c.430delG:p.A144fs is found in diplegic G2-1_S8 and B4-T1_S1. Through a study of two yeast hybrid systems, this gene was first discovered to be prominently localized in cytoplasm and perinuclear and its function is to assist in intracellular trafficking by binding itself to a galectin-3 molecule in a carbohydrate- independent manner in order to direct the lectin into the nucleus or into secretory pathways (Menon et. al., 2000). Similar to intracellular trafficking case

that has been discussed in *MAPRE3* part in the CNS category (section 4.2), we suspect that, any defect in *CYHR1* will also affect its protein binding activity, which then may lead to dysfunctional motor protein that contribute to CP features of limb paralysis. As an additional information, this *CYHR1* was reported to be implicated in tumors with high microsatellite instability (Kim et. al., 2013). Besides that, *CYHR1* mRNA expression was shown to be associated with lymph node metastasis (Desaki et. al., 2017). It is well-known that the lymph node is a part of the immune system in human body (Cottier et. al., 1973). Thus, we consider this finding of *CYHR1* as another immune-related gene that has the potential of being explored further in terms of its role in CP- immune disease association. This will indirectly increase the number of *CYHR1* study record for the future genomic database. Another immune-related gene found in our CP data is the *DLL1* which is also known as *Delta 1* at chromosome band 6q27, with a sporadic stopgain mutation of exon4:c.C663A:p.C221X that is present in diplegic B1-1_S1. This variant has never been reported in LOVD and none of the previous studies shows this gene to be associated with CP.

This *DLL1* with expression in bone marrow and dendritic cells (Sugimoto et. al., 2010; Huang et. al., 2011; Kelliher et. al., 2018) is encoding for the delta-like 1 ligand that has a binding interaction with the notch 3, one the four notch family receptors (Maekawa et. al., 2003; Amsen et. al., 2004; Mehrotra et. al., 2012). This notch-delta ligand binding induces its signaling function, which is well known as the conserved pathway in regulating cell differentiation and cell fate decisions (Greenwald, 1998; Artavanis-Tsakonas et al., 1999; Jaleco et. al., 2001; Gridley T., 2003). The delta 1-notch 3 binding is known to be involved in immune system development of T-cells and B-cells (Jaleco

et. al., 2001). Another study also showed this binding interaction plays an important role in type 1 helper T cell (Th1) development, a crucial cell that is involved in immunity system (Maekawa et. al., 2003). Few studies showed that *DLL1* alteration affects its ligand binding interaction with the notch 3 resulting in perturbed immunity responses and made it susceptible to visceral leishmaniasis, a fatal parasitic disease (Jamieson et. al., 2007; Miller et al., 2007; Fakiola et. la., 2011; Mehrotra et. al., 2012). With all the above details, it is clearly shown that DLL1 is an impactful immune-related gene, thus we presume that any alterations in its gene sequence may trigger an immunity problem in CP disorder as well since it is implicated in our CP data. The next immune-related gene in our CP data is the *GPR97* or also known as an *orphan Adhesion* G Protein-Coupled Receptor G3 (*ADGRG3*) that is present in diplegic (B2- 2_S5 and B5-1_S7) with a sporadic frameshift insertion of exon9:c.1022dupG:p.R341fs at 16q21. This gene protein is expressed in B cells, thymocytes (Sleckman et. al., 2000) and granulocytes (Hsiao et. al., 2018). GPR97 protein is one of *Adhesion* G Protein-Coupled Receptor (GPCR) members. The *Adhesion* GPCR group is the second largest subgroup of GPCR groups that contains 33 members that are characterized by their unique combination of cell adhesion and signaling properties (Bjarnadóttir et. al., 2007). The *GPR97* that was previously known as *Pb99* was first isolated from a cDNA library derived from pre-B cells and thymocytes of mouse, this study was initially focusing on its potential function in the early lymphoid development. Nevertheless, they found out that the *GPR97* loss of function had no effect on in vivo B and T cell maturation, which automatically indicated that there was likely a functional redundancy with another protein (Sleckman et. al., 2000). Hence, *GPR97* functional study was continued by the other teams, which discovered its roles to regulate the B-cell development and lymphatic endothelial cells (LECs) migration (Wang et. al., 2013; Valtcheva et. al.,

2013). The loss of function of *GPR97* was shown to be causing the imbalance ratio of cell division cycle 42 (Cdc42) protein and Ras homolog family member A (RhoA) protein levels, which is enhancing the LECs migration (Valtcheva et. al., 2013). This migration of LECs is important for vessel growth (lymphangiogenesis) and vessel remodeling processes. These two processes are responsible in modifying the lymphatic network in response to developmental or pathological demands (William et. al., 2017). Another study showed its role in regulating the antimicrobial activity in human granulocytes via its ligation by the immobilized Immunoglobulin G1 (IgG1) or monoclonal antibody G97-A mAb (Hsiao et. al., 2018). To date, we do not find any published papers showing this gene as a causative gene which is associated with any diseases or disorders. Instead, a recent study showed the *GPR97* is a potential therapeutic tool for the autoimmune disease (Wang et. al., 2018). It is well known that the *GPR97*/*ADGRG3* role is actively involved in the immune system, however there is a review paper which proposes that the *Adhesion* GPCRs groups also play roles in CNS.

According to in vivo reverse transcriptase-PCR (RT-PCR) study involving a large number of rat tissues for all the adhesion GPCRs, it showed that a large number of the newly discovered adhesion GPCRs members were abundantly expressed in CNS tissues (Bjarnadóttir et. al., 2007). Therefore, we would suggest for further characterization study to be done for this immune-related gene of *GPR97* to confirm whether or not it is involved in CNS besides its capability as therapeutic tool for autoimmune disease. Following that, other immune-related genes that are implicated in our CP data are *HLA-DRB1* and *HLA-DRB5* at 6p21.32. Shown in our data is the *HLA-DRB1* with four sporadic variants which are one frameshift insertion and three frameshift deletions. The

frameshift insertion of exon2:c.118_119insT:p.P40fs and all the frameshift deletions of exon2:c.115delC:p.Q39fs, exon2:c.111delG:p.L37fs, exon2:c.109delC:p.L37fs are present in diplegic G2-1_S8 and hemiplegic B4-T2_S2. Meanwhile, there is only one frameshift deletion of *HLA-DRB5* sporadic variant of exon2:c.171_172del:p.H57fs that is present in quadriplegic B1-3_S4 and diplegic B4-T1_S1. These *HLA-DRB1* and *HLA-DRB5* are known to be expressed in immune cells. The polymorphic HLA- DR is one of the molecules encoded by *MHC class II* which is expressed as α and β chain heterodimers (glycoproteins) on the cell surface. The glycoproteins are involved in binding and presentation of extracellular proteins peptides to T cells (TCRs) of T lymphocytes, whereby this activity is responsible to mediate the responses of cell-mediated immune and humoral (Cullen et. al., 1997; Van den Elsen et. al., 2011; Van den Elsen et. al., 2004).

For the record, this *HLA-DRB1* is the first genotype identified to be the most potential genetic risk factor for rheumatoid arthritis (Gregersen et. al., 1987; Diogo et. al., 2014). *HLA-DRB1* was also found to be associated with several immune diseases such as multiple sclerosis (Marrosu et. al., 2001; Rubio et. al., 2002; Dyment et. al., 2004) and systemic juvenile idiopathic arthritis (Ombrello et. al., 2015). However, its high expression in B-cells also was discovered to regulate the Chemokine Receptor Type 4 (CXCR4) expression on memory T helper cells (CD4+ T). This indicates that it has a potential of becoming therapeutic target for rheumatoid arthritis (Nagafuchi et. al., 2016). Meanwhile for *HLA-DRB5*, it was reported to be associated with scleroderma-related interstitial lung disease with high expression in peripheral blood mononuclear cell (Odani et. al, 2012), human systemic lupus erythematosus (SLE) (Wu et. al., 2013)

and amyotrophic lateral sclerosis (ALS) (Yang et. al., 2017). The *HLA-DRB5* variant in this study is a new variant since it has not been reported in LOVD. Meanwhile, one of four *HLA-DRB1* mutations (c.118_119insT) is clinically classified as benign in in LOVD, however there is no further information available. According to ACMG guidelines (Richard et. al., 2015), all of these sporadic variants of *HLA-DRB1* and *HLA-DRB5* are classified as likely pathogenic.

LYST is the next immune-related gene found in our CP data, with a sporadic stopgain mutation of exon40:c.C9558A:p.Y3186X at chromosome band 1q42.3 in B1-3_S4 (quadriplegia) and B2-2_S5 (diplegia). As an immune-related gene, recently it was discovered that this gene is required for the perforin-containing granules terminal maturation process into secretory, cytotoxic granules (Sepulveda et. al., 2015). This gene has never been reported to be involved in CP cases. Over the past two decades, this human *LYST* with 98.6% similar identity was identified as a homolog for mouse Beige gene, and it was also previously known as chediak-higashi syndrome 1 (*CHS1*) gene. This gene which is expressed in cytoplasm of perinuclear region is known to be regulating the lysosome fission (Perou et. al., 1997). Further studies was done later showing that it was molecularly functioning to encode for regulatory protein of intracellular trafficking process (Spritz R.A., 1998; Kaplan et. 2008). Chediak- higashi syndrome (CHS) is an autosomal recessive genetic disorder due to severe immune deficiency that is associated with lymphoproliferative syndrome, bleeding tendency, oculocutaneous albinism, and neurologic abnormalities (Spritz R.A., 1998; Introne et. al., 1999; Shiflett et. al., 2002; Ward et. al., 2002; Holt et. al., 2006; Kaplan et. al., 2008).

The encoded regulatory protein which previously called CHS1 protein contains a number of smaller sequence motifs comprised of a series ARM (ARMadillo) motif to mediate membrane associations (Peifer et. al., 1994) and repeat motifs of "HEAT" (huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A) regulatory A subunit, yeast-kinase (TOR1)) that were associated with vesicle transport (Andrade et. al., 1995). Meanwhile, the C-terminal region of this regulatory protein contains seven consecutive motifs called WD40 domain to mediate the protein-protein interactions (Neer et. al., 1994). Hence, we anticipate that, the mutation in *LYST* might impact the regulatory protein structure, whereby these perturbed motifs function may influence the regulatory protein pathway in intracellular trafficking process. Interrupted intracellular trafficking process is suspected to contribute to limb paralysis in CP. This anticipation is supported by a study that showed a missense mutation in *LYST* is implicated in HSP (Shimazaki et. al., 2014), a disorder that has similar limb paralysis feature with CP.

MUC16 at 19p13.2, was found in our CP data with two de novo mutations which are a frameshift deletion of exon8:c.597_600del:p.K199fs, and a frameshift insertion of exon8:c.595_596insTCGG:p.K199fs that are present in three diplegic CP subjects which are G2-1_S8 and both CP children in Family 5, B5-1_S7 and G5-2_S10. These variants have never been reported. Previously, it was confirmed that *MUC16* expresses the repeating peptide Cancer Antigen (CA125) molecule, a large glycoprotein of Mucin family (type 1 transmembrane) (O' Brien et. al., 2001; Felder et. al., 2014). *MUC16* which is also known as *CA125* was previously discovered to be promoting ovarian cancer cell proliferation (Rump et. al., 2004; Bast et. al., 2011). In corneal epithelial cells, the MUC16-Galectin-3 interaction acts as a barrier for bacterial and viral

infections in ocular epithelia (Argueso et. al., 2006; Argueso et. al., 2009). It was also shown that at the mucosal surfaces it serves as a protective, lubricating barrier against the particles and infectious agents (Gunn et. al., 2016). A few in vivo studies to investigate *MUC16* pathogenesis were done, unfortunately, there was no obvious functional defect shown after knocking down this MUC16 gene (Wang et. al. 2008; Cheon et. al., 2009), indicating that its pathological impact is still unclear. However, since *MUC16* is implicated in our CP data, with its involvement as protective agent in immune activities (Felder et. al., 2014), giving us a notion that this *MUC16* might has potential as a therapeutic target for any possible infections toward any CP patients with a weak immunity system.

In accordance to that, further studies are suggested to be conducted to reveal the CP-immune system association which may add a new insight to the prediction of how likely CP patients to be susceptible to immune disease as secondary condition to them. Besides that, all of these nine immune-related genes with total 13 variants are all 100% de novo mutations whereby they are only present in CP affected subjects while are totally absent in unaffected parents. Thus, we anticipate that these mutated immune-related genes might have potential in causing immune system problems in those CP affected individuals. This is supported by a study of Trisomy 21 case or also known as Down Syndrome (DS). This disorder is categorized as a neurological disorder, which is the same disorder classification as CP. However, recently this T21 chromosome aberration was found to be causing the immune dysregulation which may lead to chronic autoinflammation (Sullivan et. al., 2017). Having said that, we anticipate that besides

DS, CP also has the possibility of being an immune dysregulation-related neurological disorder.

4.4 Potentially Immune-related Genes Category

The genes (*ANKRD36*, *DNAH17* and *TTC13*) with ambiguous or unknown molecular function and/or have less reported studies were shown to be participating in some cases related to immune system. These genes are grouped as potentially immune-related genes to distinguish them from the confirmed nine immune-related genes (section 4.3) according to previous reported literature. Each of *DNAH17* and *TTC13* has one de novo mutation (n=1) whereas *ANKRD36* (n=2) has one de novo mutation and one inherited mutation. The *ANKRD36* at chromosome band 2q11.2 is the only gene that is implicated in our CP data with both types of sporadic and inherited variants of total three different mutations. For inherited, its mutations are exon14:c.1183_1184del:p.A395fs (frameshift deletion) and exon14:c.1186_1187insTT:p.V396fs (frameshift insertion) that are present in all subjects including affected CP subjects and unaffected parents, while its sporadic mutation is a stopgain of exon19:c.C1441T:p.Q481X that is present in diplegic B2- 2_S5 only.

More than 20 years ago, this *ANKRD36* was first obtained from the human cDNA and fetal brain and was termed as *KIAA1641* (Nagase et. al., 2000). However, it is reported to be upregulating in the lymph node, making itself as a biomarker for the endometrical cancer (Sudo et. al., 2011). Generally, the ANKRD is well known as a large group domain to maintain the protein-protein interactions that are involved in cell signaling,

nucleic acid management, cell cycle regulation, skeletal muscle development and immune responses (Nagase et. al., 1996; Nagase et. al., 2000; Voronin et. al., 2008; Tee et. al., 2010). However specifically for *ANKRD36*, its function is unknown (Chen et. al., 2016). Regardless of unknown function, this gene is reported to be associated with Parkinson's (Chew et. al., 2019), and Alzheimer's (Swaminathan et. al., 2010) whereby both are autoimmune diseases. Besides not knowing its molecular function, this ANKRD of specific domain 36 has never been reported to be associated with CP. Of note, the domain 15 of the same protein group (ANKRD) that is expressed by *ANKRD15* at locus 9p24.3, with its 225 kb deletion was discovered to be the causative gene for a familial CP in a Jewish family in a case-control study (Lerer et. al., 2005). This is giving us a hint that the ANKRD group and/or *ANKRD36* might have an impact on CP pathogenesis.

DNAH17 was found in our CP data with a sporadic frameshift deletion of exon27:c.423delA:p.K1408fs at 17q25.3 that is present in two diplegic subjects (B2-2_S5 and B5-1_S7). This variant has never been reported in LOVD. *DNAH17* expresses its protein called heavy chain 17 that is associated with axonemal dynein motor protein (Milisav et. al., 1998), a motor protein that is involved in mediating cilia or flagella movement by sliding the microtubules along each other (Lindemann et. al., 2010). *DNAH17* as potentially immune-related gene is expressedin adipocytes indicating its involvement in adipogenesis (Sohle et. al., 2012). Adipocytes via secreting factors play a role in mediating the physiology of immunology process (Lefterova et. al., 2009). There was a recent study that showed the loss of function in *DNAH17* caused isolated male infertility due to asthenozoospermia and they described that *DNAH17* was

responsible to encode for a sperm-specific axonemal outer dynein arm heavy chain (Whitfield et. al., 2019). No paper that describes *DNAH17* mechanism in human body to be associated with CP pathology has been found, thus further *DNAH17* pathogenesis study is suggested. Another suggested potential future analyses would be the *DNAH17* implication in immune system pathway. Next potentially immune-related gene is known to be encoding for tetratricopeptide repeat domain protein 13. This *TTC13* of 1q42.2 in our CP data is present with a sporadic frameshift deletion of exon12:c.1339delC:p.L447fs in B1-3_S4 (quadriplegia) and B2- 2_S5 (diplegia). For the purpose of knowing its pathogenesis potential and molecular function, we browsed through several reliable databases such as PubMed, Genecards, Uniprot and Google Scholar search engines. Unfortunately, unlike the other genes that are described before, this *TTC13* entry is hardly found.

Thus, we did not manage to collect information for this gene molecular function. We expect either this gene has not widely been studied or maybe its previous studies have not been published publicly. However, an in vivo study showed that TTC13 protein was highly expressed in kidney, liver and moderately expressed in the spleen, lymph node, T and B cells, which led to an assumption of this *TTC13* may has a role in the immune system (Shannon, 2012, *https://mospace.umsystem.edu/xmlui/handle/10355/43184*). Based on these threegenes that are grouped under potentially immune-related genes, we suggested for more studies in order to confirm their role in immune system and their impact in CP pathogenesis since currently the available literatures are too limited.

4.5 Other Gene Categories

These categories in the current study database would certainly covers eight genes. There are five genes with one sporadic variant (n=1) which are *ANO5*, *ASTE1*, *BEST3*, *DCHS2* and *GLYR1* and three genes with inherited variants are *CTBP2* (n=1), *CTDSP2* (n=2) and *KRTAP19-6* (n=1). ANO5 found in our CP data with a stopgain mutation of exon20:c.C2295A:p.Y765X at 11p14.3 is present in two diplegic subjects (B1-1_S1 and G5-2_S10). Via an animal study model, it is shown that this gene is expressed in skeletal muscle in the esophagus (Song et. al., 2014). As for its functional mechanism, it is reported to be involved in cellular structure and metabolism due to its active involvement in intracellular calcium activated chloride channel activity (Hartzell et. al., 2009). This gene nor its variant has never been reported to be associated with CP. Apart from that, its recessive mutation was also reported to be the causative variant for proximal limb-girdle muscular dystrophy (LGMD2L) and distal non-dysferlin miyoshi myopathy (MMD3) (Bolduc et. al., 2010).

Next is the *ASTE1* which is also a cellular structure and metabolism gene. Its novel frameshift mutation of exon6:c.1894delA:p.R632fs of 3q22.1 has never been reported. A functional study was done for this gene that is expressed in using the drosophila model, whereby the finding showed that this gene is involved in epidermal growth factor (EGF) signaling (Kotarski et. al., 1998; Stelzer et. al., 2016). The EGF signaling is a pathway that controls various biologic responses such as cell proliferation, cell differentiation and cell motility signaling (Oda et. al., 2005). For the current record, this gene was found to cause plantar fibromatosis disease (Zaragoza et. al., 2017). Another cellular structure and metabolism gene is *DCHS2* that encodes for a photocadherin

protein (Hong et. al., 2004). This gene with a stopgain mutation of exon5:c.C3293A:p.S1098X at 4q31.3 is present in two diplegic subjects (B1-1_S1 and B4-T1_S1). The cadherin family member gene (*DCHS2*) which is also known as *PCDH-J* is discovered to be expressing its protein in cerebral cortex (Hong et. al., 2004). Despite its unknown biological function, its SNP variant is discovered to be potentially associated with amnestic mild cognitive impairment and Alzheimer's disease (Vieira et. al., 2016). As for CP case itself, the *DCHS2* has never been reported, plus its stopgain variant in our study is a novel variant according to LOVD database screening. *DCHS2* interacts with Fat3 (atypical cadherin family member) to regulate the network controlling cartilage differentiation and polarity during vertebrate craniofacial development (Le pabic et. al, 2014).

Following that, the sporadic variant of frameshift mutation, exon7:c.773delT:p.L258fs at 12q15 of **BEST3** was found in our CP data in B1-3_S4 (quadriplegia) and B4- T2_S2 (hemiplegia) subjects. This cardiovascular system gene when its protein is observed to be expressed in vascular smooth muscle cells (VSMC) and was confirmed to be implicated in vitelliform macular dystrophy 2 disorder (Matchkov et. al., 2008). **BEST3** is responsible to regulate the membrane potential of cyclic guanosine monophosphate-dependent Calcium-Activated Chloride (cGMP dependent Ca2+- activated Cl−) conductance in VSMC Cells by coupling intracellular calcium and nitric oxide (NO)/cGMP pathways (Matchkov et. al., 2004; Matchkov et. al., 2008). This novel variant of BEST3 had never been reported to be associated with CP before. The last gene with sporadic variant in our data is *GLYR1* with a frameshift deletion, exon13:c.1140delG:p.G390fs at 16p13.3 that is found in B4-T1_S1 (diplegia) and G5-

2_S10 (diplegia). It has never been reported. This cellular structure and metabolism gene is characterized as a specific cofactor of the histone-lysine-specific demethylase, (LSD2/KDM1b) that stimulates histone demethylase activity by deriving the short peptides that can modulate LSD2/KDM1b's enzymatic activity and biologicalfunction (Fang et. al., 2013). The LSD2 is well-known as an epigenetic regulator that is involved in many biological processes including oocytogenesis, gene transcription and gene activation (Ciccone et al., 2009; Van Essen et al., 2010), meanwhile the lysine (K)-specific demethylase (KDM) family members play important roles in biological processes such as gene expression, cell growth, differentiation, development, and disease pathogenesis (Egger et al., 2004; Bhaumik et al., 2007; Esteller, 2008; Nottke et al., 2009). As additional information, another disease that is related to GLYR1 is a microsatellite-unstable colorectal cancer (Alhopuro et. al., 2012).

The *CTBP2* with an inherited variant of exon5:c.2228_2259del:p.G743fs, the frameshift deletion of 32 nucleotide at 10q26.3 is another cellular structure and metabolism gene that is present in all affected and unaffected subjects of our CP data. According to Katsanis et. al., 1998, CTBP2 protein is highly expressed in heart, skeletal muscle and pancreas. *CTBP2* encodes for two isoforms that are recognized as common isoform (48kD) and ribeye isoform (120kD). These isoforms enrich the *CTBP2* functions in cellular programs (Jung et. al., 2013) and transcriptional regulator (Chinnadurai, 2013) as available at *https://www.ncbi.nlm.nih.gov/books/NBK6557/*. An animal study showed that the *CTBP2* defect causes embryonic lethal in mouse (Hilderbrand et. al., 2002), and it is also contributing to cancer (Deng et. al., 2010; Guan et. al., 2013). *CTDSP2* is another cellular structure and metabolism gene. The variants

are two non-SNV variants (missense mutation) of exon4:c.C322A:p.L108I and exon4:c.G302C:p.R101T at 12q14.1, that are classified as uncertain significance due to insufficient criteria fulfilled according to ACMG guideline. It was found that this gene expression is in non-neuronal cells (Yeo et. al., 2005). It plays a role in the cell cycle arrest (Kashuba et. al., 2004; Zhu et. al., 2012) when its high expression decreases the number of cells in S phase (Kloet et. al., 2015). We hardly found any studies that show this gene is a causative gene, but it is proven that its involvement in the cell cycle arrest in hepatocellular carcinoma (Zhu et. al., 2012) and renal carcinoma (Kashuba et. al., 2004) making itself a worth tumor suppressor gene. This gene also has never been reported to be implicated in CP and its specific variant found in our study also has never been reported in any disorders or diseases. *KRTAP19-6* with an inherited variant of frameshift mutation of exon1:c.171delC:p.F57fs at 21q22.11.

The *KRTAP19-6* is described as one of the KRTAP family members (Rogers et. al., 2002). It is present in the hair cortex, hair keratin intermediate filaments that are embedded in an interfilamentous matrix. *KAP* polymorphisms are believed to be influencing the difference in the hair fibre structure across the individuals and population (Kariya et. al., 2005; Rogers et. al., 2006; Shimomura et. al., 2005). The *KRTAP19-6* has never been reported to be associated with CP cases before and its specific variant also has not been reported in any studies. Plus, its publication is also hard to be found, probably the gene has yet to be extensively studied for its genomic functionality.

CHAPTER FIVE

CONCLUSION

5.1 Summary and Conclusion

Literature has shown that CP cases with unknown causes are contributed by genetic factors since several genetic alterations have been discovered. Accordingly, this study explored the genomics of CP families that have multiple CP affected children with unknown causes. This has given new insights into CP genetics in selected CP families in Kelantan, Malaysia. Beginning with the clinical features, it is shown that CP phenotypes in this current study are clinically heterogenous for reasons such as siblings of the same parental line are classified with different phenotypes, and some CP subjects of the same CP phenotype also potray some differences in terms of clinical features, comorbidities and GMFCS. The clinical findings also proved that the CP condition is not their secondary medical condition but a possible result of genetic underlying factors since CP subjects show no dysmorphic features, no known genetic disorders, and they were born from uneventful pregnancies and uneventful deliveries.

Boinformatics analysis produced a total of 37 variants from 29 genes that were classified for their pathogenicities according to ACMG standard guidelines. They are mostly new variants that have never been reported except for a variant in *RBMX* (c.905_906insCC:p.P 302fs) and *HLA-DRB1* (c.118_119insT:p.P40fs). From this variant data, familial genetic CP is shown to be genetically heterogenous which both inherited and sporadic variants being implicated. We hypothesize that these genetic changes may contain important predictive information on this neurological disorder. Herein, exome sequencing identified a series of likely pathogenic null variants

implicating nervous and immune system genes in patients clinically diagnosed with familial CP. Interestingly, many mutated immune-related genes are found in CP affected subjects only which possibly indicates that the CP individuals may potentially have compromised immune system. Thus, CP could also be categorized as immune dysregulation-related neurological disorder. Identification of multiple new variants opens up for future studies to further understand the contribution of the related genes to CP. This could help future studies on molecular targeted drug development and CP therapeutics intervention.

5.2 Study Limitations and Perspectives

This current preliminary study is limited by the small sample size and the genetic alterations presented in our data need to be further validated using Sanger sequencing or any functional genomic assays. Future studies with larger sample size, more validation analysis and functional analysis are also suggested. Association of mutated immune-related genes with CP should also be studied to add new insights to the predictions of potential susceptibility to immune diseases as secondary medical condition and/or contribution of mutated immune-related genes towards CP susceptibility.

Another limitation in this study is the lack of resources and proper facilities in conducting the NGS study in this campus which include a server for big data storage and bioinformatician with expertise in NGS data analysis. Thus, the NGS and bioinformatics analysis were carried out in collaboration with Perdana University, Kuala Lumpur.

REFERENCES

Journal Articles

1. Abou Jamra, R., Philippe, O., Raas-Rothschild, A., et. al., (2011). Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. Am Journal Hum Genet, 88(6), p.788–95.

2. Agarwal, A., and Verma, I., (2012). Cerebral palsy in children: An overview. Journal of Clinical Orthopaedics and Trauma, 3(2), p.77-81.

3. Aguiar, S. M. H. C. Á., and Alves-Rezende, M. C. R., (2013). Dental care for cerebral palsy patients using auxiliary resources to inhibit pathological tonic reflexes. Archives of Health Investigation, 2(4), p.39-44.

4. Agulhon, C., Blanchet, P., Kobetz, A., Marchant, D., Faucon, N., Sarda, P., and Abitbol, M., (1999). Expression of FMR1, FXR1, and FXR2 genes in human prenatal tissues. Journal of Neuropathology and Experimental Neurology, 58(8), p.867-880.

5. Alhopuro, P., Sammalkorpi, H., Niittymaki, I., Bistrom, M., Raitila, A., Saharinen, J., Nousiainen, K., Lehtonen, H. J., Heliovaara, E., Puhakka, J., Tuupanen, S., Sousa, S., Seruca, R., et al., (2012). Candidate driver genes in microsatellite- unstable colorectal cancer. Int J Cancer., 130(7), p.1558–66.

6. Alkuraya, F.S., (2010). Homozygosity mapping: one more tool in the clinical geneticist's toolbox. Genetics in Medicine, 12(4), p.236-239.

7. Al-Rajeh, S., Bademosi, O., Awada, A., Ismail, H., Al-Shammasi, S., and Dawodu, A., (1991). Cerebral palsy in Saudi Arabia: a case-control study of risk factors. Dev Med Child Neurol., 33(12), p.1048–52.

8. Amoghimath, R., Gowda, V. K., and Benakappa, A., (2016). Comorbidities and their relationship to subtype of cerebral palsy in a tertiary care hospital in South India. Indian Journal of Cerebral Palsy, 2(2), p.90-93.

9. Amor, D. J., Craig, J. E, Delatycki, M. B., and Reddihough, D., (2001). Genetic factors in athetoid cerebral palsy. Journal of Child Neurol 16(11), p.793–97.

10. Amsen, D., Blander, J. M., Lee, G. R., Tanigaki, K., Honjo, T., and Flavell, R. A., (2004). Instruction of distinct CD4 T helper cell fates by different notch ligands on antigen-presenting cells. Cell, 117(4), p.515–526.

11. Andersen, G. L., Irgens, L. M., Haagaas, I., Skranes, J.S., Meberg, A.E. and Vik, T., (2008). Cerebral palsy in Norway: prevalence, subtypes and severity. European Journal of Paediatric Neurology, 12(1), p.4-13.

12. Andrade, M. A., and Bork, P., (1995). HEAT repeats in the Huntington's disease protein. Nature Genet, 11(2), p.115-116.

13. Andronic, J., Bobak, N., Bittner, S., Ehling, P., Kleinschnitz, C., Herrmann, A. M., and Sukhorukov, V. L., (2013). Identification of two-pore domain potassium channels as potent modulators of osmotic volume regulation in human T lymphocytes. Biochimica et Biophysica Acta (BBA)-Biomembranes, 1828(2), p.699–707.

14. Argueso, P., Guzman-Aranguez, A., Mantelli, F., Cao, Z., Ricciuto, J., and Panjwani, N, (2009). Association of cell surface mucins with galectin-3 contributes to the ocular surface epithelial barrier. Journal of Biol Chem. 284 (34), p.23037-23045.

15. Argueso, P., Tisdale, A., Spurr-Michaud, S., Sumiyoshi, M., and Gipson, I. K., (2006). Mucin characteristics of human corneal-limbal epithelial cells that exclude the rose bengal anionic dye. Invest Ophthalmol Vis Sci., 47 (1), p.113-119.

16. Artavanis-Tsakonas, S., Rand, M. D., and Lake, R. J., (1999). Notch signaling: cell fate control and signal integration in development. Science, 284(5415), p.770–776.

17. Ashcroft, F. M. (2005). ATP-sensitive potassium channelopathies: focus on insulin secretion. Journal of Clinical Investigation, 115(8), p.2047–2058.

18. Ashwal, S., Russman, B. S., Blasco, P. A., Miller, G., Sandler, A., Shevell, M., and Stevenson, R., (2004). Practice parameter: diagnostic assessment of the child with cerebral palsy: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child Neurology Society. Neurology, 62(6), p.851-863.

19. Atwal, G., Sun, M., Yang, X., Sun, Y., Chen, J. and Komarova, Y., (2017). The role of microtubule–associated protein EB3 in Pathogenesis of Pulmonary Hypertension. The FASEB Journal, 31(1_supplement), p.1016-21.

20. Bao, R., Huang, L., Andrade, J., Tan, W., Kibbe, W. A., Jiang, H., and Feng, G., (2014). Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. Cancer Informatics, 13(s2), p.67–82.

21. Barboza-Cerda, M. C., Campos-Acevedo, L. D., Rangel, R., Martínez-de-Villarreal, L. E., and Déctor, M.A., (2013). A novel phenotype characterized by digital abnormalities, intellectual disability, and short stature in a Mexican family maps to Xp11. 4–p11. 21. American Journal of Medical Genetics Part A, 161(2), p.237-243.

22. Barolat-Romana, G., and Davis, R., (1980). Neurophysiological mechanisms in abnormal reflex activities in cerebral palsy and spinal spasticity. Journal of Neurol.Neurosurg. Psychiatry, 43(4), p.333–342.

23. Bast, R. C., Spriggs, D. R., (2011). More than a biomarker: CA125 may contribute to ovarian cancer pathogenesis. Gynecol Oncol., 121(3), p.429-430.

24. Beligere, N., and Rao, R., (2008). Neurodevelopmental outcome of infants with meconium aspiration syndrome: report of a study and literature review. Journal of Perinatology, 28(3), p.S93-S101.

25. Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., Shang, L., Boisson, B., Casanova, J. L., and Abel, L., (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proceedings of the National Academy of Sciences, 112(17), p.5473-5478.

26. Bennett, F.C., (1999). Diagnosing cerebral palsy--the earlier the better. Contemporary Pediatrics, 16(7), p.65-65.

27. Bhaumik, S.R., Smith, E., and Shilatifard, A., (2007). Covalent modifications of histones during development and disease pathogenesis. Nat. Struct. Mol. Biol. 14(11), p.1008–1016.

28. Biervert, C., Schroeder, B. C., Kubisch, C., Berkovic, S. F., Propping, P., Jentsch, T. J., and Steinlein, O. K., (1998). A potassium channel mutation in neonatal human epilepsy. Science, 279(5349), p.403-406.

29. Bilgüvar, K., Öztürk, A.K., Louvi, A., Kwan, K.Y., Choi, M., Tatlı, B., Yalnızoğlu, D., Tüysüz, B., Çağlayan, A.O., Gökben, S. and Kaymakçalan, H., (2010). Whole exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature, 467(7312), p.207-210.

30. Bjarnadóttir, T. K., Fredriksson, R., and Schiöth, H. B., (2007). The adhesion GPCRs: a unique family of G protein-coupled receptors with important roles in both central and peripheral tissues. Cell. Mol. Life Sci., 64(16), p.2104–2119.

31. Blair, E., Al Asedy, F., Badawi, N., and Bower, C., (2007). Is cerebral palsy associated with birth defects other than cerebral defects? Dev Med Child Neurol, 49(4), p.252–58.

32. Blair, E., and Stanley F,. (1993). When can cerebral palsy be prevented? The generation of causal hypotheses by multivariate analysis of a case-control study. Paediatr Perinat Epidemiol., 7(3), p. 272–301.

33. Blair, E. and Stanley, F.J., (1988). Intrapartum asphyxia: a rare cause of cerebral palsy. The Journal of Pediatrics, 112(4), p.515-519.

34. Boddaert, N., Desguerre, I., Bahi-Buisson, N., Romano, S., Valayannopoulos, V., Saillour, Y., and Zilbovicius, M., (2010). Posterior fossa imaging in 158 children with ataxia. Journal of Neuroradiology, 37(4), p.220-230.

35. Bolduc, V., Marlow, G., Boycott, K. M., Saleki, K., Inoue, H., Kroon, J., and Mizuta, K. (2010). Recessive mutations in the putative calcium-activated chloride channel Anoctamin 5 cause proximal LGMD2L and distal MMD3 muscular dystrophies. The American Journal of Human Genetics, 86(2), p.213-221.

36. Bontekoe, C. J., McIlwain, K. L., Nieuwenhuizen, I. M., Yuva-Paylor, L. A., Nellis, A., Willemsen, R., and Cheng, N. C., (2002). Knockout mouse model for Fxr2: a model for mental retardation. Human Molecular Genetics, 11(5), p.487-498.

37. Boullerne, A. I., and Benjamins, J. A. (2006). Nitric Oxide Synthase Expression and Nitric Oxide Toxicity in Oligodendrocytes. Antioxidants & Redox Signaling, 8(5-6), p.967–980.

38. Brookes A. J. (1999). The essence of SNPs. Gene. (234), p.177–86.

39. Bu, D. F., Erlander, M. G., Hitz, B. C., Tillakaratne, N. J., Kaufman, D. L., Wagner-McPherson, C. B., Evans, G. A., and Tobin, A. J., (1992). Two human glutamate decarboxylases, 65-kDa GAD and 67-kDa GAD, are each encoded by a single gene. Proceedings of the National Academy of Sciences, 89(6), p.2115-2119.

40. Burton, M. J., (2008). The Surgical Management of Drooling. Developmental Medicine & Child Neurology, 33(12), p.1110–1116.

41. Bush, W. S. and Moore, J. H., (2012). Genome-wide association studies. PLoS Comput Biol, 8(12), p.e1002822.

42. Camacho-Gonzalez, A., Spearman, P. W. and Stoll, B. J., (2013). Neonatal infectious diseases: evaluation of neonatal sepsis. Pediatric Clinics of North America, 60(2), p.367.

43. Canning, P., Park, K., Gonçalves, J., Li, C., Howard, C. J., Sharpe, T. D., Holt, L. J., Pelletier, L., Bullock, A. N., and Leroux, M. R., (2018). CDKL family

kinases have evolved distinct structural features and ciliary function. Cell Reports, 22(4), p.885-894.

44. Chaki, M., Airik, R., Ghosh, A. K., Giles, R. H., Chen, R., Slaats, G. G., Wang, H., Hurd, T. W., Zhou, W., Cluckey, A. et al. (2012). Exome capture reveals ZNF423 and CEP164 mutations, linking renal ciliopathies to DNA damage response signaling. Cell 150(3), p.533-548.

45. Charlier, C., Singh, N. A., Ryan, S. G., Lewis, T. B., Reus, B. E., Leach, R. J., and Leppert, M. (1998). A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family. Nature Genetics, 18(1), p.53.

46. Chen, A., Dyck, H. S., Oskoui M, and Shevell, M., (2020). Canadian Cerebral Palsy Registry. Losing a diagnosis of cerebral palsy: a comparison of variables at 2 and 5 years. Dev Med Child Neurol., 62(1), p.83–88.

47. Chen, B. J., Mills, J. D., Takenaka, K., Bliim, N., Halliday, G. M., and Janitz, M., (2016). Characterization of circular RNA s landscape in multiple system atrophy brain. Journal of Neurochemistry, 139(3), p.485-496.

48. Cheon, D. J., Wang, Y., Deng, J. M., Lu, Z., Xiao, L., Chen, C. M., Bast, R. C., and Behringer, R. R., (2009). CA125/MUC16 is dispensable for mouse development and reproduction. PLoS One., 4(3), p.e4675.

49. Cherki, R., Luques, L., Anis, Y., and Meir, A., (2006). Ion Channels in Endocrine Pancreatic Cells and their Role in Diabetes. Modulator (21), p.16-21.

50. Chew, E. G. Y., Liany, H., Tan, L. C., Au, W. L., Prakash, K. M., Annuar, A. A., Chan, A. Y., Lim, S. Y., Mok, V., Chung, S. J., and Song, K., (2019). Evaluation of novel Parkinson's disease candidate genes in the Chinese population. Neurobiology of Aging, 74(February), p.235-e1.

51. Christiano, A. M., Greenspan, D. S., Lee, S., and Uitto, J. (1994b). Cloning of human type VII collagen. Complete primary sequence of the alpha 1 (VII)

chain and identification of intragenic polymorphisms. Journal of Biological Chemistry, 269(32), p.20256-20262.

52. Christiano, A. M., Anhalt, G., Gibbons, S., Bauer, E. A. and Uitto, J., (1994a). Premature termination codons in the type VII collagen gene (COL7A1) underlie severe, mutilating recessive dystrophic epidermolysis bullosa. Genomics, 21(1), p.160-168.

53. Chung, H. J., and Uitto, J. (2010). Type VII collagen: the anchoring fibril protein at fault in dystrophic epidermolysis bullosa. Dermatologic clinics, 28(1), p.93-105.

54. Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T., (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature 461(7262), p.415–418.

55. Clark, S. L., and Hankins, G. D., (2003). Temporal and demographic trends in cerebral palsy—fact and fiction. Am Journal of Obstet Gynecol, 188(3), p.628–33.

56. Condit, C. M., Achter, P. J., Lauer, I., and Sefcovic, E. (2002). The changing meanings of‘mutation:” A contextualized study of public discourse. Hum Mutat. (19), p.69–75.

57. Cottier, H., Turk, J. and Sobin, L., (1973). A proposal for a standardized system of reporting human lymph node morphology in relation to immunological function. Journal of Clinical Pathology, 26(5), p.317-331.

58. Crosby, A. H., and Proukakis, C., (2002). Is the transportation highway the right road for hereditary spastic paraplegia? The American Journal of Human Genetics, 71(5), p.1009-1016.

59. Cullen, M., Noble, J., Erlich, H., Thorpe, K., Beck, S., Klitz, W., Trowsdale, J. and Carrington, M., (1997). Characterization of recombination in the HLA class II region. American Journal of Human Genetics, 60(2), p.397.

60. Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., and Batzoglou, S., (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Computational Biology, 6(12), p.e1001025.

61. De Ligt, J., Willemsen, M. H., Van Bon, B. W., Kleefstra, T., Yntema, H. G., Kroes, T., Vulto-van Silfhout, A. T., Koolen, D. A., De Vries, P., Gilissen, C. and del Rosario, M., (2012). Diagnostic exome sequencing in persons with severe intellectual disability. New England Journal of Medicine, 367(20), p.1921-1929.

62. Delbridge, M. L., Lingenfelter, P. A., Disteche, C. M., and Graves, J. A. M., (1999). The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. Nature genetics, 22(3), p.223-224.

63. Deng Y., J. Liu, G. Han et. al., (2010). Redox-dependent Brca1 transcriptional regulation by anNADH-sensor CtBP1. Oncogene, 29(50), p.6603–6608.

64. Desaki, R., Sawada, G., Okumura, H., Ikeda, R., Tanabe, K., Komatsu, H., Mimori, K., Mori, M., Kita, Y., Uchikado, Y., Arigami, T., Uenosono, Y., Owaki, T., Ishigami, S., and Natsugoe, S., (2017). As a Novel Prognostic Marker, Cysteine/histidine-rich 1 (CYHR1) is a Therapeutic Target in Patients with Esophageal Squamous Cell Carcinoma. Annals of Surgical Oncology, 24(2), p.586- 593.

65. Devlin, L. A., Ramsbottom, S. A., Overman, L. M., Lisgo, S.N., Clowry, G., Molinari, E., Powell, L., Miles, C.G. and Sayer, J.A., (2020). Embryonic and foetal expression patterns of the ciliopathy gene CEP164. PloS one, 15(1), p.e0221914.

66. Diogo, D., Okada, Y. and Plenge, R. M. (2014). Genome-wide association studies to advance our understanding of critical cell types and pathways in rheumatoid arthritis: recent findings and challenges. Curr Opin Rheumatol 26(1), p.85–92.

67. Djukic, M., Gibson, C. S., MacLennah, A. H., Goldwater, P. N., Haan, E. A., McMichael, G., Priest, K., Dekker, G. A., Hague, W. M., Chan, A. and Rudzki, Z., (2009). Genetic susceptibility to viral exposure may increase the risk of

cerebral palsy. Australian and New Zealand Journal of Obstetrics and Gynaecology, 49(3), p.247-253.

68. Domagalska, M., Szopa, A., Czupryna, K., Nowotny, J. and Matyja, M., (2006). "The inhibiting casts" used in cerebral palsed children. Ortopedia, Traumatologia, Rehabilitacja, 8(3), p.291-299.

69. Dyment, D. A., Ebers, G. C., and Sadovnick, A. D., (2004). Genetics of multiple sclerosis. Lancet Neurol. 3(February), p.104–110.

70. Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature 429(6990), p.457–463.

71. Elliott, D. J., Dalgliesh, C., Hysenaj, G., and Ehrmann, I., (2019). RBMX family proteins connect the fields of nuclear RNA processing, disease and sex chromosome biology. The International Journal of Biochemistry & Cell Biology, 108(March), p.1-6.

72. Emamghorashi, F., Mahmoodi, N., Tagarod, Z. and Heydari, S.T., (2012). Maternal urinary tract infection as a risk factor for neonatal urinary tract infection. Iranian Journal of Kidney Diseases (IJKD), 6(3), p.178-180.

73. Enyedi, P. and Czirják, G., (2010). Molecular background of leak K+ currents: two-pore domain potassium channels. Physiological Reviews, 90(2), p.559-605.

74. Erkin, G, Delialioglu, S. U., Ozel, S., Culha, C., and Sirzai, H., (2008). Risk factors and clinical profiles in Turkish children with cerebral palsy: analysis of 625 cases. International Journal of Rehabil Res, 31(1): p.89–91.

75. Erlich, Y., Edvardson, S., Hodges, E., Zenvirt, S., Thekkat, P., Shaag, A., Dor, T., Hannon, G.J. and Elpeleg, O., (2011). Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. Genome Research, 21(5), p.658-664.

76. Esposito, S., Guez, S., Orenti, A., Tadini, G., Scuvera, G., Corti, L., Scala, A., Biganzoli, E., Berti, E. and Principi, N., (2016). Autoimmunity and cytokine imbalance in inherited epidermolysis bullosa. International Journal of Molecular Sciences, 17(10), p.1625.

77. Esteller, M. (2008). Epigenetics in cancer. N. Engl. J. Med. 358(11), p.1148–1159.

78. Fakiola, M., Miller, E. N., Fadl, M., Mohamed, H. S., Jamieson, S. E., Francis, R. W., and Elhassan, A. (2011). Genetic and functional evidence implicating DLL1 as the gene that influences susceptibility to visceral leishmaniasis at chromosome 6q27. Journal of Infectious Diseases, 204(3), p.467-477.

79. Fang, C. L., Uen, Y. H., Chen, H. K., Hseu, Y. C., Lin, C. C., Hung, S. T., and Lin, K. Y. (2018). Loss of cyclin-dependent kinase-like 2 predicts poor prognosis in gastric cancer, and its overexpression suppresses cells growth and invasion. Cancer Medicine, 7(7), p.2993-3002.

80. Fang, R., Chen, F., Dong, Z., Hu, D., Barbera, A. J., Clark, E. A., Fang, J., Yang, Y., Mei, P., Rutenberg, M. and Li, Z., (2013). LSD2/KDM1B and its cofactor NPAC/GLYR1 endow a structural and molecular model for regulation of H3K4 demethylation. Molecular Cell, 49(3), p.558-570.

81. Felder, M., Kapur, A., Gonzalez-Bosquet, J. et al. (2014). MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. Mol Cancer 13(1), p.1-15.

82. Fernando, T., and Goldman, R. D., (2019). Management of gastroesophageal reflux disease in pediatric patients with cerebral palsy. Canadian Family Physician, 65(11), p.796-798.

83. Figueroa, K. P., Waters, M. F., Garibyan, V., Bird, T. D., Gomez, C. M., Ranum, L. P., Minassian, N. A., Papazian, D. M. and Pulst, S. M., (2011). Frequency of KCNC3 DNA variants as causes of spinocerebellar ataxia 13 (SCA13). PloS One, 6(3), p.e17811.

84. Fitzgerald, J., Holden, P. and Hansen, U., (2013). The expanded collagen VI family: new chains and new questions. Connective tissue research, 54(6), p.345- 350.

85. Fitzgerald, J., Rich, C., Zhou, F. H., and Hansen, U., (2008). Three novel collagen VI chains, α4 (VI), α5 (VI), and α6 (VI). Journal of Biological Chemistry, 283(29), p.20170-20180.

86. Fletcher, N. A., and Foley, J., (1993). Parental age, genetic mutation, and cerebral palsy. Journal of Med Genet, 30(1), p.44–46.

87. Fokkema, I. F., den Dunnen, J. T. and Taschner, P. E., (2005). LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. Human Mutation, 26(2), p.63-68.

88. Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F. and den Dunnen, J. T., (2011). LOVD v. 2.0: the next generation in gene variant databases. Human Mutation, 32(5), p.557-563.

89. Frost, P. S., Barros-Aragão, F., da Silva, R. T., Venancio, A., Matias, I., e Silva, N. M. L., Kincheski, G. C., Pimentel-Coelho, P. M., De Felice, F. G., Gomes, F. C., and Ferreira, S. T., (2019). Neonatal infection leads to increased susceptibility to Aβ oligomer-induced brain inflammation, synapse loss and cognitive impairment in mice. Cell Death & Disease, 10(4), p.1-13.

90. Futagi, Y., and Abe, J. (1985). H-reflex study in normal children and patients with cerebral palsy. Brain Dev. 7(4), p.414–420.

91. Gambhira, R., Ledet, E. M., Dotiwala, A., Mandal, D. and Sartor, O., (2015). Evaluating cell-free DNA genetic alterations in men with metastatic castrate resistant prostate cancer. Journal of Clinical Oncology, 33(7), p.244-244.

92. Ganatra, H.A., and Zaidi, A.K., (2010). Neonatal infections in the developing world. In Seminars in Perinatology, 34(6), p.416-425.

93. Garne, E., Dolk, H., Krageloh-Mann, I., Holst, R. S., and Cans, C., (2008). Cerebral palsy and congenital malformations. Eur Journal of Paediatr Neurol, 12(2): p.82– 88.

94. Gibson, C. S., MacLennan, A. H. , Dekker, G. A., Goldwater, P. N., Sullivan, T. R., Munroe, D. J., Tsang, S., Stewart, C., and Nelson, K. B., (2008). Candidate genes and cerebral palsy: a population-based study. Pediatrics, 122(5), p.1079- 1085.

95. Gibson, C. S., MacLennan, A. H., Goldwater, P. N., and Dekker, G. A., (2003). Antenatal causes of cerebral palsy: associations between inherited thrombophilias, viral and bacterial infection, and inherited susceptibility to infection. Obstetrical & Gynecological Survey, 58(3), p.209-220.

96. Gibson, C. S., MacLennan, A. H., Goldwater, P. N., Haan, E. A., Priest, K., Dekker, G. A., and South Australian Cerebral Palsy Research Group, (2006). The association between inherited cytokine polymorphisms and cerebral palsy. American Journal of Obstetrics and Gynecology, 194(3), p.674-e1.

97. Gilstrap, L. C., Leveno, K. J., Burris, J., Lynne Williams, M., & Little, B. B. (1989). Diagnosis of birth asphyxia on the basis of fetal pH, Apgar score, and newborn cerebral dysfunction. American Journal of Obstetrics and Gynecology, 161(3), p.825–830.

98. Graser, S., Stierhof, Y. D., Lavoie, S. B., Gassner, O. S., Lamla, S., Le Clech, M. and Nigg, E. A., (2007). CEP164, a novel centriole appendage protein required for primary cilium formation. The Journal of Cell Biology, 179(2), p.321-330.

99. Greenwald, I., (1998). LIN-12/Notch signaling: lessons from worms and flies. Genes & Development, 12(12), p.1751–1762.

100. Gregersen, P. K., Silver, J. and Winchester, R. J., (1987). The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum, 30(11), p.1205–1213.

101. Grether, J. K., Cummins, S. K., & Nelson, K. B. (1992). The California cerebral palsy project. Paediatric and Perinatal Epidemiology, 6(3), 339-351.

102. Gridley, T., (2003). Notch signaling and inherited disease syndromes. Human Molecular Genetics, 12(suppl_1), p.R9-R13.

103. Guan, C., Shi, H., and Wang, H. et al., (2013). CtBP2 contributes to malignant development of human esophageal squamous cell carcinoma by regulation of p16 (INK4A.). Journal of Cellular Biochemistry, 114(6), p 1343–1354.

104. Gunn, B. M., Schneider, J. R., Shansab, M., Bastian, A. R., Fahrbach, K. M., Smith, I. V. A. D., Tedesco, J., et. al., (2016). Enhanced binding of antibodies generated during chronic HIV infection to mucus component MUC16. Mucosal immunology, 9(6), p.1549.

105. Gupta, R. and Appleton, R.E., (2001). Cerebral palsy: not always what it seems. Archives of Disease in Childhood, 85(5), p.356-360.

106. Gustavson, K. H., Hagberg, B., and Sanner, G., (1969). Identical syndromes of cerebral palsy in the same family. Acta Paediatr Scand, 58(4), p.330–40.

107. Hack, M., Flannery, D.J., Schluchter, M., Cartar, L., Borawski, E. and Klein, N., (2002). Outcomes in young adulthood for very-low-birth-weight infants. New England Journal of Medicine, 346(3), p.149-157.

108. Hadjipanayis, A., Hadjichristodoulou, C., and Youroukos, S., (1997). Epilepsy in patients with cerebral palsy. Dev Med Child Neurol, 39(10), p.659–63.

109. Han, T. R., Bang, M. S., Lim, J. Y., Yoon, B. H., & Kim, I. W. (2005). Risk factors of cerebral palsy in preterm infants.American Journal of Physical Medicine and Rehabilitation, 81(4), p.297-303.

110. Hartzell, H. C., Yu, K., Xiao, Q., Chien, L. T., and Qu, Z. (2009). Anoctamin/TMEM16 family members are Ca2+-activated Cl− channels. The Journal of Physiology, 587(10), p.2127-2139.

111. Heazell, A. (2012). Abnormalities of the placenta. BMC Pregnancy and Childbirth, 12(Suppl 1), p.A2.

112. Hegde, A. M., and Pani, S. C., (2009). Drooling of saliva in children with cerebral palsy-etiology, prevalence, and relationship to salivary flow rate in an Indian population. Special Care in Dentistry, 29(4), p.163–168.

113. Hemminki, K., Li, X., Sundquist, K., and Sundquist, J., (2007) High familial risks for cerebral palsy implicate partial heritable aetiology. Paediatr Perinat Epidemiol, 21(13), p.235–41.

114. Hildebrand, J. D., and Soriano, P., (2002). Overlapping and unique roles for C- terminal binding protein 1 (CtBP1) and CtBP2 during mouse development. Molecular and Cellular Biology, 22(15), p.5296-5307.

115. Hoare, B. J., Wasiak, J., Imms, C., and Carey, L. (2007). Constraint-induced movement therapy in the treatment of the upper limb in children with hemiplegic cerebral palsy. The Cochrane Library, Cochrane Database of Systematic Reviews, Art. No cd004149, issue 2.

116. Holt, O. J., Gallo, F., and Griffiths G. M., (2006). Regulating secretory lysosomes. Journal of Biochem (Tokyo), 140(1), p.7–12.

117. Höng, J. C., Ivanov, N. V., Hodor, P., Xia, M., Wei, N., Blevins, R., Gerhold, D., Borodovsky, M. and Liu, Y., (2004). Identification of new human cadherin genes using a combination of protein motif search and gene finding methods. Journal of Molecular Biology, 337(2), p.307-317.

118. Hovnanian, A., Rochat, A., Bodemer, C., Petit, E., Rivers, C. A., Prost, C., Fraitag, S., Christiano, A. M., Uitto, J., Lathrop, M., Barrandon, Y., and de Prost, Y., (1997). Characterization of 18 new mutations in COL7A1 in recessive dystrophic epidermolysis bullosa provides evidence for distinct molecular mechanisms underlying defective anchoring fibril formation. American Journal of Human Genetic, 61(3), p.599–610.

119. Hsiao, C. C., Chu, T. Y., Wu, C. J., van den Biggelaar, M., Pabst, C., Hébert, J., Kuijpers, T. W., Scicluna, B. P., Chen, T. C., Liebscher, I. and Hamann, J., (2018). The adhesion G protein-coupled receptor GPR97/ADGRG3 is expressed in human granulocytes and triggers antimicrobial effector functions. Frontiers in Immunology, 9(December), p.2830.

120. Huang, Y., Lin, L., Shanker, A., Malhotra, A., Yang, L., Dikov, M. M., et al. (2011). Resuscitating cancer immunosurveillance: selective stimulation of DLL1-Notch signaling in T cells rescues T-cell function and inhibits tumor growth. Cancer Res, 71(19), p.6122–31.

121. Huang, L., Chardon, J. W., Carter, M. T., Friend, K. L., Dudding, T. E., Schwartzentruber, J., Zou, R., Schofield, P. W., Douglas, S., Bulman, D.E. and Boycott, K. M., (2012). Missense mutations in ITPR1 cause autosomal dominant congenital nonprogressive spinocerebellar ataxia. Orphanet Journal of Rare Diseases, 7(1), p.67.

122. Hunter, J. M., Ahearn, M. E., Balak, C. D., Liang, W. S., Kurdoglu, A., Corneveaux, J. J., Russell, M., Huentelman, M. J., Craig, D. W., Carpten, J. and Coons, S. W., (2015). Novel pathogenic variants and genes for myopathies identified by whole exome sequencing. Molecular Genetics & Genomic Medicine, 3(4), p.283-301.

123. Hyde, T. M., Lipska, B. K., Ali, T., Mathew, S. V., Law, A. J., Metitiri, O. E., andKleinman, J. E., (2011). Expression of GABA Signaling Molecules KCC2, NKCC1, and GAD1 in Cortical Development and Schizophrenia. Journal of Neuroscience, 31(30), p.11088–11095.

124. Illingworth, R. S., (1966). The diagnosis of cerebral palsy in the first year of life. Developmental Medicine & Child Neurology, 8(2), p.178-194.

125. Imms, C., (2011). Bracing and splinting interventions in the upper limbs of people with cerebral palsy. Developmental medicine and Child Neurology, 53(4), p.293.

126. Ingham, P. W., (1998). Transducing Hedgehog: the story so far. The EMBO Journal, 17(13), p.3505-3511.

127. Introne, W., Boissy, R. E., and Gahl, W. A., (1999). Clinical, molecular, and cell biological aspects of Chediak-Higashi syndrome. Mol Genet Metab, 68(2), p.283–303.

128. Ishii, N., Recke, A., Mihai, S., Hirose, M., Hashimoto, T., Zillikens, D., and Ludwig, R. J., (2011). Autoantibody-induce intestinal inflammation and weight loss in experimental epidermolysis bullosa acquisita. Journal Pathol., 22(2), p.234–244.

129. Izuora, G. I., and Okora, A. B. O., (1981). Some clinical aspects of cerebral palsy among Nigerian Ibo children. Centr Afr Journal Med, 27(8), p.155- 159.

130. Jacobs, S. E., Berg, M., Hunt, R., Tarnow-Mordi, W. O., Inder, T. E., and Davis, P. G. (2013). Cooling for newborns with hypoxic ischaemic encephalopathy. The Cochrane Library, Cochrane Database of Systematic Reviews, Art. No.: CD003311, issue 1, p.1-112.

131. Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., Beckmann, J. S., Rosenfeld, J. A. and Eichler, E. E., (2014). A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. The American Journal of Human Genetics, 94(3), p.415-425.

132. Jaleco, A. C., Neves, H., Hooijberg, E., Gameiro, P., Clode, N., Haury, M., Henrique, D., and Parreira, L., (2001). Differential effects of Notch ligands Delta-1 and Jagged-1 in human lymphoid differentiation. Journal of Exp Med., 194(7), p.991–1002.

133. Jamieson, S. E., Miller, E. N., Peacock, C. S., Fakiola, M., Wilson, M. E., Bales-Holst, A., Shaw, M. A., Silveira, F., Shaw, J. J., Jeronimo, S. M., and Blackwell, J. M., (2007). Genome-wide scan for visceral leishmaniasis susceptibility genes in Brazil. Genes Immun., 8(1), p.84–90.

134. Jamra, R. A., Philippe, O., Raas-Rothschild, A., Eck, S. H., Graf, E., Buchert, R., Borck, G., Ekici, A., Brockschmidt, F. F., Nöthen, M. M. and Munnich, A., (2011). Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. The American Journal of Human Genetics, 88(6), p.788-795.

135. Jan, M. M., (2006). Cerebral palsy: comprehensive review and update. Annals of Saudi Medicine, 26(2), p.123–132.

136. Jarvis, S., Glinianaia, S. V., Arnaud, C., Fauconnier, J., Johnson, A., McManus, V., Topp, M., Uvebrant, P., Cans, C., and Krägeloh-Mann, I., (2005). Case gender and severity in cerebral palsy varies with intrauterine growth. Arch Dis Child 90(5), p.474–479.

137. Johnson, J. A., Tough, S., Wilson, R. D., Audibert, F., Blight, C., BrockS, J. A., and Okun, N., (2012). Delayed Child-Bearing. Journal of Obstetrics and Gynaecology Canada, 34(1), p.80–93.

138. Jones, M. W., Morgan, E., Shelton, J. E. and Thorogood, C., (2007). Cerebral palsy: introduction and diagnosis (part I). Journal of Pediatric Health Care, 21(3), p.146-152.

139. Jung, S. B., and Kevin, G., (2013). C-Terminal Binding Protein: A Molecular Link between Metabolic Imbalance and Epigenetic Regulation in Breast Cancer. International Journal of Cell Biology, 2013(Article ID: 647975), p.1-14.

140. Kaminker, J. S., Zhang, Y., Watanabe, C., and Zhang, Z., (2007). Can Predict: a computational tool for predicting cancer-associated missense mutations. Nucleic Acids Research, 35(suppl_2), p.W595-W598.

141. Kanaumi, T., Takashima, S., Iwasaki, H., Itoh, M., Mitsudome, A., and Hirose, S. (2008). Developmental changes in KCNQ2 and KCNQ3 expression in human brain: possible contribution to the age-dependent etiology of benign familial neonatal convulsions. Brain and Development, 30(5), p.362-369.

142. Kaplan, J., De Domenico, I., and Ward, D. M., (2008). Chediak-Higashi syndrome. Curr. Opin. Hematol. 15(1065-6251), p.22–29.

143. Kariya, N., Shimamura, Y., and Itom, (2005). Size polymorphisms in the human ultrahigh sulfur hair keratin-associated protein 4, KAP4, gene family. Journal of Invest Dermatol, 124(6), p.1111–1118.

144. Karki, R., Pandya, D., Elston, R.C. and Ferlini, C., (2015). Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Medical Genomics, 8(1), p.37.

145. Kashuba, V. I., Li, J., Wang, F., Senchenko, V. N., Protopopov, A., Malyukova, A., Kutsenko, A. S., Kadyrova, E., Zabarovska V. I., Muravenko, O. V., et al. (2004). RBSP3 (HYA22) is a tumor suppressor gene implicated in major epithelial malignancies. Proc. Natl. Acad. Sci. U.S.A., 101(14), p.4906–4911.

146. Katsanis, N., and Fisher, E. M., (1998) . A Novel C-Terminal Binding Protein (CTBP2) Is Closely Related to CTBP1, an Adenovirus E1A-Binding Protein, and Maps to Human Chromosome 21q21. 3. Genomics, 47(2), p.294-299.

147. Kawai, M., Kawahara, H., Hirayama, S., Yoshimura, N., Ida, S., (2004). Effect of baclofen on emesis and 24-hour esophageal pH in neurologically impaired children with gastroesophageal reflux disease. Journal of Pediatr Gastroenterol Nutr, 38(3), p.317-23.

148. Kelliher, M. A., and Roderick, J. E., (2018). NOTCH signaling in T-cell-mediated anti-tumor immunity and T-cell-based immunotherapies. Frontiers in Immunology, 9(1718), p.1-6.

149. Khandaker, G., Muhit, M., Karim, T., Smithers-Sheedy, H., Novak, I., Jones, C. and Badawi, N., (2019). Epidemiology of cerebral palsy in Bangladesh: a population-based surveillance study. Developmental Medicine & Child Neurology, 61(5), p.601-609.

150. Khankhanian, P., Baranzini, S. E., Johnson, B. A. et al. (2013). Sequencing of the IL6 gene in a case–control study of cerebral palsy in children. BMC Med Genet 14(126), p.1-8.

151. Kilpatrick, D. C., Starrs, L., Moore, S., Souter, V. and Liston, W. A., (1999). Mannan binding lectin concentration and risk of miscarriage. Human Reproduction, 14(9), p.2379-2380.

152. Kim, W. K., Park, M., Park, M., Kim, Y. J., Shin, N., Kim, H. K., and Kim, H. (2013). Identification and selective degradation of neopeptide-containing truncated mutant proteins in the tumors with high microsatellite instability. Clinical Cancer Research, 19(13), p.3369-3382.

153. Kimberlin, D., (2004). Herpes simplex virus, meningitis and encephalitis in neonates. Herpes: The Journal of The IHMF, 11(Jun), p.65A.

154. Kinney, H. C., (2006). The near-term (Late Preterm) human brain and risk for periventricular leukomalacia: a review. Semin Perinatol, 30(2), p.81–8.

155. Kirby, R. S., Wingate, M. S., Van Naarden Braun K., et al., (2011). Prevalence and functioning of children with cerebral palsy in four areas of the United States in 2006: a report from the Autism and Developmental Disabilities Monitoring Network. Res Dev Disabil, 32(2), p.462–69.

156. Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J., (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics, 46(3), p.310.

157. Klein, C. S., Zhou, P., & Marciniak, C. (2015). Excitability properties of motor axons in adults with cerebral palsy. Frontiers in Human Neuroscience, 9 (329), p.1-13.

158. Kloet, D. E., Polderman, P. E., Eijkelenboom, A., Smits, L. M., van Triest, M. H., van den Berg, M. C., and Burgering, B. M. (2015). FOXO target gene CTDSP2 regulates cell cycle progression through Ras and p21 (Cip1/Waf1). The Biochemical Journal, 469(2), p.289–298.

159. Knezević-Pogancev, M., (2010). Cerebral palsy and epilepsy. Medicinski Pregled, 63(7-8), p.527-530.

160. Knutson, L. M., and Clark, D. E., (1991). Orthotic Devices for Ambulation in Children with Cerebral Palsy and Myelomeningocele. Physical Therapy, 71(12), p.947–960.

161. Koh, T. H., Aynsley-Green, A., Tarbit, M., et al. (1988). Neural dysfunction during hypoglycaemia. Arch Dis Child, 63(11), p.1353-1358.

162. Koivisto, M., Blanco-Sequeiros, M., and Krause, U., (1972). Neonatal symptomatic and asymptomatic hypoglycaemia: a follow-up study of 151 children. Dev Med Child Neurol, 14(5), p.603–614.

163. Kotarski, M. A., Leonard, D. A., Bennett, S. A., Bishop, C. P., Wahn, S. D., Sedore, S. A., and Shrader, M. (1998). The Drosophila gene asteroid encodes a novel protein and displays dosage-sensitive interactions with Star and Egfr. Genome, 41(2), p.295–302.

164. Koterazawa, K., Okada, Y., Miyata, H., (2016). Incidence of cerebral palsy in Himeji over a 25-year period. No To Hattatsu= Brain and Development, 48(1), p.14–9.

165. Kruer, M. C., Jepperson, T., Dutta, S., Steiner, R. D., Cottenie, E., Sanford, L., Merkens, M., Russman, B. S., Blasco, P. A., Fan, G. and Pollock, J., (2013). Mutations in gamma adducin are associated with inherited cerebral palsy. Annals of neurology, 74(6), p.805-814.

166. Kulecka, M., Habior, A., Paziewska, A., Goryca, K., Dąbrowska, M., Ambrozkiewicz, F., Walewska-Zielecka, B., Gabriel, A., Mikula, M. and Ostrowski, J., (2017). Clinical applicability of whole-exome sequencing exemplified by a study in young adults with the advanced cryptogenic cholestatic liver diseases. Gastroenterology Research and Practice, 2017(May), p.1-8.

167. Kuo, H. J., Maslen, C. L., Keene, D. R. and Glanville, R. W. (1997). Type VI collagen anchors endothelial basement membranes by interacting with type IV collagen. Journal of Biol. Chem. 272(42), p.26522-26529.

168. Kwong, K. L, Wong, S. N., and So, K. T., (1998). Epilepsy in children with cerebral palsy. Pediatr Neurol, 19(1), p.31–6.

169. Lafrenière, R. G., Cader, M. Z., Poulin, J. F., Andres-Enguix, I., Simoneau, M., Gupta, N., Boisvert, K., Lafrenière, F., McLaughlan, S., Dubé, M.P. and Marcinkiewicz, M. M., (2010). A dominant-negative mutation in the TRESK potassium channel is linked to familial migraine with aura. Nature Medicine, 16(10), p.1157.

170. Laisram, N., Srivastava, V. K., and Srivastava, R. K., (1992). Cerebral palsy—An etiological study. The Indian Journal of Pediatrics, 59(6), p.723–72.

171. Lang, T. C., Fuentes-Afflick, E., Gilbert, W. M., Newman, T. B., Xing, G. and Wu, Y. W., (2012). Cerebral palsy among Asian ethnic subgroups. Pediatrics, 129(4), p.e992-e998.

172. Lataste, X., Emre, M., Davis, C. and Groves, L., (1994). Comparative profile of tizanidine in the management of spasticity. PubMed.gov, 44(11 Suppl 9), p.S53-9.

173. Le Pabic, P., Ng, C., and Schilling, T. F., (2014). Fat-Dachsous signaling coordinates cartilage differentiation and polarity during craniofacial development. PLoS Genet. 10(10), p.e1004726

174. Lee, J. E., and Gleeson, J. G. (2011). Cilia in the nervous system: linking cilia function and neurodevelopmental disorders. Current opinion in Neurology, 24(2), p.98.

175. Lefterova, M. I., and Lazar, M. A., (2009). New developments in adipogenesis. Trends Endocrinol. Metab., 20(3), p.107-114.

176. Lerer, I., Sagi, M., Meiner, V., Cohen, T., Zlotogora, J. and Abeliovich, D., (2005). Deletion of the ANKRD15 gene at 9p24.3 causes parent-of-origin-dependent inheritance of familial cerebral palsy. Human Molecular Genetics, 14(24), p.3911-3920.

177. Lewis, T. B., Leach, R. J., Ward, K., O'Connell, P., and Ryan, S. G. (1993). Genetic heterogeneity in benign familial neonatal convulsions: identification

of a new locus on chromosome 8q. American Journal of Human Genetics, 53(3), p.670.

178. Li, L., Liu, C., Amato, R. J., Chang, J. T., Du, G., and Li, W., (2014). CDKL2 promotes epithelial-mesenchymal transition and breast cancer progression. Oncotarget, 5(21), p.10840.

179. Lindemann, C. B., and Lesich, K. A., (2010). Flagellar and ciliary beating: the proven and the possible. Journal of Cell Sci, 123(4), p.519–528.

180. Lingenfelter, P. A., Delbridge, M. L., Thomas, S., Hoekstra, H. E., Mitchell, M. J., Graves, J. A. M., and Disteche, C. M., (2001). Expression and conservation of processed copies of the RBMX gene. Mammalian Genome, 12(7), p.538-545.

181. Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Akha, E. S., Schnekenberg, R. P., Suminaite, D., Hope, J., Baker, I., Gregory, L. and Green, A., (2012). Recessive mutations in SPTBN2 implicate β-III spectrin in both cognitive and motor development. PLoS Genetics, 8(12), p.e1003074.

182. Liu, J. M., Li, S., Lin, Q., and Li, Z., (1999). Prevalence of cerebral palsy in China., International Journal of Epidemiology, 28(5), p.949–954.

183. Lucas, A., Morley, R. and Cole, T. J., (1988). Adverse neurodevelopmental outcome of moderate neonatal hypoglycaemia. British Medical Journal, 297(6659), p.1304-1308.

184. Ludwig, R. J., and Zillikens, D., (2011). Pathogenesis of epidermolysis bullosa acquisita. Dermatol. Clin., 29 (issue 2011), p.493–501.

185. Lynex, C. N., Carr, I. M., Leek, J. P., Achuthan, R., Mitchell, S., Maher, E. R., Woods, C. G., Bonthon, D. T. and Markham, A. F., (2004). Homozygosity for a missense mutation in the 67 kDa isoform of glutamate decarboxylase in a family with autosomal recessive spastic cerebral palsy: parallels with Stiff-Person Syndrome and other movement disorders. BMC Neurology, 4(1), p.20.

186. MacDonald, P. E., and Rorsman, P., (2006). Oscillations, intercellular coupling, and insulin secretion in pancreatic β cells. PLoS Biology, 4(2), p.e49.

187. Maekawa, Y., Tsukumo, S., Chiba, S., et al., (2003). Delta1-Notch3 interactions bias the functional differentiation of activated CD41 T cells. Immunity, 19(4), p.549–59.

188. Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., and Jabado, N.n (2011). What can exome sequencing do for you? Journal of Med Genet, 48(9): p.580–589.

189. Maljevic, S., and Lerche, H., (2013). Potassium channels: a review of broadening therapeutic possibilities for neurological diseases. Journal of Neurol., 260(9), p.2201–2211.

190. Marrosu, M. G. et al. (2001). Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. Hum. Mol. Genet. 10(25), p.2907–2916.

191. Martin, D. (2000). Structural features and regulatory properties of the brain glutamate decarboxylases. Neurochemistry International, 37(2-3), p.111–119.

192. Massé, M., Cserhalmi-Friedman, P. B., Falanga, V., Celebi, J. T., Martinez-Mir, A., and Christiano, A. M., (2005). Identification of novel type VII collagen gene mutations resulting in severe recessive dystrophic epidermolysis bullosa. Clin. Exp. Dermatol., 30(3), p.289–293.

193. Matchkov, V. V., Aalkjaer, C., and Nilsson, H., (2004). A cyclic GMP-dependent calcium-activated chloride current in smooth-muscle cells from rat mesenteric resistance arteries. Journal of Gen Physiol., 123(2), p.121–134.

194. Matchkov, V. V., Larsen, P., Bouzinova, E. V., Rojek, A., Boedtkjer, D. M. B., Golubinskaya, V., and Nilsson, H. (2008). Bestrophin-3 (vitelliform macular dystrophy 2–like 3 protein) is essential for the cGMP-dependent calcium-activated chloride conductance in vascular smooth muscle cells. Circulation Research, 103(8), p.864-872.

195. Matsuda, S. and Yuzaki, M., (2009). Polarized sorting of AMPA receptors to the somatodendritic domain is regulated by adaptor protein AP-4. Neuroscience Research, 65(1), p.1-5.

196. May-Simera, H. L., and Kelley, M. W. (2012). Cilia, Wnt signaling, and the cytoskeleton. Cilia, 1(1), p.7.

197. McCombie, W.R., McPherson, J.D. and Mardis, E.R., (2019). Next-generation sequencing technologies. Cold Spring Harbor Perspectives in Medicine, 9(11), p.a036798.

198. McGinniss, M.J. and Kaback, M.M., (2013). Heterozygote testing and carrier screening. In Emery and Rimoin's Principles and Practice of Medical Genetics (pp. 1-10). Academic Press.

199. McHale, D. P., Jackson, A. P., Campbell, D. A., Levene, M. I., Corry, P., Woods, C. G., Lench, N. J., Mueller, R. F. and Markham, A. F., (2000). A gene for ataxic cerebral palsy maps to chromosome 9p12–q12. European Journal of Human Genetics, 8(4), p.267-272.

200. McIntyre, S., Taitz, D., Keogh, J., Goldsmith, S., Badawi, N. and Blair, E. V. E., (2013). A systematic review of risk factors for cerebral palsy in children born at term in developed countries. Developmental Medicine & Child Neurology, 55(6), p.499-508.

201. McMichael, G., Bainbridge, M. N., Haan, E., Corbett, M., Gardner, A., Thompson, S., Van Bon, B. W. M., Van Eyk, C. L., Broadbent, J., Reynolds, C. and O'callaghan, M. E., (2015). Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. Molecular Psychiatry, 20(2), p.176-182.

202. Mehrotra, S., Fakiola, M., Mishra, A., Sudarshan, M., Tiwary, P., Rani, D.S., Thangaraj, K., Rai, M., Sundar, S. and Blackwell, J. M., (2012). Genetic and functional evaluation of the role of DLL1 in susceptibility to visceral leishmaniasis in India. Infection, Genetics and Evolution, 12(6), p.1195-1201.

203. Menon, R. P., Strom, M., and Hughes, R. C. (2000). Interaction of a novel cysteine and histidine-rich cytoplasmic protein with galectin-3 in a carbohydrate-independent manner. FEBS Letters, 470(3), p.227-231.

204. Miceli, F., Striano, P., Soldovieri, M.V., Fontana, A., Nardello, R., Robbiano, A., Bellini, G., Elia, M., Zara, F., Taglialatela, M. and Mangano, S., (2015). A novel KCNQ3 mutation in familial epilepsy with focal seizures and intellectual disability. Epilepsia, 56(2), p.e15-e20.

205. Milanesi, L., Petrillo, M., Sepe, L., Boccia, A., D'Agostino, N., Passamano, M., Di Nardo, S., Tasco, G., Casadio, R. and Paolella, G., (2005). Systematic analysis of human kinase genes: a large number of genes and alternative splicing events result in functional and structural diversity. BMC Bioinformatics, 6(S4), p.S20.

206. Milisav, I., and Affara, N. A., (1998). A potential human axonemal dynein heavy-chain gene maps to 17q25. Mammalian Genome 9 (January), p.404-407.

207. Miller, E. N., Fadl, M., Mohamed, H. S., El Zein, A., Jamieson, S. E., Cordell, H. J., Peacock, C. S., Fakiola, M., Raju, M., Khalil, E. A., El Hassan, A. M., Ibrahim, M. E., Blackwell, J. M., (2007). Y chromosome lineage- and village-specific genes on chromosomes 1p22 and 6q27 that control visceral leishmaniasis in The Sudan. PLoS Genet., 3(5), p.e71.

208. Montassir, H., Maegaki, Y., Ogura, K., Kurozawa, Y., Nagata, I., Kanzaki, S. and Ohno, K., (2009). Associated factors in neonatal hypoglycemic brain injury. Brain and Development, 31(9), p.649-656.

209. Moreno-De-Luca, A., Helmers, S. L., Mao, H., Burns, T. G., Melton, A. M., Schmidt, K. R., Fernhoff, P. M., Ledbetter, D. H. and Martin, C. L., (2011). Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. Journal of Medical Genetics, 48(2), p.141-144.

210. Moreno-De-Luca, A., Ledbetter, D. H. and Martin, C. L., (2012). Genetic insights into the causes and classification of the cerebral palsies. The Lancet Neurology, 11(3), p.283-292.

211. Morris, C., (2002). A review of the efficacy of lower-limb orthoses used for cerebral palsy. Developmental Medicine and Child Neurology, 44(03), p.205.

212. Mukhopadhyay, S., and Rohatgi, R. (2014). G-protein-coupled receptors, Hedgehog signaling and primary cilia. Semin. Cell Dev. Biol. 33(September), 63–72. Academic Press.

213. Murphy, D. J., Johnson, A. M., Sellers, S. and MacKenzie, I. Z., (1995). Case-control study of antenatal and intrapartum risk factors for cerebral palsy in very preterm singleton babies. The Lancet, 346(8988), p.1449-1454.

214. Nagafuchi, Y., Shoda, H., Sumitomo, S., Nakachi, S., Kato, R., Tsuchida, Y., and Kanda, H. (2016). Immunophenotyping of rheumatoid arthritis reveals a linkage between HLA-DRB1 genotype, CXCR4 expression on memory CD4+ T cells, and disease activity. Scientific Reports, 6(July), p.29338.

215. Nagase, T., Kikuno, R., Nakayama, M., Hirosawa, M. and Ohara, O., (2000). Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Research, 7(4), p.271-281.

216. Nagase, T., Seki, N., Ishikawa, K.I., Tanaka, A. and Nomura, N., (1996). Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1. DNA Research, 3(1), p.17-24.

217. Nasim, M. T., Chernova, T. K., Chowdhury, H. M., Yue, B. G., and Eperon, I. C., (2003). HnRNP G and Tra2beta: opposite effects on splicing matched by antagonism in RNA binding. Hum Mol Genet, 12(11), p.1337–1348.

218. Neer, E. J., Schmidt, C.J., Nambudripad, R., and Smith, T.F., (1994). The ancient regulatory-protein family of WD-repeat proteins. Nature, 371(6495), p.297-300.

219. Nelson, K. B., Lynch, J. K., (2004). Stroke in newborn infants. Lancet Neurol, 3(3), p.150-158.

220. Nelson, K. B., Dambrosia, J. M., Grether, J. K., and Phillips, T. M., (1998). Neonatal cytokines and coagulation factors in children with cerebral palsy. Annals of Neurology. Official Journal of the American Neurological Association and the Child Neurology Society, 44(4), p.665-675.

221. Nelson, K. B., and Ellenberg, J. H., (1985). Antecedents of cerebral palsy: I. Univariate analysis of risks. American Journal of Diseases of Children, 139(10), p.1031-1038.

222. Ng, P. C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. Genome Research, 11(5), p. 863-874.

223. Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., and Lee, C., (2010a). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature Genetics, 42(9), p.790.

224. Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A. and Shendure, J., (2010b). Exome sequencing identifies the cause of a mendelian disorder. Nature Genetics, 42(1), p.30.

225. Nirschl, J. J., Ghiretti, A. E. and Holzbaur, E. L., (2017). The impact of cytoskeletal organization on the local regulation of neuronal transport. Nature Reviews Neuroscience, 18(10), p.585.

226. Niwa, S., Nakajima, K.,Miki, H.,Minato, Y.,Wang, D., and Hirokawa, N. (2012). KIF19A is a microtubule-depolymerizing kinesin for ciliary length control. Dev. Cell, 23(6), p.1167–1175.

227. Nottke, A., Colaia´ covo, M.P., and Shi, Y. (2009). Developmental roles of the histone lysine demethylases. Development, 136(6), 879–889.

228. O'Callaghan, M., and MacLennan, A. (2013). Cesarean Delivery and Cerebral Palsy. Obstetrics & Gynecology, 122(6), p.1169–1175.

229. O'Sullivan, M. C., Miller, S., Ramesh, V., Conway, E., Gilfillan, K., Mcdonough, S., and Eyre, J. A., (1998). Abnormal development of biceps brachii phasic stretch reflex and persistence of short latency heteronymous reflexes from biceps to triceps brachii in spastic cerebral palsy. Brain, 121(Pt12), p.2381–2395.

230. O'Brien, T. J., Beard, J. B., Underwood, L. J., Dennis, R. A., Santin, A. D., York, L., (2001). The CA 125 gene: an extracellular superstructure dominated by repeat sequences. Tumour Biol., 22(6), p.348-366.

231. O'Callaghan, M. E., MacLennan, A. H., Gibson, C. S., McMichael, G. L., Haan, E. A., Broadbent, J. L., and Dekker, G. A. (2011). Epidemiologic Associations With Cerebral Palsy. Obstetrics & Gynecology, 118(3), p.576–582.

232. Oda, K., Matsuoka, Y., Funahashi, A. and Kitano, H., (2005). A comprehensive pathway map of epidermal growth factor receptor signaling. Molecular Systems Biology, 1(1), p.2005-0010.

233. Odani, T., Yasuda, S., Ota, Y., Fujieda, Y., Kon, Y., Horita, T., and Koike, T., (2012). Up-regulated expression of HLA-DRB5 transcripts and high frequency of the HLA-DRB5* 01: 05 allele in scleroderma patients with interstitial lung disease. Rheumatology, 51(10), p.1765-1774.

234. Odding, E., Roebroeck, M. E., Stam, H. J., (2006). The epidemiology of cerebral palsy: incidence, impairments and risk factors. Disabil Rehabil, 28(4), p.183–91.

235. Oh, E.C., and Katsanis, N., (2012). Cilia in vertebrate development and disease. Development, 139(3), p.443–448.

236. Ohgiya, Y., Nobusawa, H., Seino, N., Miyagami, O., Yagi, N., Hiroto, S., Munechika, J., Hirose, M., Takeyama, N., Ohike, N. and Matsuoka, R., (2016). MR imaging of fetuses to evaluate placental insufficiency. Magnetic Resonance in Medical Sciences, 15(2), p.212-219.

237. Ombrello, M. J., Remmers, E. F., Tachmazidou, I., Grom, A., Foell, D., Haas, J.P., and Zeft, A. S. (2015). HLA-DRB1* 11 and variants of the MHC class II locus are strong risk factors for systemic juvenile idiopathic arthritis. Proceedings of the National Academy of Sciences, 112(52), p.15970-15975.

238. Omura, Y., Nishio, Y., Takemoto, T., Ikeuchi, C., Sekine, O., Morino, K., Maeno, Y., Obata, T., Ugi, S., Maegawa, H. and Kimura, H., (2009). SAFB1, an RBMX-binding protein, is a newly identified regulator of hepatic SREBP-1c gene. Bmb Rep, 42(4), p.232-237.

239. Pakula, A.T., Braun, K.V.N. and Yeargin-Allsopp, M., (2009). Cerebral palsy: classification and epidemiology. Physical Medicine and Rehabilitation Clinics, 20(3), p.425-452.

240. Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E. and Galuppi, B., (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. Developmental Medicine & Child Neurology, 39(4), p.214-223.

241. Palmer L, Petterson B, Blair E, and Burton P., (1994). Family patterns of gestational age at delivery and growth in utero in moderate and severe cerebral palsy. Dev Med Child Neurol, 36(12), p.1108-19.

242. Parolin Schnekenberg, R., Perkins, E. M., Miller, J. W., Davies, W. I., D'Adamo, M. C., Pessia, M., Fawcett, K. A., Sims, D., Gillard, E., Hudspith, K. and Skehel, P., (2015). De novo point mutations in patients diagnosed with ataxic cerebral palsy. Brain, 138(7), p.1817-1832.

243. Parveen, S., (2018). Management and Treatment for Cerebral Palsy in Children's. Indian Journal of Pharmacy Practice, 11(2), p.104-109.

244. Paulus, W., Baur, I., Liszka, U., Drlicek, M., Leigh, I. and Bruckner-Tuderman, L., (1995). Expression of type VII collagen, the major anchoring fibril component, in normal and neoplastic human nervous system. Virchows Archiv, 426(2), p.199-202.

245. Peifer, M., Berg, S., and Reynolds A. B., (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. Cell, 76(5), p.789 -791.

246. Perou, C. M., Leslie, J. D., Green, W., Li, L., McVey, W. D., and Kaplan, J., (1997).The beige/Chediak-Higashi syndrome gene encodes a widely expressed cytosolic protein. Journal of Biol Chem 272(47), p.29790-29794.

247. Petkov, G. V., (2011). Role of potassium ion channels in detrusor smooth muscle function and dysfunction. Nature Reviews Urology, 9(1), p.30–40.

248. Petterson, B., Stanley, F., and Henderson, D., (1990).Cerebral palsy in multiple births in Western Australia: genetic aspects. Am Journal of Med Genet, 37(3), p.346.

249. Pildes, R. S., Cornblath, M., Warren, I., Page-El, E., Di Menza, S., Merritt, D. M., et al., (1974). A prospective controlled study of neonatal hypoglycemia. Pediatrics, 54(1), p.5–14.

250. Pinder, R.M., Brogden, R.N., Speight, T. M., et al., (1977). Dantrolene Sodium. Drugs, 13(1), p.3–23.

251. Rabbani, B., Tekin, M., and Mahdieh, N., (2013). The promise of whole-exome sequencing in medical genetics. Journal of Hum Genet 59(1434-5161/13), p.5– 15.

252. Raju, U., Sondhi, V. and Patnaik, S. K., (2010). Meconium aspiration syndrome: an insight. Medical Journal Armed Forces India, 66(2), p.152-157.

253. Ramensky, V., Bork, P., and Sunyaev, S., (2002). Human non-synonymous SNPs: server and survey. Nucleic Acids Research, 30(17), p.3894-3900.

254. Ranatunga, K. W., (2011). Skeletal muscle stiffness and contracture in children with spastic cerebral palsy. The Journal of Physiology, 589(Pt 11), p.2665.

255. Rankin, J., Cans, C., Garne, E., Colver, A., Dolk, H., Uldall, P., Amar, E. and krageloh-mann, I., (2010). Congenital anomalies in children with cerebral palsy: a population-based record linkage study. Developmental Medicine & Child Neurology, 52(4), pp.345-351

256. Ravelli, A. M., Milla, P. J., (1998). Vomiting and gastroesophageal motor activity in children with disorders of the central nervous system. Journal of Pediatric Gastroenterology and Nutrition, 26(1), p.56–63.

257. Reddihough, D. S., King, J., Coleman, G., and Catanese, T. (2008b). Efficacy of programmes based on Conductive Education for young children with cerebral palsy. Developmental Medicine & Child Neurology, 40(11), p.763–770.

258. Reddihough, D., Johnson, H., Staples, M., Hudson, I., and Exarchos, H. (2008a). Use Of Benzhexol hydrochlorde to control drooling of children with cerebral palsy. Developmental Medicine & Child Neurology, 32(11), p.985–989.

259. Reid, E., Kloos, M., Ashley-Koch, A., Hughes, L., Bevan, S., Svenson, I. K., Graham, F. L., Gaskell, P. C., Dearlove, A., Pericak-Vance, M.A. and Rubinsztein, D. C., (2002). A kinesin heavy chain (KIF5A) mutation in hereditary spastic paraplegia (SPG10). The American Journal of Human Genetics, 71(5), p.1189-1194.

260. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E. and Voelkerding, K., (2015). 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), p.405-423.

261. Rogers, M. A., Langbein, L., Praetzel-Wunder, S., Winter, H., & Schweizer, J. (2006). Human Hair Keratin-Associated Proteins (KAPs). International Review of Cytology, 251(0074-7696/06), p.209–263.

262. Rogers, M. A., Langbein, L., Winter, H., Ehmann, C., Praetzel, S., and Schweizer, J. (2002). Characterization of a first domain of human high glycine-

tyrosine and high sulfur keratin-associated protein (KAP) geneson chromosome 21q22.1.]ournal of Biol. Chem. 277(20), p.48993–49002.

263. Rosenbaum, P. (2003). Cerebral palsy: what parents and doctors want to know. BMJ, 326(7396), p. 970–974.

264. Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M., Bax, M., Damiano, D., Dan, B. and Jacobsson, B., (2007). A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl, 109(suppl 109), p.8- 14.

265. Rubio, J.P., Bahlo, M., et al. (2002). Genetic dissection of the human leukocyte antigen region by use of haplotypes of Tasmanians with multiple sclerosis. Am. Journal of Hum. Genet. 70(5), p.1125–1137.

266. Rump, A., Morikawa, Y., Tanaka, M., Minami, S., Umesaki, N., Takeuchi, M., Miyajima, A., (2004). Binding of ovarian cancer antigen CA125/MUC16 to mesothelin mediates cell adhesion. Journal of Biol Chem., 279(10), p.9190-9198.

267. Rutecki, P. A. (1992). Neuronal excitability: voltage-dependent currents and synaptic transmission. Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society, 9(2), p.195-211.

268. Sabatelli, P., Gara, S.K., Grumati, P., Urciuolo, A., Gualandi, F., Curci, R., Squarzoni, S., Zamparelli, A., Martoni, E., Merlini, L. and Paulsson, M., (2011). Expression of the collagen VI α5 and α6 chains in normal human skin and in skin of patients with collagen VI-related myopathies. Journal of Investigative Dermatology, 131(1), p.99-107.

269. Salem, Y. S., Levy, A., Wiznitzer, A., Holcberg, G., Mazor, M., Sheiner, E., (2011). A significant linear association exists between advanced maternal age and adverse perinatal outcome. ArchGynecol Obstet, 283(4), p.755–9.

270. Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein,

J. L. and Walker, M. F., (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature, 485(7397), p.237-241.

271. Sanger, T. D., Chen, D., Fehlings, D. L., Hallett, M., Lang, A. E., Mink, J. W., Singer, H. S., Alter, K., Ben-Pazi, H., Butler, E. E. and Chen, R., (2010). Definition and classification of hyperkinetic movements in childhood. Movement Disorders, 25(11), p.1538-1549.

272. Sano, Y., Inamura, K., Miyake, A., Mochizuki, S., Kitada, C., Yokoi, H., Nozawa, K., Okada, H., Matsushime, H. and Furuichi, K., (2003). A novel two-pore domain K+ channel, TRESK, is localized in the spinal cord. Journal of Biological Chemistry, 278(30), p.27406-27412.

273. Seelow, D., Schuelke, M., Hildebrandt, F., and Nurnberg, P. (2009). HomozygosityMapper--an interactive approach to homozygosity mapping. Nucleic Acids Research, 37(Web Server), W593–W599.

274. Sepulveda, F. E., Burgess, A., Heiligenstein, X., Goudin, N., Ménager, M. M., Romao, M., and Ménasché, G., (2015). LYST controls the biogenesis of the endosomal compartment required for secretory lysosome function. Traffic, 16(2), p.191-203.

275. Sharifnia, P. and Jin, Y., (2015). Regulatory roles of RNA binding proteins in the nervous system of C. elegans. Frontiers in Molecular Neuroscience, 7(Jan), p.100.

276. Shevell, M., (2019). Cerebral palsy to cerebral palsy spectrum disorder: Time for a name change? Neurology, 92(5), p.233-235.

277. Shiflett, S. L., Kaplan, J. and Ward, D. M., (2002). Chediak–Higashi syndrome: a rare disorder of lysosomes and lysosome related organelles. Pigment Cell Research, 15(4), p.251-257.

278. Shimazaki, H., Honda, J., Naoi, T., Namekawa, M., Nakano, I., Yazaki, M., Nakamura, K., Yoshida, K., Ikeda, S.I., Ishiura, H. and Fukuda, Y., (2014). Autosomal-recessive complicated spastic paraplegia with a novel lysosomal

trafficking regulator gene mutation. Journal of Neurol Neurosurg Psychiatry, 85(9), p.1024-1028.

279. Shimol, S.B., Einhorn, M., and Greenberg, D., (2012). Listeria meningitis and ventriculitis in an immunocompetent child: case report and literature review. Infection, 40(2), p.207-211.

280. Shimomura, Y. and Ito, M., (2005). December. Human hair keratin-associated proteins. In Journal of Investigative Dermatology Symposium Proceedings, Elsevier, 10(3), p. 230-233.

281. Singh, N. A., Charlier, C., Stauffer, D., DuPont, B. R., Leach, R. J., Melis, R., and McHarg, M. L. (1998). A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns. Nature Genetics, 18(1), p.25.

282. Slaats, G. G., Ghosh, A. K., Falke, L. L., Le Corre, S., Shaltiel, I. A., van de Hoek, G., Klasson, T. D., Stokman, M. F., Logister, I., Verhaar, M. C. and Goldschmeding, R., (2014). Nephronophthisis-associated CEP164 regulates cell cycle progression, apoptosis and epithelial-to-mesenchymal transition. PLoS Genetics, 10(10), p.e1004594.

283. Sleckman, B. P., Khan, W. N., Xu, W., Bassing, C. H., Malynn, B. A., Copeland, N. G., Bardon, C. G., Breit, T. M., Davidson, L., Oltz, E. M., Jenkins, N. A., Berman, J. E., and Alt, F.W. (2000). Cloning and functional characterization of the early-lymphocyte-specific Pb99 gene. Mol. Cell.Biol. 20(12), p.4405–4410

284. Soderhall C, Marenholz I, Kerscher T et al. (2007). Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. PLoS Biol, 5(9), p.e242.

285. Söhle, J., Machuy, N., Smailbegovic, E., Holtzmann, U., Grönniger, E., Wenck, H., Stäb, F. and Winnefeld, M., (2012). Identification of new genes involved in human adipogenesis and fat storage. PloS One, 7(2), p.e31193

286. Song, H. Y., Tian, Y. M., Zhang, Y. M., Zhou, L., Lian, H., and Zhu, J. X. (2014). A novel finding of anoctamin 5 expression in the rodent gastrointestinal tract. Biochemical and Biophysical Research Communications, 451(2), p.258-262.

287. Soo, K. Y., Farg, M. and Atkin, J. D., (2011). Molecular motor proteins and amyotrophic lateral sclerosis. International Journal of Molecular Sciences, 12(12), p.9057-9082.

288. Soulard, M., Della V. V., and Larsen, C. J., (2002). Autoimmune antibodies to hnRNPG protein in dogs with systemic lupus erythematosus: epitope mapping of the antigen. Journal of Autoimmune, 18(3), p.221–229.

289. Soulard, M., Valle, V. D., Siomi, M. C., Pinol-Roma, S., Codogno, P., Bauvy, C., and Larsen, C. J. (1993). hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. Nucleic Acids Research, 21(18), p.4210-4217.

290. Spassky, N., Han, Y.G., Aguilar, A., et al. (2008). Primary cilia are required for cerebellar development and Shh-dependent expansion of progenitor pool. Dev Biol, 317(1), p.246—59.

291. Spritz, R. A., (1998). Genetic defects in Chediak-Higashi syndrome and the beige mouse. Journal of Clin Immunol; 18(2), p.97–105.

292. Staiano A, and Del Giudice E. (1994). Colonic transit and anorectal manometry in children with severe brain damage. Pediatrics 94(2), p.169–173.

293. Stelmach, T., Pisarev, H., and Talvik, T., (2005). Ante- and perinatal factors for cerebral palsy: case-control study in Estonia. Journal of Child Neurol, 20(8), p.654– 660.

294. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y. and Kaplan, S., (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. Current Protocols in Bioinformatics, 54(1), p.1-30.

295. Stojilkovic, S. S., Bjelobaba, I. and Zemkova, H., (2017). Ion channels of pituitary gonadotrophs and their roles in signaling and secretion. Frontiers in Endocrinology, 8(June), p.126.

296. Stoodley, C. J., and Schmahmann, J. D., (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. Neuroimage, 44(2), p.489-501.

297. Strand, K. M., Andersen, G. L., Haavaldsen, C,, Vik, T., and Eskild, A., (2016). Association of placental weight with cerebral palsy: population-based cohort study in Norway. BJOG; 123(December), p.2131–2138.

298. Su, J. and Wang, L., (2012). Research advances in neonatal hypoglycemic brain injury. Translational Pediatrics, 1(2), p.108.

299. Su, J. M., Tsamtsouris, A. and Laskou, M., (2003). Gastroesophageal reflux in children with cerebral palsy and its relationship to erosion of primary and permanent teeth. Journal of the Massachusetts Dental Society, 52(2), p.20-24.

300. Su, L. K., and Qi, Y. (2001). Characterization of human MAPRE genes and their proteins. Genomics, 71(2), p.142-149.

301. Sudo, S., Konno, Y., Odagiri, T., Kato, T., Hosaka, M., Takeda, M., and Sakuragi, N., (2011). Exploration of biomarkers for lymph node metastasis in patients with endometrial cancer using exon-expression microarray. Journal of Clinical Oncology, 29(15_suppl), p.5100-5100.

302. Sugimoto, K., Maekawa, Y., Kitamura, A., Nishida, J., Koyanagi, A., Yagita, H., et. al., (2010). Notch2 signaling is required for potent antitumor immunity in vivo. Journal Immunol, 184(April), p.4673–8.

303. Sullivan, K. D., Evans, D., Pandey, A., Hraha, T. H., Smith, K. P., Markham, N., Rachubinski, A. L., Wolter-Warmerdam, K., Hickey, F., Espinosa, J. M. and Blumenthal, T., (2017). Trisomy 21 causes changes in the circulating proteome indicative of chronic autoinflammation. Scientific Reports, 7(1), p.14818.

304. Sun, H., Xu, F., Xiong, H., et. al., (2013). Characteristics of respiratory distress syndrome in infants of different gestational ages. Lung, 191(4):425– 433.

305. Suzuki, Y., Toribe, Y., Mogami, Y., Yanagihara, K. and Nishikawa, M., (2008). Epilepsy in patients with congenital cytomegalovirus infection. Brain and Development, 30(6), p.420-424.

306. Swaminathan, S., Kim, S., Shen, L., Risacher, S. L., Foroud, T., Pankratz, N., Potkin, S. G., Huentelman, M. J., Craig, D. W., Weiner, M. W. and Saykin, A. J., (2010). Preliminary analysis of copy number variation in the ADNI cohort. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 6(4), p.S18.

307. Tagliavini, F., Pellegrini, C., Sardone, F., Squarzoni, S., Paulsson, M., Wagener, R., Gualandi, F., Trabanelli, C., Ferlini, A., Merlini, L. and Santi, S., (2014). Defective collagen VI α6 chain expression in the skeletal muscle of patients with collagen VI-related myopathies. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1842(9), p.1604-1612.

308. Taglienti, C. A., Wysk, M., and Davis, R. J. (1996). Molecular cloning of the epidermal growth factor-stimulated protein kinase p56 KKIAMRE. Oncogene, 13(12), p.2563-2574.

309. Tahmassebi, J., and Curzon, M. (2003). Prevalence of drooling in children with cerebral palsy attending special schools. Developmental Medicine & Child Neurology, 45(09), p.613-617.

310. Takeshita, K., Ando, Y., Ohtani, K., Takashima, S., (1989). Cerebral Palsy in Tottori, Japan-benefit and risks of progress in perinatal medicine. Neuroepidemiology, 8(4), p.184–192.

311. Tam V., Patel N., Turcotte M et. al., (2019). Benefits and Limitations of genome-wide association studies. Nat Rev Genet, 20, 467-484

312. Tamanini, F., Willemsen, R., van Unen, L., Bontekoe, C., Galjaard, H., Oostra, B. A., and Hoogeveen, A. T. (1997). Differential expression of FMR1, FXR1

and FXR2 proteins in human brain and testis. Human Molecular Genetics, 6(8), p.1315-1322.

313. Taub, E., Ramey, S. L., DeLuca, S., and Echols, K., (2004). Efficacy of Constraint-Induced Movement Therapy for Children With Cerebral Palsy With Asymmetric Motor Impairment. Pediatrics, 113(2), p.305–312.

314. Tee, J. M., and Peppelenbosch, M. P., (2010). Anchoring skeletal muscle development and disease: the role of ankyrin repeat domain containing proteins in muscle physiology. Critical Reviews in Biochemistry and Molecular Biology, 45(4), p.318-330.

315. Tekgul, H., Polat, M., Tosun, A., Serdaroglu, G., and Gokben, S., (2013). Electrophysiologic assessment of spasticity in children using H-reflex. Turk. Journal of .Pediatric. 55(5), p.519–523.

316. Tekin, F., Kavlak, E., Cavlak, U., and Altug, F., (2018). Effectiveness of Neuro-Developmental Treatment (Bobath Concept) on postural control and balance in Cerebral Palsied children. Journal of Back and Musculoskeletal Rehabilitation, 31(2), p.397–403.

317. Thorngren-Jerneck, K., and Herbst, A., (2006). Perinatal Factors Associated With Cerebral Palsy in Children Born in Sweden. Obstetrics & Gynecology, 108(6), p.1499–1505.

318. Thygesen, S. K., Olsen, M., Østergaard, J. R., et al., (2016). Respiratory distress syndrome in moderately late and late preterm infants and risk of cerebral palsy: a population-based cohort study. BMJ Open, 6(10), p.e011643.

319. Touyama, M.,and Touyama, J., (2008). Prevalence of cerebral palsy in Okinawa between 1995 and 2001. No To Hattatsu; 40(5), p.387–92.

320. Toyokawa, S., Maeda, E. and Kobayashi, Y., (2017). Estimation of the number of children with cerebral palsy using nationwide health insurance claims data in Japan. Developmental Medicine & Child Neurology, 59(3), p.317-321.

321. Tsend-Ayush, E., O'Sullivan, L.A., Grützner, F.S., Onnebo, S.M., Lewis, R.S., Delbridge, M.L., Marshall Graves, J.A. and Ward, A.C., (2005). RBMX gene is essential for brain development in zebrafish. Developmental Dynamics, 234(3), p.682-688.

322. Uitto J, and Richard G., (2005). Progress in epidermolysis bullosa: from eponyms to molecular genetic classification. Clin Dermatol., 23(1), p.33–40.

323. Vahidnezhad, H., Youssefian, L., Jazayeri, A. and Uitto, J., (2018). Research techniques made simple: genome-wide homozygosity/autozygosity mapping is a powerful tool for identifying candidate genes in autosomal recessive genetic diseases. Journal of Investigative Dermatology, 138(9), p.1893-1900.

324. Valtcheva, N., Primorac, A., Jurisic, G., Hollmén, M. and Detmar, M., (2013). The orphan adhesion G protein-coupled receptor GPR97 regulates migration of lymphatic endothelial cells via the small GTPases RhoA and Cdc42. Journal of Biological Chemistry, 288(50), p.35736-35748.

325. Van De Leemput, J., Chandran, J., Knight, M.A., Holtzclaw, L.A., Scholz, S., Cookson M.R., Houlden, H., Gwinn-Hardy, K., Fung, H.C., Lin, X. and Hernandez, D., (2007). Deletion at ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans. PLoS Genetics, 3(6), p.e108.

326. Van den Elsen P. J., Holling T. M., Kuipers H. F., and Van der Stoep N. (2004). Transcriptional regulation of antigen presentation. Curr. Opin. Immunol. 16(1), p.67–75.

327. Van Den Elsen, P. J., (2011). Expression regulation of major histocompatibility complex class I and class II encoding genes. Frontiers in Immunology, 2(Article 48), p.48.

328. Van Essen, D., Zhu, Y., and Saccani, S. (2010). A feed-forward circuit controlling inducible NF-kB target gene activation by promoter histone demethylation. Mol. Cell, 39(5), p.750–760.

329. Varki, R., Sadowski, S., Uitto, J. and Pfendner, E., (2007). Epidermolysis bullosa. II. Type VII collagen mutations and phenotype–genotype correlations in the dystrophic subtypes. Journal of Medical Genetics, 44(3), p.181-192.

330. Venables, J.P., Elliott, D.J., Makarova, O.V., Makarov, E.M., Cooke, H.J. and Eperon, I.C., (2000). RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2β and affect splicing. Human Molecular Genetics, 9(5), p.685-694.

331. Verkerk, A. J., Schot, R., Dumee, B., Schellekens, K., Swagemakers, S., Bertoli-Avella, A. M., Lequin, M. H., Dudink, J., Govaert, P., Van Zwol, A. L. and Hirst, J., (2009). Mutation in the AP4M1 gene provides a model for neuroaxonal injury in cerebral palsy. The American Journal of Human Genetics, 85(1), p.40-52.

332. Veugelers, R., Benninga, M. A., Calis, E. A., Willemsen, S. P., Evenhuis, H., Tibboel, D. and Penning, C., (2010). Prevalence and clinical presentation of constipation in children with severe generalized cerebral palsy. Developmental Medicine & Child Neurology, 52(9), p.e216-e221.

333. Vieira, R. N., Ávila, R., de Paula, J. J., Cintra, M. T. G., de Souza, R. P., Nicolato, R., Malloy-Diniz, L., de Miranda, D. M., de Moraes, E. N., de Marco, L. A. and Romano-Silva, M. A., 2016. Association between DCHS2 gene and mild cognitive impairment and Alzheimer's disease in an elderly Brazilian sample. International Journal of Geriatric Psychiatry, 31(12), p.1337- 1344.

334. Voronin, D. A., and Kiseleva, E. V., (2008). Functional role of proteins containing ankyrin repeats. Cell and Tissue Biology, 2(1), p.1–12.

335. Walstab, J. E., Bell, R. J., Reddihough, D. S., Brennecke, S. P., Bessell, C. K., and Beischer, N. A., (2004). Factors identified during the neonatal period associated with risk of cerebral palsy. Aust N Z Journal of Obstet Gynaecol 44(4), p.342–346.

336. Wang, Y., Cheon, D. J., Lu, Z., Cunningham, S. L., Chen, C. M., Luo, R. Z., Xing, D., Orsulic, S., Bast, R. C., and Behringer, R. R., (2008). MUC16 expression during embryogenesis, in adult tissues, and ovarian cancer in the mouse. Differentiation. 76(10), p.1081-1092.

337. Wang, J., Wang, X., Chen, X., Lu, S., Kuang, Y., Fei, J., and Wang, Z. (2018). Gpr97/Adgrg3 ameliorates experimental autoimmune encephalomyelitis by regulating cytokine expression. Acta Biochimica Et Biophysica Sinica, 50(7), p.666-675.

338. Wang, J. J., Zhang, L. L., Zhang, H. X., Shen, C. L., Lu, S. Y., Kuang, Y., Wan, Y. H., Wang, W. G., Yan, H. M., Dang, S. Y. and Fei, J., (2013). Gpr97 is essential for the follicular versus marginal zone B-lymphocyte fate decision. Cell Death & Disease, 4(10), p.e853.

339. Wang, J. L., Cao, L., Li, X. H., Hu, Z. M., Li, J. D., Zhang, J. G., Liang, Y., Li, N., Chen, S. Q., Guo, J. F. and Jiang, H., (2011). Identification of PRRT2 as the causative gene of paroxysmal kinesigenic dyskinesias. Brain, 134(12), p.3493-3501.

340. Wang, X., & Wang, Y. (2012). Gait analysis of children with spastic hemiplegic cerebral palsy. Neural Regeneration Research, 7(20), p.1578–1584.

341. Ward, D. M., Shiflett, S. L., Kaplan, J., (2002). Chediak-Higashi syndrome: a clinical and molecular view of a rare lysosomal storage disorder. Curr Mol Med, 2(5), p.469–477.

342. Waxman, S.G. and Zamponi, G.W., (2014). Regulating excitability of peripheral afferents: emerging ion channel targets. Nature neuroscience, 17(2), p.153.

343. Wheway, G., Nazlamova, L., and Hancock, J. T., (2018). Signaling through the primary cilium. Frontiers in Cell and Developmental Biology, 6(Feb), p.8.

344. Whitfield, M., Thomas, L., Bequignon, E., Schmitt, A., Stouvenel, L., Montantin, G., Tissier, S., Duquesnoy, P., Copin, B., Chantot, S. and Dastot, F., (2019). Mutations in DNAH17, encoding a sperm-specific axonemal outer dynein arm heavy chain, cause isolated male infertility due to asthenozoospermia. The American Journal of Human Genetics, 105(1), p.198-212.

345. Williams, M. C, and O'Brien, W. F., (1997). Cerebral palsy in infants with asymmetric growth restriction. Am Journal of Perinatol, 14(04), p.211-5.

346. Williams, S. P., Odell, A. F., Karnezis, T., Farnsworth, R. H., Gould, C. M., Li, J., Paquet-Fifield, S., Harris, N. C., Walter, A., Gregory, J. L. and Lamont, S. F., (2017). Genome-wide functional analysis reveals central signaling regulators of lymphatic endothelial cell migration and remodeling. Sci. Signal., 10(499), p.eaal2987.

347. Wiltgen, M. and Tilz, G.P., (2007). DNA microarray analysis: principles and clinical impact. Hematology, 12(4), pp.271-287.

348. Woehlke, G., Ruby, A. K., Hart, C. L., Ly, B., Hom-Booher, N. and Vale, R. D., (1997). Microtubule interaction site of the kinesin motor. Cell, 90(2), p.207-216.

349. Woodley, D. T., Remington, J. and Chen, M., (2007). Autoimmunity to type VII collagen: epidermolysis bullosa acquisita. Clinical Reviews in Allergy & Immunology, 33(1-2), p.78-84.

350. Wu, D., Zou, Y. F., Xu, X. Y., Feng, X. L., Yang, L., Zhang, G. C., Bu, X. S. and Tang, J. L., (2011). The Association of Genetic Polymorphisms with Cerebral Palsy: A Meta-analysis. Medicine & Child Neurology, 53(3), p.217-225.

351. Wu, L., Guo, S., Yang, D., Ma, Y., Ji, H., Chen, Y., and Liu, J. (2013). Copy number variations of HLA-DRB5 is associated with systemic lupus erythematosus risk in Chinese Han population. Acta Biochim Biophys Sin, 46(2), p.155-160.

352. Wu, Y. W., Croen, L. A., Torres, A. R., Van De Water, J., Grether, J. K. and Hsu, N. N., (2009). Interleukin-6 genotype and risk for cerebral palsy in term and near-term infants. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 66(5), p.663-670.

353. Xu, F., Wu, X., Jiang, L.-H., Zhao, H., and Pan, J. (2016). An organelle K+ channel is required for osmoregulation in Chlamydomonas reinhardtii. Journal of Cell Science, 129(15), p.3008–3014.

354. Xu, Y., Wang, H., Sun, Y., Shang, Q., Chen, M., Li, T., Zhu, D., He, L., Zhu,C. and Xing, Q., (2014). The association of apolipoprotein E gene polymorphisms with cerebral palsy in Chinese infants. Molecular Genetics and Genomics, 289(3), p.411-416.

355. Yager, J. Y., (2002). Hypoglycemia injury to the immature brain. Clin Perinatal, 29(4), p.651–74.

356. Yang, C., Wu, J., De Heus, C., Grigoriev, I., Liv, N., Yao, Y., Smal, I., Meijering, E., Klumperman, J., Qi, R. Z. and Akhmanova, A., (2017a). EB1 and EB3 regulate microtubule minus end organization and Golgi morphology. Journal of Cell Biology, 216(10), p.3179-3198.

357. Yang, X., Zheng, J., Tian, S., Chen, Y., An, R., Zhao, Q., and Xu, Y. (2017b). HLA-DRA/HLA-DRB5 polymorphism affects risk of sporadic ALS and survival in a southwest Chinese cohort. Journal of The Neurological Sciences, 373(Feb), p.124-128.

358. Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K. and Matsuoka, M., (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nature Genetics, 48(8), p.927.

359. Yap, C. C., Murate, M., Kishigami, S., Muto, Y., Kishida, H., Hashikawa, T. and Yano, R., (2003). Adaptor protein complex-4 (AP-4) is expressed in the central nervous system neurons and interacts with glutamate receptor δ2. Molecular and Cellular Neuroscience, 24(2), p.283-295.

360. Yeo, M., Lee, S. K., Lee, B., Ruiz, E. C., Pfaff, S. L. and Gill, G. N., (2005). Small CTD phosphatases function in silencing neuronal gene expression. Science, 307(5709), p.596-600.

361. Zachazewski, J. E., Eberle, E. D., Jefferies, M., (1982). Effect of tone-inhibiting casts and orthoses on gait. A case report. Phys Ther., 62(4), p.453-455.

362. Zaragoza, M., Nguyen, C., Widyastuti, H., McCarthy, L., and Grosberg, A., (2017). Dupuytren's and Ledderhose Diseases in a Family with LMNA-Related Cardiomyopathy and a Novel Variant in the ASTE1 Gene. Cells, 6(4), p.40.

363. Zhang, Y., O'Connor, J. P., Siomi, M. C., Srinivasan, S., Dutra, A., Nussbaum, R. L., and Dreyfuss, G. (1995). The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. The EMBO Journal, 14(21), p.5358-5366.

364. Zhu Y., Lu Y., Zhang Q., Liu J. J., Li T. J., Yang J. R., Zeng C., and Zhuang, S. M. (2012). MicroRNA-26a/b and their host genes cooperate to inhibit the G1/S transition by activating the pRb protein. Nucleic Acids Res., 40(10), p.4615–4625.

365. Zhu, M., and Zhao, S., (2007). Candidate gene identification approach: Progress and challenge. International Journal of Biol Sci, 3(7), p.420-7.

Books:

366. Brooker, R. J., (2012). Genetics analysis & principles. New York: McGraw Hill. P.195 & P.508.

367. Brown D.A. (2000) M-Current: From Discovery to Single Channel Currents. Tokyo: Springer. P.55-94.

368. Kumari, D. and Gazy, I. (2019). Towards Mechanism-Based Treatments for Fragile X Syndrome. Basel: MDPI. P.88.

369. Martin, E. A., and Hine, R., (2008). A dictionary of biology. Oxford: Oxford University Press. P.337.

370. Parvizi, J., (2010). High Yield Orthopaedics E-Book. Elsevier Health Sciences. Philadelphia: Saunders Elsevier. P.85-86.

371. Philpot P., Greenspan J., Aghai Z. H., (2018). Problems During Delivery as an Etiology of Cerebral Palsy in Full-Term Infants. In: Miller F., Bachrach S., Lennon N., O'Neil M. (eds) Cerebral Palsy. Springer, Cham. P.1-10.

372. Rapin, I., (2000). Static Disorders of Brain Development. Rowland, L.P. (Ed.), Merritt's Neurology, Edition 10. Baltimore: Lippincott, Williams & Wilkins. P. 476–484.

373. Stanley, F., Blair, E., and Alberman, E., (2000). Cerebral Palsies: Epidemiology and Casual Pathways. London: MacKeith Press. P.98 & P.108.

374. Volpe, J. J., (2008). Neurology of the newborn. Vol. Philadephia: Saunders Elsevier, P.130-134.

375. Yamada W. M., Koch C., and Adams P. R., (1998). Methods in Neuronal Modeling, C. Koch and Segev, Eds. Cambdrige: MIT Press. P. 97-133.

Websites with Authors:

376. Borchers, A. A., (2015). Handbook of Signs & Symptoms 2015 [Online]. [Assessed on 5th April 2020]. Available from World Wide Web: *https://studfile.net/preview/7605613/page:38/*

377. Carney, P. R., Geyer, J. D., Ringdahl, D., and Carolyn, C. (2007). Chapter 11. Cerebral Palsy [Online]. [Assessed on 5th April 2020]. Available from World WideWeb:*https://neurology.mhmedical.com/content.aspx?bookid=459&sectionid=41027 558&jumpsectionid=41030165&caclickthru=129824*

378. Chinnadurai, G., (2013). CtBP Family Proteins: Unique Transcriptional Regulators in the Nucleus with Diverse Cytosolic Functions. In: Madame Curie Bioscience Database [Online]. [Assessed on 1st December 2019]. Availablefrom World Wide Web: *https://www.ncbi.nlm.nih.gov/books/NBK6557/*

379. Lou, L. F., and Zhang, J., (2010). Current Status of Diagnosis and Treatment in Hypoglycemic Encephalopathy [Online]. [Assessed on 5th April 2020]. Available from World Wide Web: *http://en.cnki.com.cn/Article_en/CJFDTotal-YXZS201013031.htm*

380. Norton, N. S., (2007). Cerebral Palsy. xPharm: The Comprehensive Pharmacology Reference [Online]. [Assessed on 4th April 2020]. Available from World Wide Web: *https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/cerebral-palsy*

381. Shannon, S. G. (2012). The characterization of the tetratricopeptide repeat protein 13 in mammals [Online]. [Assessed on 15th November 2019]. Available from World Wide Web: *https://mospace.umsystem.edu/xmlui/handle/10355/43184*

Websites with No Author:

382. Biceps Muscle (2008) [Online], [Accessed on 6th April 2020]. Available from World Wide Web: *https://www.britannica.com/science/biceps-muscle*

383. Class SamToFastq (2018) [Online], [Accessed on 20th December 2018]. Available from World WideWeb: *https://javadoc.io/static/org.broadinstitute/gatk/4.1.4.1/picard/sam/SamToFastq.html*

384. Contractures in Cerebral Palsy (2019) [Online], [Accessed on 6th April 2020]. Available from World Wide Web: *https://en.hesperian.org/hhg/Disabled_Village_Children:Contractures_in_Cerebral_Palsy*

385. dbSNP (2017) [Online], [Accessed on 20th December 2018]. Available from World Wide Web: *http://www. ncbi.nlm.nih.gov/SNP/*

386. Disabled Village Children - CHAPTER 18 (Hip Problems) (2015) [Online], [Accessed on 6th April 2020]. Available from World Wide Web: *https://www.dinf.ne.jp/doc/english/global/david/dwe002/dwe00220.html*

387. GATK4 command-line syntax (2017) [Online], [Accessed on 20th December 2018]. Available from World Wide Web: *https://software.broadinstitute.org/gatk/documentation/article?id=11050*

388. Genome Analysis Toolkit (2018) [Online], [Accessed on 20th December 2018]. Available from World Wide Web: *https://software.broadinstitute.org/gatk/*

389. Mayo Foundation For Medical Education And Research Scoliosis (2020) [Online], [Accessed on 6th April 2020]. Available from World Wide Web: *https://www.mayoclinic.org/diseases-conditions/scoliosis/multimedia/scoliosis/img-20008437*

390. NextSeq 550 System Sequencing Performance Parameters (2017) [Online], [Accessed on 8th April 2018]. Available from World Wide Web: *https://www.illumina.com/systems/sequencingplatforms/nextseq/specifications.html*

391. OmniDB (2017) [Online], [Accessed on 20th December 2018]. Available from World Wide Web: *https://omnidb.org/en/*

392. Picard (2018) [Online], [Accessed on 20th December 2018]. Available from World Wide Web: *https://broadinstitute.github.io/picard/*

393. Rare chromosome disorder support group (2018) [Online], [Accessed on 29th June 2020]. Available from World Wide Web: *https://www.rarechromo.org/media/information/Chromosome%2017/17p13.1%20and%2017p13.2%20microdeletions%20FTNW.pdf*

394. The Human Gene Mutation Database (HGMD) at the Institute of Medical Genetics in Cardiff (2015) [Online], [Accessed on 14th February 2017]. Available from World Wide Web: *http://www.hgmd.cf.ac.uk/ac/index.php*

APPENDICES

Appendix A: The Data Collection Format Sheet Sample

Data Collection Sheet

Patient ID: _____

Three-generation Family Pedigree

Inclusion Criteria (IC)

1) Fulfil clinical diagnostic criteria for Cerebral Palsy ☐

AND

2) More than one family members (from the same parental lines) affected with similar clinical features / progression ☐

OR

3) Born from consanguineous parents ☐

OR

4) Parental age more than 40 years at the conception. ☐

5) Syndromic features with coagulopathies, prematurity and brain coagulopathies. ☐

Exclusion Criteria (EC)

1) Intra or post-partum events that may lead to hypoxia due to mechanical factor ☐

2) Other syndromic features with defined chromosomal abnormalities ☐

3) Evident infection during pregnancy ☐

4) Parental history of identifiable genetic disorder ☐

5) Family history of identifiable genetic disorder related to Cerebral Palsy ☐

Criteria for inclusion: IC1 AND any of the IC2-5

Criteria for exclusion: any of the EC1-5

Nervous Features:

Physical Examination: Done/Not Done

Developmental milestone: Normal/ Delayed

If delayed, specify:

_____

Higher Mental Function

Mental Retardation: Yes/No/Not Assessed

If yes, specify:

_____

Specific Learning Disability: Yes/No/Not Assessed

If yes, specify: _____

Other cognitive abnormalities:

Motor Function

Hypotonia/Hypertonia/Normal

Hyperreflexia/Hyporeflexia/Normal

Muscle Wasting: Yes/No

Abnormal Movement: Yes/No

If yes, specify:

_____

Muscle Weakness: Yes/No

If yes, specify:

_____

Cerebellar Dysfunction: Yes/No

If yes, specify:

_____

Sensory abnormalities: Yes/No

If yes, specify:

_____

Cranial Nerve Abnormalities: Yes/No

If yes, specify:

_____

Other nervous abnormalities:

Dysmorphic features:

Appendix B: The DNA Extraction

Table B1: The four readings (R) values of DNA concentration [DNA] (ng/µl), A260:A280 and A260:A230 for 20 samples.

ID	R1			R2			R3			R4			Average (ng/µl)
	[DNA] ng/µl	A260: A280	A260:A 230	[DNA] ng/µl	A260: A280	A260:A 230	[DNA] ng/µl	A260:A 280	A260:A 230	[DNA] ng/µl	A260: A280	A260: A230	
Family 1													
D1_S3	79.11	1.933	2.325	77.71	1.916	2.247	78.15	1.935	2.345	76.31	1.929	2.194	78.323
M1_S2	120.6	1.908	2.489	100.3	1.906	2.445	111.1	1.92	2.792	128.5	1.849	2.067	116.467
B1-1_S1	101.1	1.921	2.279	102.5	1.91	2.149	97.64	1.944	2.315	122.9	1.92	2.273	108.833
B1-3_S4	110.1	1.916	2.397	97.53	1.884	2.315	103.5	1.909	2.361	97.1	1.906	2.311	103.71
Family 2													
D2_S7	137.9	1.909	2.322	132.6	1.931	2.414	265.2	1.9	2.331	202.5	1.908	2.312	157.667
M2_S6	106.5	1.538	1.658	76.89	1.896	2.357	72.23	1.974	2.562	72.06	1.942	2.449	73.727
G2-1_S8	68.77	1.962	2.753	84.48	1.662	1.892	67.62	1.959	2.602	68.52	1.921	2.444	73.54
B2-2_S5	125.1	1.913	2.43	161.4	1.573	1.88	132.9	1.866	2.173	128.4	1.878	2.326	128.8
Family 3													
D3_S9	85.91	1.84	2.296	138.7	1.887	2.448	95.18	1.86	2.326	90.07	1.894	2.362	90.387
M3_S6	105.8	1.915	2.616	109.2	1.857	2.328	109.1	1.923	2.558	129	1.88	2.38	115.767
G3- 3_S10	91.15	1.879	2.366	90.29	1.928	2.526	112.1	1.882	2.364	105.7	1.904	2.395	102.983

Table B1: Continued

ID	R1			R2			R3			R4			Average (ng/µl)
	[DNA] ng/µl	A260: A280	A260:A 230	[DNA] ng/µl	A260: A280	A260:A 230	[DNA] ng/µl	A260:A 280	A260:A 230	[DNA] ng/µl	A260: A280	A260: A230	
Family 4													
D4_S4	127.8	1.883	2.402	123.1	1.913	2.508	129.3	1.898	2.414	124.3	1.922	2.49	127.133
M4_S3	76.21	1.847	1.932	75.47	1.881	1.923	78.28	1.904	1.959	84.07	1.763	1.793	76.653
B4-T1_ S1	94.15	1.936	2.563	94.64	1.931	2.337	94.28	1.913	2.432	90.75	1.909	2.337	93.223
B4-T2_ S2	55.08	1.90	2.403	56.03	1.931	2.422	60.66	1.901	2.409	65.05	1.923	2.342	60.58
Family 5													
D5_S9	73.68	2.002	2.605	77.00	1.920	2.255	74.3	2.002	2.478	74.97	1.928	2.266	74.9875
M5_S8	110.7	1.897	2.200	104.4	1.932	2.326	104	1.919	2.283	101.7	1.953	2.348	105.2
B5-1_S7	252.8	1.922	2.366	238.7	1.886	2.219	199.2	1.869	2.153	163.2	1.911	2.27	213.475
G5- 2_S10	228.7	1.926	2.357	148.3	1.909	2.249	133.6	1.897	2.198	120.1	1.929	2.279	157.675
Family 6													
B6-2_S5	61.62	1.908	2.444	64.13	1.782	1.563	61.07	1.861	2.001	61.04	1.867	2.014	61.243

Appendix C: DNA Library Preparation

Table C1: The list of 20 samples with their assigned Illumina default Index sequences (PCR primer) of Index 5 [i5] and Index 7 [i7] that were determined using IEM software (refer Method section 2.9.1 and 2.9.4).

Case	Description	[i7]_ID	Index Sequence	[i5]_ID	Index Sequence	Pool (PO1) or (PO2)
Fam 1						
D1_S3	Normal father	N703	AGGCAGAA	E505	CTCCTTAC	PO1
M1_S2	Normal mother	N702	CGTACTAG	E505	CTCCTTAC	PO1
B1-1_S1	CP son	N701	TAAGGCGA	E505	CTCCTTAC	PO1
B1-3_S4	CP son	N711	AAGAGGCA	E505	CTCCTTAC	PO1
Fam 2						
D2_S7	Normal father	N706	TAGGCATG	E505	CTCCTTAC	PO1
M2_S6	Normal mother	N705	GGACTCCT	E505	CTCCTTAC	PO1
G2-1_S8	CP son	N707	CTCTCTAC	E506	TATGCAGT	PO1
B2-2_S5	CP son	N704	TCCTGAGC	E505	CTCCTTAC	PO1
Fam 3						
D3_S9	Normal father	N706	TAGGCATG	E506	TATGCAGT	PO1
M3_S6	Normal mother	N705	GGACTCCT	E506	TATGCAGT	PO2
G3-3_S10	CP daughter	N712	GTAGAGGA	E506	TATGCAGT	PO1

Table C1: Continued.

Case	Description	[i7]_ID	Index Sequence	[i5]_ID	Index Sequence	Pool (PO1or PO2)
Fam 4						
D4_S4	Normal father	N709	GCTACGCT	E505	CTCCTTAC	PO2
M4_S3	Normal mother	N710	CGAGGCTG	E505	CTCCTTAC	PO2
B4-T1_S1	CP son	N707	CTCTCTAC	E505	CTCCTTAC	PO2
B4-T2_S2	CP son	N708	CAGAGAGG	E505	CTCCTTAC	PO2
Fam 5						
D5_S9	CP father	N710	CGAGGCTG	E506	TATGCAGT	PO2
M5_S8	Normal mother	N709	GCTACGCT	E506	TATGCAGT	PO2
B5-1_S7	CP son	N708	CAGAGAGG	E506	TATGCAGT	PO2
G5-2_S10	CP daughter	N711	AAGAGGCA	E506	TATGCAGT	PO2
Fam 6						
B6-1_S5	CP son	N712	GTAGAGGA	E505	CTCCTTAC	PO2

Appendix D: The full scripted commands that had been executed successfully indicated by "done" echo for each execution step during Bioinformatics analyses.

Appendix D.1 Raw Read Analyses for 2 arrays (array=batch 1 and array2=batch 2)

D.1.1 First FastQC

```
!/bin/bash

./fastqc/file directory for batch 1 raw/*.fastq.gz

./fastqc/file directory for batch 2 raw/*.fastq.gz
```

D.1.2 Trim Galore: Trimming the adapter and clipping the front and end

```
!/bin/bash

array=("B2-2_S5" "Undetermine(control)" "G2-1_S8" "B1-1_S1" "M2_S6" "G3-3_S10" "D2_S7" "D1_S3" "B1-3_S4" "D3_S9" "M1_S2" ) for element in ${array [@]}

do

for i in {1..4}; do

./trim_galore /file directory for batch 1/${element}_L00${i}_R1_001.fastq.gz /file directory for batch 1/${element}_L00${i}_R2_001.fastq.gz

-a CTGTCTCTTATACACATCT--paired--path_to_cutadapt /file directory for adapter --length 100--clip_R1 17 --three_prime_clip_R1 5 –three_prime_clip_R2 5 \

-o /file directory for clipped batch 1 output

done

done

echo "Completed for all arrays!"
```

D.1.3 The commands were repeated for batch 2 reads

```
!/bin/bash

array2=(“D4_S4” “M4_S3” “B4-T2_S2” “B4-T1_S1” “D5_S9” “G5-2_S10” “B5-1_S7” “M5_S8” “M3_S6” “B6-1_S5”) for element in ${array2 [@]}

do

for i in {1..4}; do

./trim_galore /file directory for batch 2/${element}_L00${i}_R1_001.fastq.gz /file directory for batch 2/${element}_L00${i}_R2_001.fastq.gz

-a CTGTCTCTTATACACATCT--paired--path_to_cutadapt /file directory for adapter --length 100--clip_R1 17 --three_prime_clip_R1 5 –three_prime_clip_R2 5 \

-o /file directory for clipped batch 2 output

done

done

echo “Completed for all arrays!”
```

D.1.4 Second FastQC

```
!/bin/bash

./fastqc/file directory for clipped batch 1/*.fq.gz

./fastqc/file directory for clipped batch 2/*.fq.gz
```

Appendix D.2: Reference Mapping (Assembly)

{element}.gz= refers to each sample ID in (“ ”) batch 1 and {element2}.gz= refers to each sample ID in (“ ”) in batch 2.

D.2.1 Tools and software loading

```
!/bin/bash
module load programming/jdk1.8
module load bioinformatics/bwa-0.7.17
module load bioinformatics/picard
module load bioinformatics/samtools-1.5
module load bioinformatics/GATK
Generate dict file for indexing
```

D.2.2 Generate dictionary (dict.) file for indexing

```
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar
CreateSequenceDictionary \
R=/file directory of hg19 database.fa \
O=/ file directory of hg19 database.dict
```

D.2.3 Transformed to unmapped BAM (uBAM) files

```
array=( “B2-2_S5” “G2-1_S8” “B1-1_S1” “M2_S6” “G3-3_S10” “D2_S7” “D1_S3” “B1-3_S4” “D3_S9” “M1_S2”)
array2=( “D4_S4” “M4_S3” “B4-T2_S2” “B4-T1_S1” “D5_S9” “G5-2_S10” “B5-1_S7” “M5_S8” “M3_S6” “B6-1_S5”
```

D.2.3 Continued

```
for element in ${array[@]}
do
for i in {1..4}; do
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar FastqToSam \
FASTQ=/file directory for {element}_L00${i}_R1_001_val_1.fq.gz \
FASTQ2=/file directory for {element}_L00${i}_R2_001_val_2.fq.gz \
OUTPUT=/file directory for {element}_L004{I}.bam \
LIBRARY_NAME=Nextera READ_GROUP_NAME=L00${i}
SAMPLE_NAME=${element} \
PLATFORM=illumina SEQUENCING_CENTER=UMBI
done
done
echo “Batch 1 unmapped BAM done”
```

```
for element2 in ${array2[@]}
do
for i in {1..4}; do
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar FastqToSam \
FASTQ=/file directory for {element2}_L00${I}_R1_001_val_1.fq.gz \
FASTQ2=/ file directory for {element2}_L00${i}_R2_001_val_2.fq.gz \
OUTPUT=/file directory for {element2}_L00${i}.bam \
LIBRARY_NAME=Nextera READ_GROUP_NAME=L00${i}
SAMPLE_NAME=${element2} \
PLATFORM=illumina SEQUENCING_CENTER=UMBI
done
done
echo “Batch 2 unmapped BAM done”
```

D.2.4 Adapter Marking

This will mark illumina adapters in the reads as XT

D.2.4.1 For batch 1 {array}

```
for element in ${array[@]}
do
for i in {1..4}; do
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar MarkIlluminaAdapters \
INPUT=/file directory for {element}_L00${i}.bam \
OUTPUT=/file directory for {element}_L00${i}.bam \
METRICS=/file directory for {element}_L00${i}.metrics.txt
done
done
echo “Batch 1 mark adapter done”
for element in ${array[@]}
do
for i in {1..4}; do
gzip/file directory for {element}_L00${i}.metrics.txt &
done
done
echo “gzip batch 1 metrics done”
```

D.2.4.2 For batch 2 {array2}

```
for element2 in ${array2[@]}
do
for i in {1..4}; do
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar MarkIlluminaAdapters \
INPUT=/file directory for {element}_L00${i}.bam \
OUTPUT=/file directory for {element2}_L00${i}.bam \
METRICS=/file directory for {element2}_L00${i}.metrics.txt
done
done
echo “Batch 2 mark adapter done”
for element in ${array[@]}
do
for i in {1..4}; do
gzip/file directory for {element2}_L00${i}.metrics.txt &
done
done
echo “gzip batch 2 metrics done”
echo “End: No Trim 17 Front 5 End”
```

D.2.4.3 Read Quality Changing

1) Change read quality of XT reads to 2

2) This will effectively mask the XT reads

3) Shows up as ####### in the quality score

4) Write as interleaved fastqc files

D.2.4.3 (a) For batch 1 read quality changing

```
for element in ${array[@]}
do
for i in {1..4}; do
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar SamToFastq \
INPUT=/file directory for{element}_L00${i}.bam \
CLIPPING_ATTRIBUTE=XT CLIPPING_ACTION=2 INTERLEAVE=true
NON_PF=true \
TMP_DIR=/file directory for {element}_L00${i}.fq.gz
done
done
echo “Batch 1 done for change read quality”
```

D.2.4.3 (b) For batch 2 read quality changing

```
for element2 in ${array2[@]}
do
for I in {1..4}; do
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar SamToFastq \
INPUT=/file directory for {element2}_L00${i}.bam \
CLIPPING_ATTRIBUTE=XT CLIPPING_ACTION=2 INTERLEAVE=true
NON_PF=true \
TMP_DIR=/file directory for {element2}_L00${i}.fq.gz
done
done
echo “Batch 2 done for change read quality”
```

D.2.5 Alignment mapping and merging with unmapped

D.2.5.1 For batch 1

```
for element in ${array[@]}
do
for i in {1..4}; do
zcat /file directory for {element}_L00${i}.fq.gz | \
bwa mem –M –t 6 –p /file directory for hg19 database/hg19.sorted.fa – 2>/file directory for {element}_L00${i}.log | \
java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar MergeBamAlignment \
R=/file directory for hg19.sorted.fa \
UNMAPPED_BAM=/file directory for /${element}_L00${i}.bam
ALIGNED_BAM=/dev/stdin \
O=/file directory for {element}_L00${i}_merged.bam \
CREATE_INDEX=true ADD_MATE_CIGAR=true CLIP_ADAPTERS=false \
CLIP_OVERLAPPING_READS=true
INCLUDE_SECONDARY_ALIGNMENTS=true
MAX_INSERTIONS_OR_DELETIONS=-1 \
PRIMARY_ALIGNMENT_STRATEGY=MostDistant
ATTRIBUTES_TO_RETAIN=XS TMP_DIR=/file directory
for{element}_L00${i}.log
done
done
```

D.2.5.2 For batch 2

```
for element2 in ${array2[@]}

do

for I in {1..4}; do

zcat /file directory for/${element2}_L00${i}.fq.gz | \

bwa mem –M –t 6 –p /file directory for hg19 database/hg19.sorted.fa – 2>/file location for /merged_BAM/Error_report/bwa_error_${element2}_L00${i}.log | \

java –Xmx20G –jar /opt/biotools/picard-tools-2.15.0/picard.jar MergeBamAlignment \

R=/file directory for hg19.sorted.fa \UNMAPPED_BAM=/file directory for/${element2}_L00${i}.bam \

ALIGNED_BAM=/dev/stdin \

O=/file directory for/${element2}_L00${i}_merged.bam \

CREATE_INDEX=true ADD_MATE_CIGAR=true CLIP_ADAPTERS=false \

CLIP_OVERLAPPING_READS=true

INCLUDE_SECONDARY_ALIGNMENTS=true

MAX_INSERTIONS_OR_DELETIONS=-1 \

PRIMARY_ALIGNMENT_STRATEGY=MostDistant

ATTRIBUTES_TO_RETAIN=XS TMP_DIR=/file directory for

{element2}_L00${i}.log

done

done

echo “Batch 2 done for map and merge alignment with unmapped”
```

Appendix D.3: Duplicate Marking

D.3. 1 For batch 1

```
for element in ${array[@]}

do

java -Xmx20G -jar /opt/biotools/picard-tools-2.15.0/picard.jar MarkDuplicates \

INPUT=/file directory for {element}_L001_merged.bam \

INPUT=/file directory for {element}_L002_merged.bam \

INPUT=/file directory for {element}_L003_merged.bam \

INPUT=/file directory for {element}_L004_merged.bam \

OUTPUT=/file directory for {element}_L1-4.bam \

METRICS_FILE=/file directory for {element}_L1-4.txt \

CREATE_INDEX=true \

TMP_DIR=/file directory for merged_BAM/tempo

done
```

D.3.2 For batch 2

```
do

java -Xmx20G -jar /opt/biotools/picard-tools-2.15.0/picard.jar MarkDuplicates \

INPUT=/file directory for {element2}_L001_merged.bam \

INPUT=/file directory for {element2}_L002_merged.bam \

INPUT=/file directory for {element2}_L003_merged.bam \

INPUT=/file directory for {element2}_L004_merged.bam \

OUTPUT=/file directory for {element2}_L1-4.bam \

METRICS_FILE=/file directory for {element2}_L1-4.txt \

CREATE_INDEX=true \

TMP_DIR=/file directory for merged_BAM/tempo

done

done
```

Appendix D.4: Base Recalibration (Run BQSR)

D.4.1 Task 1: Analyze patterns of covariation in the sequence dataset

```
for element in ${array[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T BaseRecalibrator \
-R /file directory for hg19 database/hg19.sorted.fa \
-I /file directory for {element}_L1-4.bam \
-L /file directory for /nexterarapidcapture_expandedexome_targetedregions.bed \
-knownSites /file directory for hg19 database/dbsnp_138.hg19.vcf \
-knownSites /file directory for hg19 database/Mills_and_1000G_gold_standard.indels.hg19.sites.vcf \
-o / file directory for {element}_L1-4_recal_data.table
done
```

```
for element2 in ${array2[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T BaseRecalibrator \
-R /file directory for hg19 database/hg19.sorted.fa \
-I /file directory for {element2}_L1-4.bam \
-L /file directoryfor nexterarapidcapture_expandedexome_targetedregions.bed \
-knownSites /file directory for hg19 database/dbsnp_138.hg19.vcf \
-knownSites /file directory for hg19 database/Mills_and_1000G_gold_standard.indels.hg19.sites.vcf \
-o / file directory for {element2}_L1-4_recal_data.table
done
```

D.4.2 Task 2: Do a second pass to analyze covariation remaining after recalibration

```
for element in ${array[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \i
-T BaseRecalibrator \
-R /file directory for hg19 database/hg19.sorted.fa \
-I file directory for {element}_L1-4.bam \
-L /file directory for nexterarapidcapture_expandedexome_targetedregions.bed \
-knownSites /file directory for hg19 database/dbsnp_138.hg19.vcf \
-knownSites /file directory for
hg19database/Mills_and_1000G_gold_standard.indels.hg19.sites.vcf \
-BQSR /file directory for /expandedexome_targetedregions/${element}_L1-
4_recal_data.table
-o /file directory for {element}_L1-4_post_recal_data.table
done
```

```
for element2 in ${array2[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \i
-T BaseRecalibrator \
-R /file directory for hg19 database/hg19.sorted.fa \
-I file directory for {element2}_L1-4.bam \
-L /file directory for nexterarapidcapture_expandedexome_targetedregions.bed \
-knownSites /file directory for hg19 database/dbsnp_138.hg19.vcf \
-knownSites /file directory for
hg19database/Mills_and_1000G_gold_standard.indels.hg19.sites.vcf \
-BQSR /file directory for {element2}_L1-4_recal_data.table
-o /file directory for {element2}_L1-4_post_recal_data.table
done
```

D.4.3 Task 3: Generate before/after plots

```
for element in ${array[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T AnalyzeCovariates \
-R /file directoryfor hg19/hg19.sorted.fa \
-L /file directory for nexterarapidcapture_expandedexome_targetedregions.bed \
-before /file directory for {element}_L1-4_recal_data.table \
-after /file directory for {element}L1-4_post_recal_data.table \
-plots /file directory for /expandedexome_targetedregions/${element}L1-4_recalibration_plots.pdf
done
```

```
for element2 in ${array2[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T AnalyzeCovariates \
-R /file directory for hg19/hg19.sorted.fa \
-L /filedirectoryfor nexterarapidcapture_expandedexome_targetedregions.bed \
-before /file directory for {element2}_L1-4_recal_data.table \
-after /file directory for {element2}L1-4_post_recal_data.table \
-plots /file directory for {element2}L1-4_recalibration_plots.pdf
done
```

D.4.4 Task 4: Generate before/after plots

```
for element in ${array[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T PrintReads \
-R /file directory for hg19 database/hg19.sorted.fa \
-I /file directory for {element}_L1-4.bam \
-L /file directory for nexterarapidcapture_expandedexome_targetedregions.bed \
-BQSR /file directory for {element}_L1-4_recal_data.table
-o //file directory for {element}_L1-4_recal_reads.bam
done
```

```
for element2 in ${array2[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T PrintReads \
-R /file directory for hg19 database/hg19.sorted.fa \
-I /file directory for {element2}_L1-4.bam \
-L /file location for nexterarapidcapture_expandedexome_targetedregions.bed \
-BQSR /file directory for {element2}_L1-4_recal_data.table
-o //file directory for {element2}_L1-4_recal_reads.bam
done
```

Appendix D.5: Variant Calling and Joint Genotyping

D.5.1: Variants Calling with HaplotypeCaller

```
for element in ${array[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T HaplotypeCaller \
-R /file directory for hg/hg19database9.sorted.fa \
-I file directory {element}_L1-4_recal_reads.bam \
-L /file directory for nexterarapidcapture_expandedexome_targetedregions.bed \
--genotyping_mode DISCOVERY \
-stand_emit_conf 10 \
-stand_call_conf 30 \
-o file directory for {element}_L1-4_raw_snps_indels.g.vcf \
-ERC GVCF \
-variant_index_type LINEAR -variant_index_parameter 128000
done
```

```
for element2 in ${array2[@]}
do
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T HaplotypeCaller \
-R /file directory for hg/hg19database9.sorted.fa \
-I file directory for {element2}_L1-4_recal_reads.bam \
-L /file directory for nexterarapidcapture_expandedexome_targetedregions.bed \
--genotyping_mode DISCOVERY \
-stand_emit_conf 10 \
-stand_call_conf 30 \
-o file directory for {element2}_L1-4_raw_snps_indels.g.vcf \
-ERC GVCF \
-variant_index_type LINEAR -variant_index_parameter 128000
done
```

Appendix D.5.2: Join Genotyping

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T GenotypeGVCFs \
-R /file directory for hg19 database/hg19.sorted.fa \
-stand_call_conf 4 -stand_emit_conf 4 \
-V /file directory for B2-2_S5_L1-4_raw_snps_indels.g.vcf \
-V /file directory for G2-1_S8_raw_snps_indels.g.vcf \
-V / file directory for B1-1_S1_L1-4_raw_snps_indels.g.vcf \
-V /file directory for M2_S6_L1-4_raw_snps_indels.g.vcf \
-V /file directory for G3-3_S10_L1-4_raw_snps_indels.g.vcf \
-V /file directory for D2_S7_L1-4_raw_snps_indels.g.vcf \
-V / file directory for D1_S3_L1-4_raw_snps_indels.g.vcf \
-V / file directory for B1-3_S4_L1-4_raw_snps_indels.g.vcf \
-V / file directory for D3_S9_L1-4_raw_snps_indels.g.vcf \
-V / file directory for M1_S2_L1-4_raw_snps_indels.g.vcf \
-V / file directory for D4_S4_L1-4_raw_snps_indels.g.vcf \
-V / file directory for M4_S3_L1-4_raw_snps_indels.g.vcf \
-V / file directory for B4-T2_S2_L1-4_raw_snps_indels.g.vcf \
-V / file directory for B4-T1_S1_L1-4_raw_snps_indels.g.vcf \
-V / file directory for D5_S9_L1-4_raw_snps_indels.g.vcf \
-V / file directory for G5-2_S10_L1-4_raw_snps_indels.g.vcf \
-V / file directory for B5-1_S7_L1-4_raw_snps_indels.g.vcf \
-V / file directory for M5_S8_L1-4_raw_snps_indels.g.vcf \
-V / file directory for M3_S6_L1-4_raw_snps_indels.g.vcf \
-V / file directory for B6-1_S5_L1-4_raw_snps_indels.g.vcf \
-o /file directory for CombinedBatch1_2AfterJoinGeno.vcf
```

Appendix D.6: Variant Filtering

D.6.1 Task 1: Build the SNP recalibration model

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \

-T VariantRecalibrator \

-R /file directory for hg19 database/hg19.sorted.fa \

-input /file directory for CombinedBatch1_2AfterJoinGeno.vcf

-resource:hapmap,known=false,training=true,truth=true,prior=15.0 /file directory for
hg19database/hapmap_3.3.hg19.sites.vcf \

-resource:omni,known=false,training=true,truth=true,prior=12.0 / file directory for
hg19database /1000G_omni2.5.hg19.sites.vcf \

-resource:1000G,known=false,training=true,truth=false,prior=10.0 / file directory for
hg19database /Mills_and_1000G_gold_standard.indels.hg19.sites.vcf \

-resource:dbsnp,known=true,training=false,truth=false,prior=2.0 / file directory for
hg19database /dbsnp_138.hg19.vcf \

-an DP \

-an QD \

-an FS \

-an SOR \

-an MQ \

-an MQRankSum \

-an ReadPosRankSum \

-an InbreedingCoeff \

-mode SNP \

-tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0 \

-recalFile /file directory for recalibrate_SNP.recal \

-tranchesFile /file directory for recalibrate_SNP.tranches \

-rscriptFile /file directory for recalibrate_SNP_plots.R
```

D.6.2 Task 2: Apply the desired level of recalibration to the SNPs in the call set

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T ApplyRecalibration \
-R /file directory for crm_hg19/hg19.sorted.fa \
-input / file directory for CombinedBatch1_2AfterJoinGeno.vcf
-mode SNP \
--ts_filter_level 99.0 \
-recalFile /filedirectoryfor recalibrate_SNP.recal \
-tranchesFile /file directory for recalibrate_SNP.tranches \
-o /file directory for recalibrated_snps_raw_indels.vcf
```

D.6.3 Task 3: Build the Indel recalibration model

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \

-T VariantRecalibrator \

-R /file directory for hg19 database/hg19.sorted.fa \

-input /file directory for recalibrated_snps_raw_indels.vcf \

-resource:mills,known=false,training=true,truth=true,prior=12.0 /file directory for hg19 database/Mills_and_1000G_gold_standard.indels.hg19.sites.vcf \

-resource:dbsnp,known=true,training=false,truth=false,prior=2.0 file directory for crm_hg19/dbsnp_138.hg19.vcf \

-an QD \

-an DP \

-an FS \

-an SOR \

-an MQRankSum \

-an ReadPosRankSum \

-an InbreedingCoeff \

-mode INDEL \

-tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0 \

--maxGaussians 4 \

-recalFile /file directory for recalibrate_INDEL.recal \

-tranchesFile /file directory for recalibrate_INDEL.tranches \

-rscriptFile /file directory for recalibrate_INDEL_plots.R
```

D.6.4 Task 4: Apply the desired level of recalibration to the Indels in the call set

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T ApplyRecalibration \
-R /filedirectoryfor hg19 database/hg19.sorted.fa \
-input /file directory for recalibrated_snps_raw_indels.vcf \
-mode INDEL \
--ts_filter_level 99.0 \
-recalFile /file directory for recalibrate_INDEL.recal \
-tranchesFile /file directory for recalibrate_INDEL.tranches \
-o /file directory for recalibrated_variants.vcf
```

Appendix D.7: Genotype Refinement

D.7.1 Task 1: Derive posterior probabilities of genotypes

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T CalculateGenotypePosteriors \
-R /file directory for hg19 database/hg19.sorted.fa \
-V file directory for recalibrated_variants.vcf \
-supporting /file directory for hg19 database/1000G_phase3_v4_20130502.sites.vcf \
-ped /file directory for Pedigree_Trio.ped \
-o /file directory for recalibratedVariants.postCGP.vcf
```

D.7.2 Task 2: Filter low quality genotypes

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T VariantFiltration \
-R /file directory for hg19 database/hg19.sorted.fa \
-V /file directory for recalibratedVariants.postCGP.vcf \
-G_filter "GQ < 20.0" -G_filterName lowGQ \
-o /file directory for recalibratedVariants.postCGP.Gfiltered.vcf
```

D.7.3 Task 3: Annotate possible de novo mutations

```
java -jar /opt/biotools/GenomeAnalysisTK.jar \
-T VariantAnnotator \
-R /file directory for hg19 database/hg19.sorted.fa \
-V /file directory for recalibratedVariants.postCGP.Gfiltered.vcf \
-A PossibleDeNovo -ped /filedirectoryfor Pedigree_Trio.ped \
-o /file directory for recalibratedVariants.postCGP.Gfiltered.deNovos.vcfdu
```

Appendix D.8: Variant Annotation

```
perl /file directory for recalibratedVariants.postCGP.Gfiltered.deNovos.annovar

perl /file directory for recalibratedVariants.postCGP.Gfiltered.deNovos.annovar /file directory for annovar/humandb/ -buildver hg19 -out CP_SNP_INDEL -remove -protocol cytoBand, 1000g2015aug_all, avsift, cg69, clinvar_20140929, clinvar_20170130, cosmic70, esp6500siv2_all, exac03, exac03nontcga, hrcr1, icgc21, kaviar_20150923, kgXref, nci60, snp137, snp138, ljb26_all, ensGene, knownGene, refGene –operation r,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,g,g,g -nastring .

perl /file directory for annovar/convert2annovar.pl -format vcf4 /file directory for recalibratedVariants.postCGP.Gfiltered.deNovos.vcf -outfile /file directory for recalibratedVariants.postCGP.Gfiltered.deNovos.annovar -allsample -includeinfo -comment
```

```
for element in ${array[@]}

do

perl /file directory for annovar/table_annovar.pl /file directory for recalibratedVariants.postCGP.Gfiltered.deNovos.annovar.${element}.avinput /file directory for annovar/humandb/ -buildver hg19 -out CP_SNP_INDEL_${element} -remove -protocol cytoBand, 1000g2015aug_all, avsift, cg69, clinvar_20140929, clinvar_20170130, cosmic70, esp6500siv2_all, exac03, exac03nontcga, hrcr1, icgc21, kaviar_20150923, kgXref, nci60, snp137, snp138, ljb26_all, ensGene, knownGene, refGene -operation r,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,g,g,g –nastring .

done
```

```
for element2 in ${array2[@]}

do

perl /file directory for geno_refinement/recalibratedVariants.postCGP.Gfiltered.deNovos.annovar.${element2}.avinput /file directory for annovar/humandb/ -buildver hg19 –out CP_SNP_INDEL_${element2} -remove -protocol cytoBand, 1000g2015aug_all, avsift, cg69, clinvar_20140929, clinvar_20170130, cosmic70, esp6500siv2_all, exac03, exac03nontcga, hrcr1, icgc21, kaviar_20150923, kgXref, nci60, snp137, snp138, ljb26_all, ensGene, knownGene, refGene -operation r,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,g,g,g –nastring .

done
```

```
perl /file directory for recalibrated_variants_withoutped.vcf -format vcf4old -includeinfo -outfile /file directory for recalibrated_variants_withoutped.annovar

perl /file directory for recalibrated_variants_withoutped.annovar /file directory for annovar/humandb/ -buildver hg19 -out CP_SNP_INDEL -remove -protocol cytoBand, 1000g2015aug_all, avsift, cg69, clinvar_20140929, clinvar_20170130, cosmic70, esp6500siv2_all, exac03, exac03nontcga, hrcr1, icgc21, kaviar_20150923, kgXref, nci60, snp137, snp138, ljb26_all, ensGene, knownGene, refGene -operation r,f,f,f,f,f,f,f,f,f,f,f,f,f,f,f,g,g,g –nastring .
```

Appendix E: List of Presentation

1. Oral presentation

Nur Atikah Binti Nor Azhar, Hui San Ong, N.S. Yaacob, W.Y. Low, S.A. Razak, N. Yahaya, T.H. Sasongko: Bioinformatics pipeline: a preliminary study of malay idiopathic cerebral palsy patients in Kelantan

List of Oral Presentations

1. Gerard Kian-Meng Goh, Keith Dunker, James A. Foster, Vladimir N. Uversky: The Viral Shapeshifters: Using Disorder Analysis to Solve the HIV Vaccine Mystery
2. Ahmed Salisu , Nurulfiza, M.I., Mariatulqabtiah, A.R.,Hair-Bejo, M., Omar, A. R., Aini, I.: Molecular Detection And Phylogenetic Analysis Of UPMT27 Field Isolate Of Malaysia Fowl Adenovirus Associated With Inclusion Body Hepatitis
3. Gerard Kian-Meng Goh, Keith Dunker, James A. Foster, Vladimir N. Uversky: Using Disorder Analysis to Study the Nipah Virus
4. Linda Erlina, Arry Yanuar, Fadilah, Rafika Indah Paramita: Molecular Docking And Molecular Dynamics Simulation Analysis For Finding HDAC Class IIA Inhibitor As New Antidiabetic Drug
5. Jason S.E. Loo, Abigail L. Emtage, Sze Siew Lee, Alvina L.W. Kueh: Microsecond Molecular Dynamics Simulations Of The Active And Inactive-State CB1 Cannabinoid Receptor Following Cross-Docking: Agonist/Antagonist Discrimination In Virtual Screening
6. Nur Atikah Binti Nor Azhar, Hui San Ong, N.S. Yaacob, W.Y. Low, S.A. Razak, N. Yahaya, Teguh Haryo Sasongko: Bioinformatics Pipeline: A Preliminary Study Of Malay Idiopathic Cerebral Palsy Patients In Kelantan
7. Wan Fahmi Wan Mohamad Nazarie, A Rahman A Jamal, Nurul-Syakima Ab Mutalib: Ovarian Serous Cystadenocarcinoma Drug-Treatment Transcriptome Analysis Of TCGA Data Using Deseq2
8. Muchamad Dafip: Identification Of Gene Mutation And Protein Structure Of LIPL32 In Human-Pathogenic Leptospira : A Bioinformatic Approach
9. Vannajan Sanghiran Lee, Chanat Thanavanich, Sila Kittiwachana, Ramtin Ranji, Chee Sun Liew, Wei Lim Chong, Luelak Lomlim, Noorsaadah Abd Rahman, Sharifuddin Md Zain: Cheminformatics Knime Workflow For Alzheimer's Disease
10. Reza Febrian, Hilyatuz Zahroh, Sandra Hermanto: *In Silico* Approach To Discover Bioactive Peptides From Goat Milk Casein Hydrolysates As Angiotensin Converting Enzyme Inhibitor
11. Rafika Indah Paramita, Fadilah Fadilah, Hadi Poerwono, Linda Erlina: Ligand-Based Drug Design Of Pinostrobin Derivatives By Pharmacophore Modelling As Wild-Type And Mutant HER2 Inhibitor)
12. Puteri Nur Sarah Diana Engku Baharuddin, Mohd Shahir Shamsir: Structural Insights Of Human Aquaglyceroporin 7 And Its Selectivity Filter
13. Muhammad Alwee Fikri Abdullah, Farah Najihah Isa'ali, 3Siti Azma Jusoh: Potential Interactions Of Seven Known Drugs To The 4-Hydroxytamoxifen Secondary Binding Site Of Estrogen Receptor-β

CERTIFICATE
OF ACHIEVEMENT

THIS CERTIFICATE IS PROUDLY PRESENTED TO

NUR ATIKAH BINTI NOR AZHAR

InSyB 2018

As an Oral Presenter at the
2nd International Symposium on Bioinformatics
'Advancing Bioinformatics: The Future of Life Sciences'

December 20-21, 2018
Perdana University, Serdang, Malaysia

Yam Wai Keat, PhD
Co-Chair, InSyB 2018

Mohammad Asif Khan, PhD
Co-Chair, InSyB 2018

APBioNET

PERDANA UNIVERSITY

MyBioinformatics