

CHAPTER III

RESEARCH METHODS

3.1. Research Field

This research was in the field of molecular genetics, interrelated with clinical genetics and pediatric neurology.

3.2. Research location

Students with mental retardation from several special schools had already been selected from SLB Negri, SLB Pelita Ilmu, SLB Hj. Soemiyati in a previous study cohort in Semarang and Bandung (Mundhofir, 2008). DNA analysis of MCPH and *PTEN* genes was performed in DNA diagnostics laboratory of Radboud University Nijmegen Medical Center, The Netherlands.

3.3. Research period

The research was conducted between December 2010 and September 2011.

3.4. Research Design

This research was an exploration study with molecular genetic analysis (DNA sequencing)

3.5. Population and Samples

3.5.1. Population

The study population was taken from a previous study cohort consisted of 527 individuals with mental retardation from special schools in Semarang and Bandung. This initial cohort could only obtain history of pre-, ante-, and postnatal trauma and history of asphyxia at birth from few parents of the included

individuals. Physical examination, including dysmorphology assessment and measurement of the head circumference, was already performed for all the individuals before they were included in the previous study cohort. Routine chromosome analysis and DNA extraction was performed in Molecular and Cytogenetic laboratory of Center for Biomedical Research, Faculty of Medicine Diponegoro University Semarang in 2007 by the previous investigation. MLPA technique with P070 probe-kit for detecting subtelomeric deletions and duplication, as well as fragment length analysis for *FMR-1* gene were performed in the previous study (Mundhofir, 2008). Analysis of *SCN1A*, *LGII*, *ARX* and *STXBPI* gene was performed on the same cohort with epilepsy with negative finding (Afadiyanti, 2011).

3.5.2. Sample selection

The samples were selected from the previous cohort of 527 individuals with mental retardation (Mundhofir, 2008). The individuals with the occipito-frontal circumference (OFC) measurement of less than -2 standard deviation (microcephaly) as well as individuals with $OFC > 2$ standard deviation (macrocephaly) measured with the Nellhaus charts (Appendix II) was selected.

Measurement of OFC was done by 6 research assistants trained in dysmorphology examination in the previous study (Mundhofir, 2008). Each measurement was taken three times consecutively by three different examiners in one session, and an average of these three measurements was taken as the OFC measurement result and recorded in examination form. However, kappa reliability test was not performed in the previous study and not available for calculation in the present study. The secondary data available for this study is the average OFC of each subject.

The inclusion criteria are the followings:

- a. Mental retardation
- b. Microcephaly; defined as measurement of OFC less than -2 standard deviation or macrocephaly; defined as measurement of OFC more than 2 standard deviation based on the Nelhaus chart.
- c. Consented to join the reseach by signing informed consent

The exclusion criteria are children with chromosomal abnormality, Fragile X, sub-telomeric deletion or duplication diagnosed/confirmed in the previous study (Mundhofir, 2008).

Based on the criteria, 48 (23 female and 25 male) microcephalic individuals with mental retardation would undergo DNA sequencing for MCPH genes, while 10 macrocephalic individuals with mental retardation would undergo *PTEN* gene sequencing.

3.6. Variables of Study and Operational Definitions

3.6.1. List of Variables

a. Microcephaly

Microcephaly is the presence of occipito-frontal circumference (OFC) of the head that is less than -2SD from the normal range

b. Macrocephaly

Macrocephaly is the presence of occipito-frontal circumference (OFC) of the head that is more than -2SD from the normal range.

c. Mental retardation

As defined by the American Associaton on Intellectual Developmental Disabilities, the criteria are as follows (APA, 1994): IQ<70, concomitant limitations of two or more areas of adaptive skills, onset before the age of 18.

d. Genetic variants

The genetic variants are variation found from the sequencing of molecular genetic testing of the *MCPHI*, *WDR62*, *STIL*, *CENPJ*, *ASPM*, *CDK5RAP2* and *PTEN* gene. This will include any missense, nonsense, splicing, small deletions, small insertions, that can be found in the *ASPM*, *WDR62*, *CENPJ*, *CDK5RAP2*, *MCPHI* and *STIL* genes; as well as any pathogenic missense, nonsense, splicing, small deletions, small insertions, and large deletions that can be found in the *PTEN* gene. The results were categorized as genetic variants, pathogenic mutations, and unclassified variants.

3.6.3. Operational definitions

a. Microcephaly

The occipito-frontal circumference (OFC) measurement smaller than -2 standard deviations. Measurement of OFC was done using measuring tape which encircles the glabella and most prominent part of the posterior occiput. OFC measurement was performed in a previous study (Mundhofir, 2008). The data was recorded in centimetres in the previous study, taken as secondary data and plotted on the Nellhaus head chart (Nellhaus, 1968) for the present study.

Scale: Nominal

b. Macrocephaly

The occipito-frontal circumference (OFC) measurement larger than 2 standard deviations. Measurement of OFC was done using measuring tape which encircles the glabella and most prominent part of the posterior occiput. OFC measurement was performed in a previous study (Mundhofir, 2008). The data was recorded in centimetres in the previous study, taken as secondary data and plotted on the Nellhaus head chart (Nellhaus, 1968) for the present study.

Scale: Nominal

c. Mental retardation

as defined by the American Association on Intellectual Developmental Disabilities, the criteria are as follows (APA, 1994)

- IQ < 70
- Concomitant limitations of two or more areas of adaptive skills
- Onset before the age of 18

d. Genetic variants

Genetic variation in DNA sequence of the human genome as compared to reference sequence in *Ensembl* assembly GRCh37 (GCA_000001405.6) from the Genome Reference Consortium (The Genome Reference Consortium, 2011). The results of the genetic variants will be pathogenic mutation or unclassified variants.

1) Pathogenic mutation

Genetic variants found in the sequencing would be compared with known pathogenic mutations as described in the literature and *Human Gene Mutation Database* (HGMD, 2011). When these variants are previously described as causative for MCPH or macrocephaly and mental retardation, they were termed as “pathogenic mutations”.

Scale: Nominal

The mutations are categorized as follows:

- a)** Missense: a nucleotide change is causing replacement of one amino acid with a different amino acid into the protein sequence.
- b)** Nonsense: a nucleotide change is causing replacement of one amino acid with a stop codon into the protein sequence.
- c)** Splicing: a mutation that alters a specific sequence denoting the site at which the splicing of an intron takes place.
- d)** Small deletions: is a mutation in which a part of a sequence of DNA nucleotides is missing.

- e) Small insertions: is a mutation in which a one or more nucleotides are inserted into a DNA sequence.
- f) Large deletions: a mutation in which a large part of genetic material is missing, for example a whole exon or a whole gene

2) Unclassified variants:

Unclassified variants are genetic variants in the form of point mutation that is found in DNA sequencing. These genetic variants, when found, would undergo protein prediction programs, namely SIFT, Polyphen-2 and Align-GVGD. SIFT would conclude whether the amino acid change is “tolerated” or “not tolerated”. Polyphen-2 would conclude whether the amino acid change is “benign”, “possibly damaging”, or “damaging”. Align-GVGD would calculate the class of amino acid change as between C0 to C65, depending on the level of risk estimates.

The results of the protein prediction programs will result in scores, and together will form the conclusion on their likelihood of pathogenicity.

Scale: Nominal

The unclassified variants were categorized as (Bell, 2007):

- a) Class 1 (UV1): Certainly non pathogenic
- b) Class 2 (UV2): Unlikely to be pathogenic
- c) Class 3 (UV3): Likely to be pathogenic
- d) Class 4 (UV4): Certainly pathogenic

The genes examined in this study are as follows:

- a) *MCPH1* is the gene coding the Microcephalin protein and located on chromosome 8p23.
- b) *MCPH2*, also called *WDR62*, is the gene coding the WD repeat domain 62 protein and located on 19q13.12.
- c) *MCPH3*, also known as *CDK5RAP2*, is the gene coding the Cyclin-dependent kinase 5 regulatory associated protein 2 and located on chromosome 9q33.3.

- d) *MCPH5*, also known as *ASPM*, is the gene coding the Abnormal spindle like microcephaly associated protein and located on chromosome 1q31.
- e) *MCPH6*, also known as *CENPJ*, is the gene coding the Centromeric protein J and located on chromosome 13q12.2.
- f) *MCPH7*, or *STIL*, is the gene coding the SCL/TAL1 interrupting locus and located on chromosome 1p32.

3.7. Method

3.7.1. Patients collection

From a population of mentally retarded (MR) individuals in the previous study, 48 MR individuals with microcephaly and 10 MR individuals with macrocephaly was collected. The previous cohort study team performed the physical examination, dysmorphology assessment, blood collection, cytogenetic analysis, DNA isolation, Fragile X screening and MLPA for subtelomeric deletions and duplications screening on the study population (Mundhofir, 2008). The previous study has performed fragment length analysis for *FMR-1* gene as well as analysis of subtelomeric deletions and duplication using MLPA with P070 probe-kit (Mundhofir, 2008). Analysis of *SCN1A*, *LGII*, *ARX* and *STXBPI* gene was performed on the same cohort with epilepsy with negative finding (Afadiyanti, 2011). DNA extraction was performed using the salting out method based on the manual of the Center for Biomedical Research Semarang (Faradz, 2004) (Appendix I).

OFC measurement data of the study population was collected in the previous study, and only plotted on the Nellhaus chart for this study due to its broad age range of 0-18 years old and the unavailability of head chart from a South East Asian population with the same age range. The individuals with microcephaly or macrocephaly were included in the current study. The individuals with microcephaly will undergo PCR and continued with sequencing of *MCPHI*, *WDR62*, *STIL*, *CENPJ*, *ASPM*, and *CDK5RAP2* genes. The individuals with macrocephaly will undergo PCR continued with sequencing of *PTEN* gene.

3.7.2. General laboratory technique

DNA sequencing for the most prevalent MCPH genes, *ASPM* and *WDR62*, was performed in all microcephalic individuals. Considering the low prevalence of the remaining MCPH genes (*MCPHI*, *CDK5RAP2*, *CENPJ*, *STIL*) as a pathogenic cause, the sequencing of these genes was conducted only on individuals with the lowest OFC (below -4SD).

The large deletions and duplications, which explains less than 10% of all causative pathogenic mutations on MCPH genes, can only be detected with MLPA, however, the MLPA kit for MCPH genes had not yet been available on the market up to the analysis period of this study and therefore not performed.

Investigation for *PTEN* gene mutation was conducted in individuals in the mental retardation study collection with macrocephaly using DNA sequencing. MLPA for the *PTEN* gene can be used in combination with DNA sequencing in order to find exonic and whole gene deletion that could be missed by sequencing, however, MLPA technique for *PTEN* gene is not included in this study.

3.7.3. New primer design and testing

ASPM and *PTEN* gene sequencing were part of the genetic testing services provided by the DNA Diagnostics in the Department of Human Genetics, RUNMC and the primers which were used for this research are those routinely used for diagnostics.

Newly designed primers for the *WDR62*, *MCPHI*, *CDK5RAP2*, *CENPJ* and *STIL* was used (Appendix III). Before being put into use, these new primers were tested with control DNA using gel electrophoresis and DNA sequencing to ensure that they work properly. For the primer design, all forward and reverse primers designed must be 18-20 bases long and must contain 8-10 CG bases. When there is failure of primer to work properly, new primers were designed.

The procedure for primer design and testing was as follows:

- 1) Reference sequence of the gene was noted, including the NM number (The National Center for Biotechnology Information reference sequence accession number). The reference sequence was found through the UCSC genome browser at <http://genome.ucsc.edu/> (UCSC version hg19, release date Feb, 2009).
- 2) From the website, access was gained by filling in the gene name (for example, WDR62), and afterwards accessing the “exon primers” link.
- 3) The following parameters was adjusted online on the UCSC website: minimum distance between primer and exon/ intron boundary is 60 bp; primer region is 70 bp, maximal target (exon) site 500 bp; overlap 50 bp; GC clamp off, annealing temp 60 C. A list of suggested primers can be generated.
- 4) Forward and reverse primers was checked with the complete sequence of the possible overlapping with known single nucleotide polymorphisms (SNP). In the case of SNP overlapping, a new primer sequence was generated.
- 5) Primers were ordered from “Biolegio” in The Netherlands. When available, primer working solution was made from the forward and reverse primers stock for every amplicon (exon segment for amplification) by adding 10 ul of forward primer stock, 10 ul of reverse primer stock, and 80 ul of 0.1X TE buffer.
- 6) The primer working solution was tested on control DNA by performing PCR and running the PCR product on agarose gel electrophoresis. Working primers will show bands on the gel.

3.7.4. DNA amplification (PCR)

For amplification of each gene amplicon with PCR, 100 ng DNA in a working solution of 1 ul was added with PCR mix solution consisting 7.6 ul of 360 PCR mix, 0.5 ul of primers working solution, and 6 ul of MQ.

Amplification was performed using PCR System 9700 (Applied Biosystem) with the following protocol (Table 9):

Table 9. PCR protocol

| PCR | Time | Temp | Cycle |
|---------------------|------------|------|-------|
| Denaturation | 10 minutes | 95°C | |
| Denaturation | 30 seconds | 95°C | |
| Annealing | 30 seconds | 60°C | 35 |
| Elongation | 1 minute | 72°C | |
| Elongation | 7 minute | 72°C | |
| | | 10°C | |

Purification of the PCR product

After amplification with PCR, 15 ul of DNA product was put on filter membrane trays for purification, added with 25 ul of MQ. The tray was vacuumed until all the liquid are filtered. Another 40 ul of MQ was added, and the vacuum procedure was repeated. The purified product will then be diluted in 40 ul of TE 0.1x buffer and stored in 96 wells trays.

Gel electrophoresis

Gel electrophoresis was performed to confirm the presence of amplification product before they were sent for sequencing using the ABI 3730 XL sequencer (Applied Biosystem).

Gel electrophoresis was performed with the following procedure:

1. Agarose solution 1% was prepared by adding 5 grams of agarose in 500 ml of TBE. Firstly, only 300 ml of TBE and 5 gr of agarose was mixed in Erlenmeyer flask. The solution was mixed evenly and then put in the microwave for 4 minutes until it boils. Afterwards the flask was taken out. More TBE was added until the solution volume reached 500 ml. NANCY, the fluorescent stain is used for visualization of PCR product on gel, should be added and mixed evenly.
2. The gel tray was prepared by tightening the rubber side and inserting the combs for making the wells. Vertical distance between the wells needs to be the same length.
3. Gel solution was poured on the tray, and air bubbles must be removed from the wells. The solution was left form solid agar. After the solid agar was formed, it should be moved to the electrophoresis container and ready for use.
4. The purified product (5 ul) was added with 5 ul of orange G loading buffer, and the total volume of 10 ul was inserted into the wells on the solid agar with the markers (100bp marker) on each side of the well rows.
5. The electrophoresis machines should be turned on and the voltage was set to 200 volts. The process should be checked again after about 10-20 minutes. The electrophoresis should run until the product went halfway between the well rows.
6. When the electrophoresis is finished, the gel should be taken out, and picture of the gel was taken.

3.7.5. Sequencing

After confirmation of the presence of PCR product with gel electrophoresis, the purified PCR product was sent for sequencing. One microlitre of purified PCR product was mixed with M13 forward and reverse primer and later was sent for sequencing on 3730 XL sequencer (Applied Biosystem).

3.7.6. Sequencing Data analysis

The DNA sequencing data was analysed with SEQPilot software. From the software, data of the variants was obtained in the form of base changes in the DNA sequence.

The DNA sequencing data was analysed with SEQPilot software version 3.2.1.0. Variants was further analysed using Alamut 2.0 mutation interpretation software. The impact of protein alteration was calculated using Align-GVGD method, SIFT, and Polyphen computational approaches, as an accepted mathematical and statistical method of molecular genetics investigation.

After obtaining the protein prediction information, the found variants were then classified based on the guideline by the UK Molecular Genetics Society (CMGS) and Dutch Society of Clinical Genetic Laboratory Specialists (Vereniging Klinisch Genetische Laboratoriumspecialisten). The components analyzed include the following points:

- a. Whether the variant was previously described in the literature or in the mutation database containing information on the variant at RNA and protein level, as well as clinical interpretation of the named mutation. The core databases are HGMD, OMIM, dbSNP/Ensembl, and Pubmed.
- b. Test of matched controls
- c. Any co-occurrence with a known deleterious mutation. However, parental sample is required in these cases.
- d. Co-segregation of disease within the family
- e. New variant concurrent with the sporadic incidence of the disease.
- f. Prediction of pathogenic effect based on protein prediction calculations.
- g. Presence or absence in SNP databases

The variants was classified into:

- 1) Class 1 (UV1): Certainly non pathogenic
- 2) Class 2 (UV2): Unlikely to be pathogenic

- 3) Class 3 (UV3): Likely to be pathogenic
- 4) Class 4 (UV4): Certainly pathogenic

3.7.8. Identifiler test

Identifiler test is a direct PCR Amplification Kit is a short tandem repeat (STR) multiplex that allow direct amplification. This kit tests genetic variants of STR at 15 loci. STR are multiple copies of identical DNA sequence arranged consecutively in a particular region of a chromosome. Each individuals with a unique DNA sequence will have a unique number of STR in each loci. This test is used to obtain information about family relationships, for example homozygosity status of twins or paternity and maternity status.

The kit used in this study was AmpFISTR Identifiler Direct PCR Amplification Kit from Applied Biosystem. The laboratory procedures followed the manufacturer's manual (Life Technologies, 2011).

3.7.9. Data analysis

After genetic analysis of the variants, the results was presented in tables and figures. Subjects with variants classified as "likely to be pathogenic" received follow-up visit and had the family members' DNA taken for follow-up molecular genetic examination. Results which requires further description was discussed further in detail.

3.7.10. Research ethics

This study is a continuation of a mental retardation study that had been started in 2008, and ethical clearance had therefore been given as a part of this previous study (Mundhofir, 2008) (Appendix V).

3.8. Research Flowchart

