CHAPTER II

LITERATURE REVIEW

Mental retardation (MR) is often found together with abnormality of the head circumference, microcephaly or macrocephaly, as an associated sign. Both microcephaly and macrocephaly increases the risk for mental retardation, and there is a correlation between the degree of head circumference abnormality and the severity of cognitive impairment. Microcephalic individuals are more likely to have MR, while MR individuals are at least twice more likely to have macrocephaly compared to intellectually normal peers (Field, 2007; Ashwal, 2009). There is a high likelihood that mental retardation and abnormality of head circumference is caused by the same explanation, for example, a genetic mutation that is causing both mental retardation and microcephaly or macrocephaly (Abuelo, 2007; Olney, 2007).

2.1. Microcephaly

2.1.1. Definition and etiology

Microcephaly is the reduced occipitofrontal circumference of the head smaller than -2 standard deviation. The incidence of microcephaly at birth is between 1.3 and 150 per 100,000 live births. The rate of the incidence depends on the population and the threshold in defining microcephaly (Kaindl, 2010).

Head circumference is determined by brain size and postnatal head growth is considered an important clinical indicator of brain growth (Cheong, 2008). Comparisons of head circumference from different countries show that its growth may differ in population of one country to another (Ayatollahi, 2006), and that the smaller head circumference of some ethnicities reflects the smaller body stature (Tsuzaki, 1990).

Microcephaly is determined using measurement of occipito-frontal circumference (OFC) by placing a measuring tape around the cranial vault which includes the widest part of the forehead and the most prominent part of the occipital area (Abuello, 2007). The widely accepted definition of a ‘‘normal’’ OFC measurement includes observations within 2 standard deviations from the mean (Rollins 2010). Several standard charts are available for monitoring of head circumference, such as those from the CDC, WHO, Fells, United States Head Circumference Growth Reference Charts, and Nellhaus (Roche, 1987; Rollins, 2010).

The choice of using the Nellhaus head circumference charts is often made based on their frequent use in clinical settings, the age range covered by these charts, and their mixed racial population sample (Miles, 2000). The Nellhaus charts, compiled from 14 different studies of head circumference, are used for children from 0 to 18 years old (Nellhaus, 1968). As mentioned before, studies in different countries revealed that head charts from different countries have their differences in what is considered a normal range of head circumference. For example, the head circumference values of Turkish boys were similar to those of Japanese and Irish children, in girls they were closer to the values of Irish and English girls (Karabiber, 2001); while a study by Ayatollahi in Iran (2006) showed that head circumference of Iranian boys were lower than their counterparts in Turkey, Japan, Ireland and USA. The closest available chart to Indonesian stature is the Singaporean chart, however, it is only available for children age 0-6 years (National Healthcare Group Polyclinics, 2000).

Recently, a preliminary study conducted in Semarang, Central Java (Mundhofir, unpublished) indicated that the mean value of head circumference of normal schoolchildren age 7-12 years is smaller compared to the normal range of the Nellhaus charts.

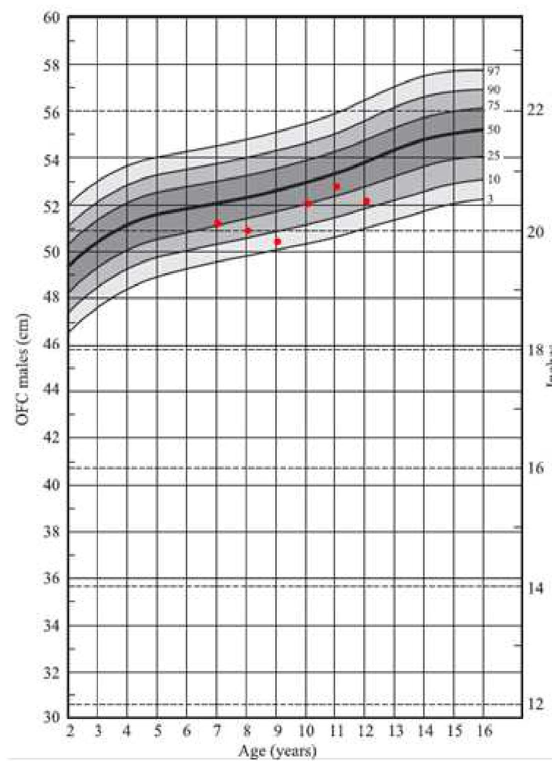

Figure 1. Head chart plotting of Semarang normal boys. The red dots show the mean OFC of boys between age 7 and 12 years old.

The boys mean values are between the 3rd and under 50th percentile (Figure 1). while the girls are between the 10th and under 50th percentile (Figure 2) (Mundhofir, unpublished).

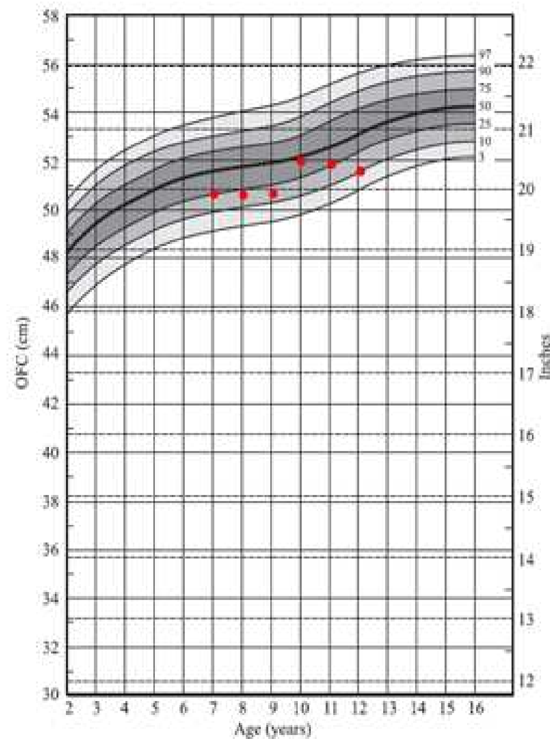

Figure 2. Head chart plotting of Semarang normal girls. The red dots show the mean OFC of boys between age 7 and 12 years old.

The interpretation of head circumference must be considered using the specific head charts appropriate for application in the different countries. It has also been suggested that the head circumference standard applied in each country should be reevaluated and updated periodically (Karabiber, 2001).

Microcephaly can be distinguished into primary microcephaly, and secondary microcephaly. Primary microcephaly, also called congenital microcephaly, is considered a static developmental anomaly, present at birth or as early as 32 weeks of gestation (Woods, 2004). Secondary microcephaly, or acquired microcephaly, is a progressive neurodegenerative condition, where the head circumference at birth is within normal range but later fall below it. There are several nongenetic and genetic causes of primary microcephaly with mental retardation, such as congenital toxoplasmosis, maternal alcohol overconsumption at pregnancy and Rubinstein Taybi syndrome (Woods, 2005). Some of the environmental and genetic causes of microcephaly are listed below (Table 2) (Abuelo, 2007).

Table 2. Causes of microcephaly (adapted from Abuelo, 2007; Jacob, 2009; Mahmood, 2011)

Environmental and genetic causes of microcephaly
Environmental causes of microcephaly
Hypoxic-ischemic encephalopathy
Intrauterine infections
Teratogen (alcohol, hydantoin, radiation)
Intrauterine growth retardation (IUGR)
Neural tube defects (NTD)
Maternal Phenylketonuria (PKU)
Poorly controlled maternal diabetes
Chronic malnutrition
Genetic causes of microcephaly
Primary hereditary microcephaly (MCPH)
Chromosome abnormalities (trisomy 13, trisomy 18, trisomy 21)
Deletion syndrome:
4p deletion (Wolf-Hirschhorn) syndrome
7q11.23 deletion (Williams) syndrome
Rett Syndrome
Other syndromes with multiple anomalies
Cornelia de Lange syndrome
Smith Lemli Opitz syndrome
Holoprosencephaly

2.1.2. MCPH

Autosomal recessive primary microcephaly, or also known as primary hereditary microcephaly (MCPH) is a neurodevelopmental disorder characterized by microcephaly present at birth and mental retardation. Its incidence is between 1:30,000 and 1:2,000,000 in non-consanguineous populations, in comparison with 1:10,000 in consanguineous Pakistani families (Mahmood, 2011). The current clinical definition for MCPH is as follows (Woods, 2005):

1. Congenital microcephaly
2. Mental retardation without other neurological signs such as spasticity or progressive cognitive decline. Seizures are unusual but do not exclude the diagnosis
3. For most MCPH patients, normal height, weight, appearance, chromosome analysis and brain scan are observed. In patients with MCPH1 mutations, reduction in height may be found but head circumference was significantly reduced compared to height.

In MCPH, the head circumference is reduced with reduction in the size of the cerebral cortex, especially in the surface area, showing a simplified gyral pattern (Desir, 2008). Table 3 shows the features of MCPH that may be found clinically.

Table 3. Overview of MCPH features (Adapted from Kaindl, 2010).

Main features	• Microcephaly (-2SD) at birth, further relative reduction in the first years of life • Reduction of cerebral cortex volume • Simplification of gyral pattern • Mild to severe mental retardation (normal IQ possible)
Inconsistent features	
• Neurological/neuropsychological	• Delay of motor milestones • Pyramidal sign • Speech delay • Hyperactivity and attention deficit • Aggressiveness • Sleep disorder • Seizures
• Further intracranial malformations	• Agenesis of corpus callosum • Focal dysplasia • Focal microgyria • Dysmorphic or large lateral ventricles • Reduction of white matter
• Endocrinology	• Short stature • Early puberty

2.1.3. Pathogenesis of MCPH

Brain size at birth is determined by the rate of proliferation and cell death during neurogenesis. There are two types of proliferative division of the neuroepithelial (NE) cells progenitors, symmetrical and asymmetrical division. In symmetrical division, the NE cells divide to produce daughter cells which will adopt a NE cell fate. In asymmetrical division, only one daughter cell remains an NE cell, while the other one become a basal progenitor or neuron (Figure 3).

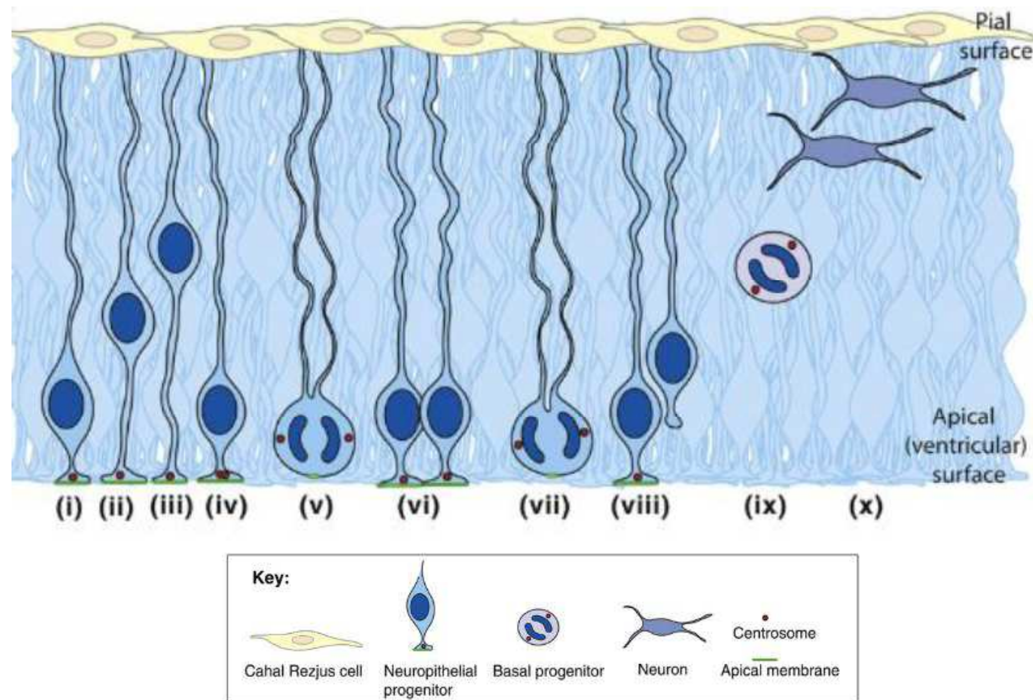

Figure 3. Development of mouse neuroepithelium. Neuroepithelial cells nuclei migrate (i,ii) to basal surface during G1 where they undergo S phase (iii). NE cells then migrate apically at G2 (iv). The centrosomes (red circles) stay at the apical membrane (green). Mitosis happens at the apical surface; the centrosomes form the spindle poles. Symmetrical division results in two identical neuroepithelial cells (v, vi) while asymmetrical division (vii) produce one neuroepithelial cell and one daughter cell which detaches and may either be a basal progenitor (ix) or neuron (x) (Adapted from Thornton, 2009).

The amount of rounds of symmetric proliferative division is considered important to determine the final number of neurons formed. The more symmetrical division occur, the more neurons was produced. For example, the study by Kaindl (2010) found that one NE progenitor with 10 rounds of asymmetric division produced 10 neurons, while when all but the last division were symmetrical, 512 neurons were produced. Spindle positioning will determine the cleavage process of NE cells. Any disruption of this process will cause a premature shift from symmetrical to asymmetric cell division, which reduces neuron production and lead to a MCPH-like phenotype. Mutations in the

MCPH genes affect fetal neurogenesis by affecting this shift towards asymmetrical division of neuron progenitors (Kaindl, 2010).

2.2. MCPH Genes

Currently, the known genetic causes of MCPH are mutations in 7 loci, with 6 identified genes: *MCPH1*, *WDR62* (MCPH2), *CDK5RAP2* (MCPH3), *ASPM* (MCPH5), *CENPJ* (MCPH6) and *STIL* (MCPH7) (Table 4). As MCPH is inherited in an autosomal recessive manner, both alleles have to be mutated, either that an individual has a homozygous mutation or two compound heterozygous mutations in order to result in affected phenotype.

Table 4. MCPH genes and function (Adapted from Thornton, 2009)

MCPH	Protein	Proportion of MCPH	Gene	Chromosome	Function
MCPH1	Microcephalin	<5%	*1.1 MCPH1*	8p23	DNA damage repair; chromosome condensation; transcriptional regulation of DNA damage genes
MCPH2	WD repeat domain 62	NA	*1.2 WDR62*	19q13.12	Proliferative division of neural precursors and neuronal migration
MCPH3	Cyclin-dependent kinase 5 regulatory associated protein 2	<5%	*1.3 CDK5RAP2*	9q33.3	Predicted role in regulating microtubule dynamics, PCM recruitment and stabilization; centrosome maturation and cohesion
MCHP4	-	NA		15q15-q21	Unknown
MCPH5	Abnormal spindle like microcephaly associated	37% - 54%	*ASPM*	1q31	Spindle pole organisation and orientation
MCPH6	Centromeric protein J	<5%	*CENPJ*	13q12.2	Centriole biogenesis and length; microtubule dynamics
MCPH7	SCL/TAL1 interrupting locus	<5%	*STIL*	1p32	Spindle organisation

NA: Not available

MCPH1

MCPH1 is located on human chromosome 8p23, encoding 835 amino acids. It has 14 exons, spanning 237 kb in the human genome (Jackson, 1998; Jackson, 2002). *MCPH1* gene encodes the protein microcephalin consisted of 835 amino acids. Microcephalin mRNA is expressed in human fetal tissue especially in the brain, liver and kidney. In murine brain, microcephalin was found to be expressed during neurogenesis especially in the proximity of lateral ventricles which holds progenitor cells that produce neuron which later form the cerebral cortex (Jackson, 2002). The MCPH phenotype may be caused by defect in the cell cycle checkpoint control and DNA repair, as well as defective centrosomal function affecting the number of symmetric cell divisions of neuronal stem cells during neurogenesis. The pathologic variants include homozygous nonsense mutations, homozygous deletions and homozygous duplications (Kaindl, 2010). The p.Pro828Ser normal variant in a Chinese population was associated with variation in cranial size (Wang, 2008).

WDR62 (MCPH2)

Recently, the MCPH2 gene was identified to be *WDR62*, the second most common cause of MCPH. *WDR62* is a gene with 32 exons and consists of 1523 amino acids. The six described mutations include four missense mutations and two duplications in 3 families. (Nicholas, 2010).

The protein produced by *WDR62* gene encodes a spindle pole protein that is expressed in neuronal precursor cells of mammalian embryonic neuroepithelium undergoing mitosis. It is thought that *WDR62* codes the key protein in the positioning of spindle poles and in prolonging the neural precursor generation that is vital to cerebral cortex growth (Nicholas 2010; Bilguvar 2010).

***CDK5RAP2* (MCPH3)**

CDK5RAP2 comprises 38 exons and spans approximately 191 kb. The human cyclin-dependent kinase 5, regulatory associated protein 2 gene encodes *CDK5RAP2* with 1893 amino acid. *CDK5RAP2* is a centrosome-associated protein, and its mRNA is widely expressed in human and in embryonal mouse tissue In murine embryos, highest levels were detected in the central nervous system (Bond et al., 2005).

Studies of *CDK5RAP2* homologues show that they regulate centrosome maturation, recruitment and strengthening of pericentriolar matrix at the centrioles and regulate centrosome cohesion. Loss of function causes failure of centrosome to mature and efficiently organize microtubules. This defect in humans may lead to spindle positioning abnormality that cannot be tolerated in neuroepithelial progenitors (Kaindl, 2010). Furthermore, *CDK5RAP2*, as well as MCPH1 was found to be essential in spindle checkpoint activation, indicating that spindle checkpoint mechanism controls the number of neurons generated by neural precursor cells (Zhang, 2009). Spindle positioning is crucial during asymmetrical division to generate daughter cells with different sizes or fates (McCarthy, 2006).

The three described mutations in the *CDK5RAP2* gene predict a premature stop codon in 4 Pakistani families, suggesting in a loss of functional protein activity that causes the MCPH phenotype (Bond, 2005; Hassan, 2007).

***ASPM* (MCPH5)**

Mutations in the *ASPM* gene on chromosome 1q31.3 have been considered the most common cause of MCPH and counts for half of Asian and European MCPH cases (Nicholas, 2009). The *ASPM* gene spans 62kb with 28 exons. *ASPM* gene is expressed in many human embryonic (e.g., liver, kidney, heart, lung, brain) and adult (e.g., breast, lung, pancreas, uterus, thyroid, liver, ovary, testis) tissues (Bond, 2002).

The *ASPM* gene encodes the abnormal spindle-like microcephaly-associated protein consisting of 3477 amino acids. The *ASPM* protein is highly expressed in progenitor cells and regulates the positioning and organization of spindle for proliferation. Expression is high on onset of neurogenesis at the early stage, and decreases as neurogenesis progresses. *ASPM* therefore determines the orientation of the cell division to be either symmetrical or asymmetrical during neurogenesis and important in maintaining the proliferative capacity of neuroepithelial cells progenitor (Thornton 2009).

The pathologic variants of the *ASPM* gene include translocation, deletion, insertion/ duplication or base substitution. The mutations were spread throughout the genes with no mutation hot-spot. The disease mechanism causing MCPH is thought to be nonsense mediated decay of the *ASPM* mRNA leading to significant reduction of the protein in neuroepithelial cells during neurogenesis (Nicholas, 2009).

***CENPJ* (MCPH6)**

The human centromeric protein J (*CENPJ*) gene spans 40 kb and comprises 17 exons. *CENPJ*, consisting of 1338 amino acids, plays a role in the control of centrosome and spindle function during neurogenic mitosis (Bond et al 2005; Kaindl et al 2010).

CENPJ is an essential regulator of centriole length during its biogenesis and functioning of centriolar microtubules. Loss of centrioles causes spindles deformation and DNA segregation defects. Loss of *CENPJ* could lead to MCPH phenotype through lack of mature centrosomes, improper generation of microtubules and impairment of spindle positioning. This in turn may lead to deviation in cleavage plane alignment in cell division (Thornton, 2009). The described pathogenic mutations include 2 homozyhous deletions and one homozygous missense mutation (Bond et al, 2005).

***STIL* (MCPH7)**

STIL, SCL/TAL1 interrupting locus gene is consisted of 1287 amino acids. *STIL* homozygous mutations causing loss of function were very recently identified as the cause of MCPH in four of 24 consanguineous Indian familiesRecent data suggest that *STIL* has similarities in function with *ASPM*; and the phenotype of MCPH was hypothesized to be caused by impairment in spindle alignment because of *STIL* mutation. The described pathogenic mutations of *STIL* are nonsense mutation, frameshift deletion and intronic splice mutation resulting in truncated protein (Kumar, 2009).

2.3. Macrocephaly

2.3.1. Definition and aetiology

Macrocephaly refers to an abnormally large head which includes the scalp, cranial bone and intracranial contents. In macrocephaly, the measured occipitofrontal circumference (OFC) is greater than the 2 standard deviation (Williams, 2008).

Macrocephaly can be distinguished as syndromic and non-syndromic. Non-syndromic macrocephaly refers to conditions where the enlarged brain is the predominant abnormality and not associated with any other physical abnormality or major malformation. Minor craniofacial changes may be present secondary to the enlargement of cranial vault. Syndromic macrocephaly means that physical or behavioral abnormalities are associated with the macrocephaly and the combination of these abnormalities can be recognized as a syndrome (Olney, 2007). Table 5 (Williams, 2008) shows the more commonly encountered conditions and not a complete list of all the genetic disorders associated with macrocephaly.

Table 5. Classification of macrocephaly conditions (Williams, 2008)

Causes of macrocephaly
Genetic causes of macrocephaly
Familial macrocephaly
Benign asymptomatic
Autism disorder (multifactorial, non-syndromic)
Syndrome associations
Bannayan –Riley –Ruvulcaba syndrome/ Cowden syndrome
Neurofibromatosis type 1
Sotos syndrome
Cardiofaciocutaneous syndrome
Costello syndrome
Fragile X
Metabolic types
Alexander disease
Canavan disease
Tay-Sachs disease
Bone dysplasia
Achondroplasia
Osteopetrosis
Non-genetic causes of macrocephaly
Hemorrhage
Infections
Post-traumatic subdural effusions
Arachnoid cyst

Children with macrocephaly should be examined for detailed history including family, developmental and growth history and physical examination for any other dysmorphic features or anomalies. For syndromic macrocephaly, associated manifestations may lead to diagnosis. Therefore, search for other congenital abnormalities should be conducted such as abdominal ultrasonography, echocardiogram, and ophthalmologic examination. When available, cytogenetic studies and molecular genetic studies should be performed as indicated (Olney, 2007).

According to Field (2007), individuals with mental retardation are more likely to have macrocephaly than intellectually normal peers. Although the neuropathological basis of macrocephaly is not well understood, the presence of

this feature may be helpful in restricting the differential diagnosis in individuals with mental retardation (Field, 2007). Table 6 shows some of the recognizable syndromic conditions and their identified genes where macrocephaly is a predominant clinical feature. However, only *PTEN* gene causes both mental retardation and macrocephaly (Barker, 2003).

Table 6. Macrocephaly syndromes with causative genes (Williams, 2008)

Syndrome/disease	**↑ OFC major clinical finding**	**↑ OFC minor clinical finding**	**Identified genes**	**Inheritance pattern**
Familial macrocephaly				
Benign asymptomatic	+		-	MF
Autism disorder (multifactorial, non-syndromic)		+	-	MF
Syndrome associations				
Bannayan –Riley – Ruvulcaba syndrome/ Cowden syndrome	+		*PTEN*	AD
Neurofibromatosis type 1		+	*NF1*	AD
Sotos syndrome	+		*NSD1*	AD
Cardiofaciocutaneous syndrome		+	*KRAS, BRAF, MEK1, MEK2*	AD
Costello syndrome		+	*HRAS*	AD
Fragile X		+	*FMR1*	XLD
Metabolic types				
Alexander disease	+		*IDS*	AR
Canavan disease	+		*IDUA*	AR
Tay-Sachs disease	+		*HEXA*	AR
Bone dysplasia				
Achondroplasia	+		*FGFR3*	AD
Osteopetrosis		+	*TCIRG1, CLCN7, OSTM1, TNFSF11*	AR, AD

MF: multifactorial disease, AD: Autosomal dominant, XLD: X-Linked, AR: Autosomal recessive

Williams (2008) published a scheme for macrocephaly evaluation. Physical examination and history may found a syndromic disorder that can later

be confirmed by an appropriate and available tests. When there is no neurological dysfunction, the possibility of familial macrocephaly should be considered. However, when there are developmental concerns, a brain MRI is usually performed. When there is no informative result from MRI, tests such as chromosome study, array-CGH and fragile X molecular screening are often performed for assessment in the presence of developmental delay and/or dysmorphism together with macrocephaly. In developed countries, MRI/CT scanning is a likely option for routine diagnosis, however this may be difficult to apply in developing countries (Figure 4).

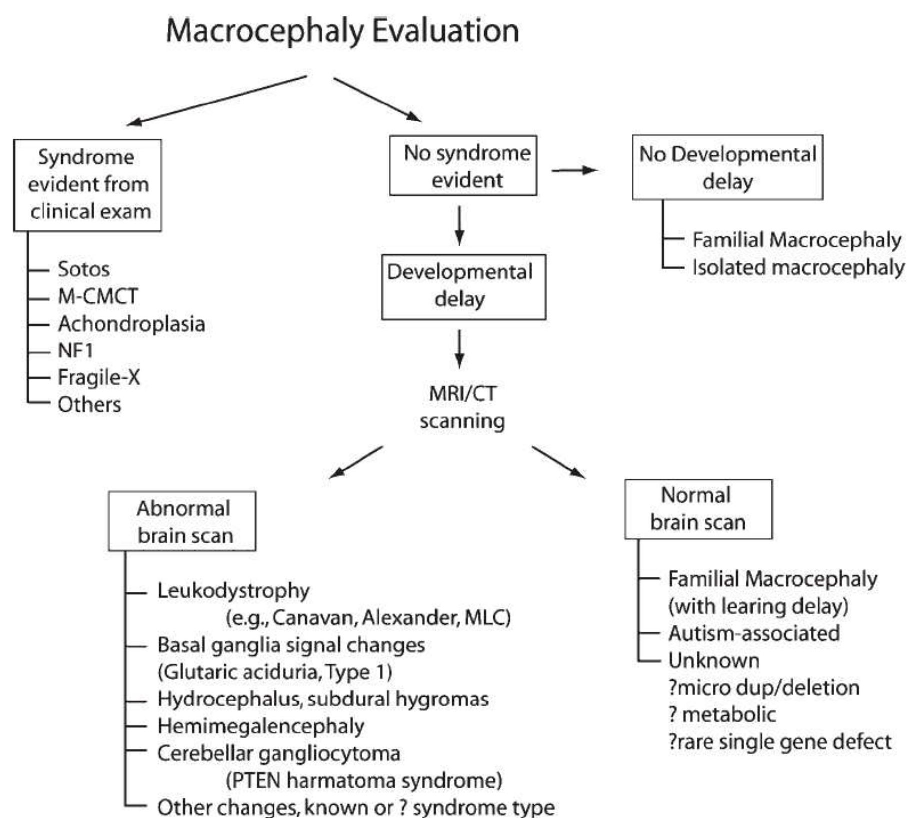

Figure 4. Macrocephaly evaluation scheme. MLC: Megalencephalic leukoencephalopathy with subcortical cysts; M-CMTC: macrocephaly cutis marmorata telangiectatica congenital (Williams, 2008).

The brain enlargement found in autism patients has been speculated to occur caused by the excess of synaptic connections. However, there is increasing evidence of the association between autism and excessive brain growth (Eigsti, 2003; Frith, 2003).

The finding of *PTEN* mutations in autism or MR patients with macrocephaly subsequently resulted in the recommendation of *PTEN* testing in these patients as identification of mutation may impact on genetic counseling in recurrence risk and medical management of patients (Orrico, 2008; Varga, 2009). Moreover, *PTEN* testing is relevant for these patients since patients with *PTEN* mutations are advocated to undergo cancer screening according to the *National Comprehensive Cancer Network* (*NCCN*) guidelines.

2.3.2. PTEN

The phosphatase and tensin homologue (*PTEN*) gene on 10q23.3 is a tumor suppressor gene that functions as a dual-specificity phosphatase involved with cellular growth. It functions primarily as a lipid phosphatase down-regulating the phosphoinositol 3-kinase/AKT pathway. This pathway is critical in normal cell proliferation and organ development, and its dysfunction has been associated with diseases characterised by overgrowth (Varga 2009; Barker, 2003). This phospatase is also found to affect cell migration and important for the maintenance of chromosomal integrity through association with centromere proteins (Li, 1997).

PTEN mutations predispose to diverse phenotypes with different disorders that have overlapping clinical features which includes mental retardation. These disorders are collectively called *PTEN* hamartoma tumour syndrome (PHTS) which encompasses Cowden syndrome, Bannayan–Riley–Ruvalcaba syndrome, Proteus syndrome and Proteus-like syndrome (Orloff, 2008; Hobert, 2009). Cowden disease in particular is a macrocephaly syndrome associated with symptoms of autism (Goffin, 2001).

Several reports have shown germline *PTEN* mutations in patients with autistic behaviours or mental retardation with the presence of macrocephaly (Zori et al 1998; Goffin et al, 1998; Parisi et al, 2001; Delatycki et al, 2003; Butler et al, 2005). On the basis of these, Varga et al (2009) performed *PTEN* genetic testing in 114 patients referred with autism spectrum disorder; from 49 patients with MR (32 patients with both MR and macrocephaly), 6 patients (12,2%) were found to have germline *PTEN* mutations (Varga, 2009).

Gene targeting studies show that mutation in *PTEN* gene leads to increase in cell numbers, decrease in cell death, and enlargement of neuronal cell size and subsequently macrocephaly (Groszer et al, 2001; McCaffery, 2005). A study showed that there is brain overgrowth in Pten haploinsufficient mice of both sexes and that there are deficits in social approach behavior of the female mice (Page, 2009).

2.4. Analysis and interpretation of genetic variants

2.4.1. General interpretation

Molecular genetic testing by DNA sequencing will find variations within the DNA sequence. Some of the detected sequence variants can easily be recognized as pathogenic mutations, but for others the pathogenicity is still unclear (also known as "variants with unknown pathogenicity"). Consequently, determination of the significance of genetic variants is needed (Bell, 2007).

DNA sequencing data is analyzed with sequence analysis software which will report a list of variants that are either pathogenic, non-pathogenic, or of unknown pathogenicity. The variants with unknown pathogenicity must be further analyzed and classified using a set of criteria. Most European laboratories use the international guidelines by UK Clinical Molecular Genetics Society and the Dutch Society of ClinicalGenetic Laboratory Specialists (Bell, 2007). Since functional studies on the proteins are usually time-consuming and not available in DNA

diagnostic laboratories, the classification of unclassified variants is usually predicted by several computer programs. Various parameters can be used to detect the structural or functional effects of amino acid substitutions, a method using mathematical calculation also known as protein prediction (Wang, 2001). The risk in the pathogenic effect of non-synonymous changes of the DNA sequence may be calculated with several bioinformatics methods, such as Grantham scoring, Align-GVGD, SIFT and Polyphen.

2.4.2. Grantham score and Align-GVGD

Grantham (1974) introduced the formula for calculating the difference between amino acids taking into account their composition (C), polarity (P) and molecular volume (V). This will in turn show how much difference exists between amino acids (Mathe, 2006).

The GVGD calculation is derived from the original Grantham formula (Tavtigian, 2005; Grantham, 1974):

$$50.723 \times \sqrt{[1.833(Ci - Cj)^2] + [0.1018(Pi - Pj)^2] + [0.000399(Vi - Vj)^2]}$$

C: Composition
P: Polarity
V: Molecular volume
i: Original amino acid
j: Mutated amino acid

The C, P and V values for the 20 common amino acids were previously determined and provided with the Grantham definition of the formula (Table 7).

Table 7. Amino acid properties for Grantham calculations (Grantham, 1974).

Amino acid	Property		
	c	*p*	*v*
Ser	1.42	9.2	32
Arg	0.65	10.5	124
Leu	0	4.9	111
Pro	0.39	8.0	32.5
Thr	0.71	8.6	61
Ala	0	8.1	31
Val	0	5.9	84
Gly	0.74	9.0	3
Ile	0	5.2	111
Phe	0	5.2	132
Tyr	0.20	6.2	136
Cys	2.75	5.5	55
His	0.58	10.4	96
Gln	0.89	10.5	85
Asn	1.33	11.6	56
Lys	0.33	11.3	119
Asp	1.38	13.0	54
Glu	0.92	12.3	83
Met	0	5.7	105
Trp	0.13	5.4	170

C: composition index, P: polarity index, V: molecular volume.

The original Grantham scores are provided in a table. The numbers show the scoring difference between the amino acid pair (vertical and horizontal axis), between 0 and 215. The higher Grantham score refers to bigger chemical difference between the amino acids (Table 8).

Table 8. The Grantham score table based on amino acid change (Grantham, 1974).

Arg	Leu	Pro	Thr	Ala	Val	Gly	Ile	Phe	Tyr	Cys	His	Gln	Asn	Lys	Asp	Glu	Met	Trp	
110	145	74	58	99	124	56	142	155	144	112	89	68	46	121	65	80	135	177	Ser
	102	103	71	112	96	125	97	97	77	180	29	43	86	26	96	54	91	101	Arg
		98	92	96	32	138	5	22	36	198	99	113	153	107	172	138	15	61	Leu
			38	27	68	42	95	114	110	169	77	76	91	103	108	93	87	147	Pro
				58	69	59	89	103	92	149	47	42	65	78	85	65	81	128	Thr
					64	60	94	113	112	195	86	91	111	106	126	107	84	148	Ala
						109	29	50	55	192	84	96	133	97	152	121	21	88	Val
							135	153	147	159	98	87	80	127	94	98	127	184	Gly
								21	33	198	94	109	149	102	168	134	10	61	Ile
									22	205	100	116	158	102	177	140	28	40	Phe
										194	83	99	143	85	160	122	36	37	Tyr
											174	154	139	202	154	170	196	215	Cys
												24	68	32	81	40	87	115	His
													46	53	61	29	101	130	Gln
														94	23	42	142	174	Asn
															101	56	95	110	Lys
																45	160	181	Asp
																	126	152	Glu
																		67	Met

Arg: Arginine, Leu: Leucine, Pro: Proline, Thr: Threonine, Ala: Alanine, Val: Valine, Gly; Glysine, Ile: Isoleucine, Phe: Phenylalanine, Tyr: Tyrosine, Cys: Cysteine, His: Histidine, Gln: Glutamine, Asn: Asparagine, Lys: Lysine, Asp: Aspartat, Glu: Glutamic acid, Met: Methionine, Trp: Tryptophan

Align-GVGD is an extended method that combines the Grantham Variation (GV) and Grantham Deviation (GD) scores. GV measures the degree of biochemical variation among amino acids found at a certain position in the multiple sequence alignment while GD reflects the biochemical distance of the mutant amino acid at a particular position. The use of multiple sequence alignment for GV and GD calculations relies on the basic logic that missense substitution in a conserved sequence, as well as missense substitution outside the sequence variations that is evolutionary tolerated, tends to alter protein function.

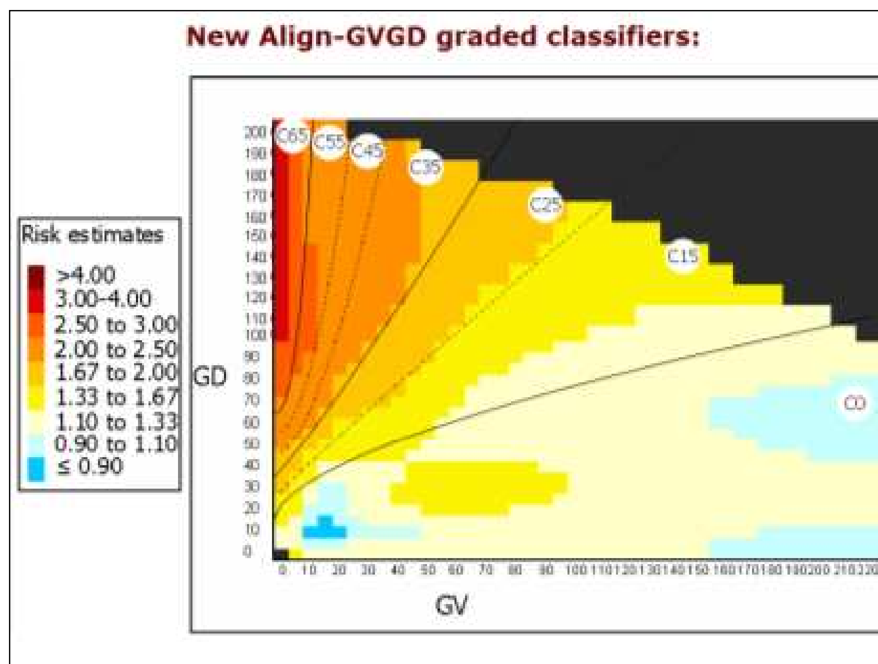

Figure 5. Align-GVGD classifiers. Using a mathematical model, GV and GD can measure the biochemical variations between the amino acids. Based on the combined GV and GD scores, missense mutations are classified between C0 to C65, interpreted as having the least to the most likelihood of interfering with function (Tavtigian, 2005).

In practice, Align-GVGD is a computational calculation that can be accessed on http://agvgd.iarc.fr/index.php.

2.4.3. Sorting Intolerant From Tolerant' (SIFT)

The 'Sorting Intolerant From Tolerant' (SIFT) uses sequence homology to predict whether the amino acid substitution will affect protein function and phenotype. SIFT prediction method assumes that important amino acid was conserved in the protein family and changes at the well conserved position tends to be predicted as deleterious. A highly conserved position is generally intolerant to substitution, while a poorly conserved position is more tolerant. For example, if a conserved position contains the hydrophobic amino acid of isoleucine, valine

and leucine, then the SIFT assumes that this position can only have the amino acids having a hydrophobic trait. Therefore when a substitution occurs, the substitution to a hydrophobic acid was predicted as tolerated, while changes to others (for example, to a polar amino acid) in this conserved position was predicted to affect protein function (Ng, 2003).

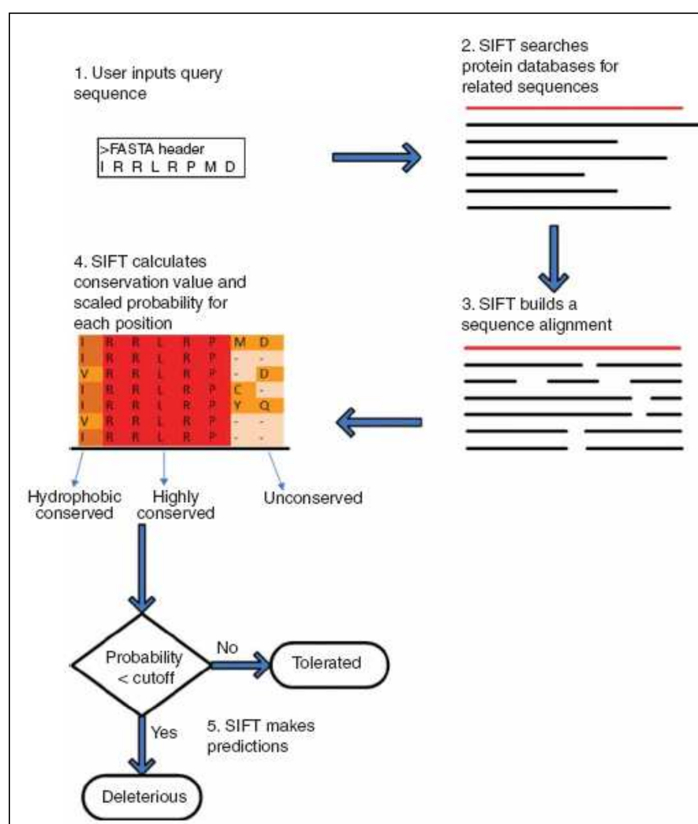

Figure 6. 'Sorting Tolerant From Intolerant' (SIFT) algorithm. The process flow is described for the amino acid sequence 'IRRLRPMD'. I: Isoleusine, R: Arginine, L: Leucine, P: Proline, M: Methionine, D: Aspartic acid (Kumar, 2009).

SIFT scores are produced by computerized calculations of position-specific scoring matrix (Kumar, 2009). Results are predicted as either "damaging"or "tolerated". "Damaging" means that the missense substitution is

likely to affect protein function while "tolerated" means that the substitution is likely to be neutral. SIFT prediction is available through http://sift.jcvi.org/.

2.4.4. PolyPhen-2

Polymorphism phenotyping (PolyPhen) is a sequence-based and structure-based method in predicting the functional effect of a mutation. Polyphen prediction relies on several strategies: (1) characterization of the substitution site, (2) calculation of position-specific independent count (PSIC), (3) mapping of the substitution to a known 3D protein structure to reveal whether the replacement has the likelihood to disturb the protein structure, or whether it tends to disrupt the protein's interactions or contacts with critical sites, (4) structural analysis by calculation of surface area, residue volume, and loss of hydrogen bond (Ramensky, 2002).

PolyPhen-2 calculates the likelihood that a mutation influences protein function, phenotype and fitness; whether it is damaging or benign. A mutation is then classified as benign, possibly damaging or probably damaging. Figure 7 shows the different steps involved in the prediction and interpretation of nonsynonymous SNP. The web interface is available at http://genetics.bwh.harvard.edu/pph2/.

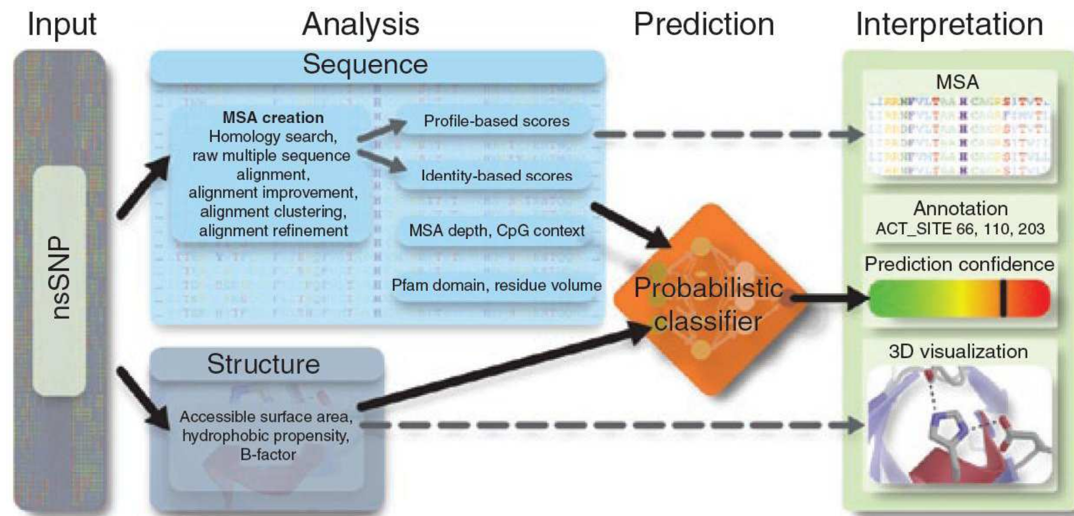

Figure 7. Scheme of Polyphen-2 prediction method. nsSNP: nonsynonymous single nucleotide polymorphism, MSA: multiple sequence alignment (Ramensky, 2002).

2.5. Theoretical Framework

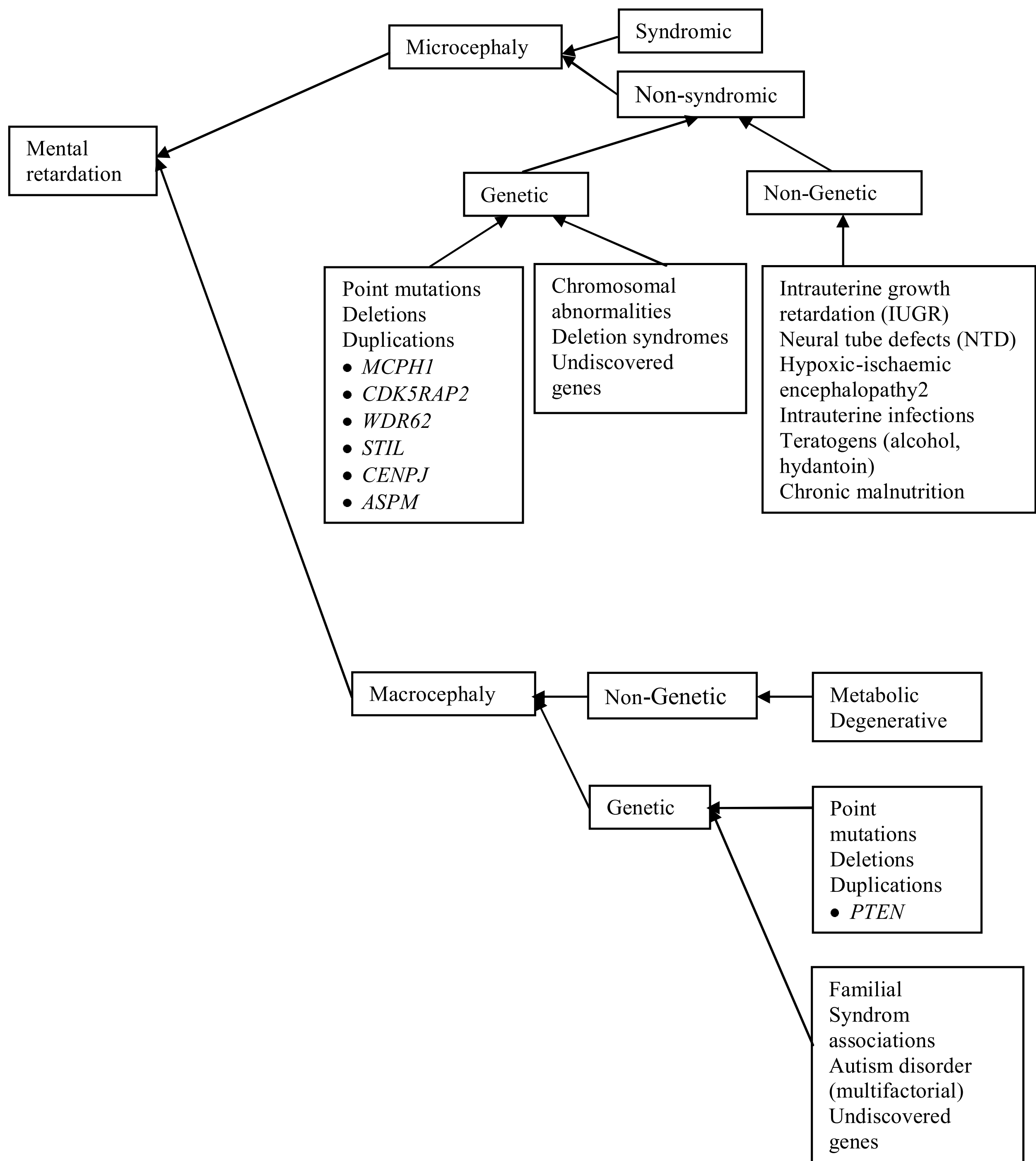

Figure 8. Theoretical Framework

2.6. Conceptual Framework

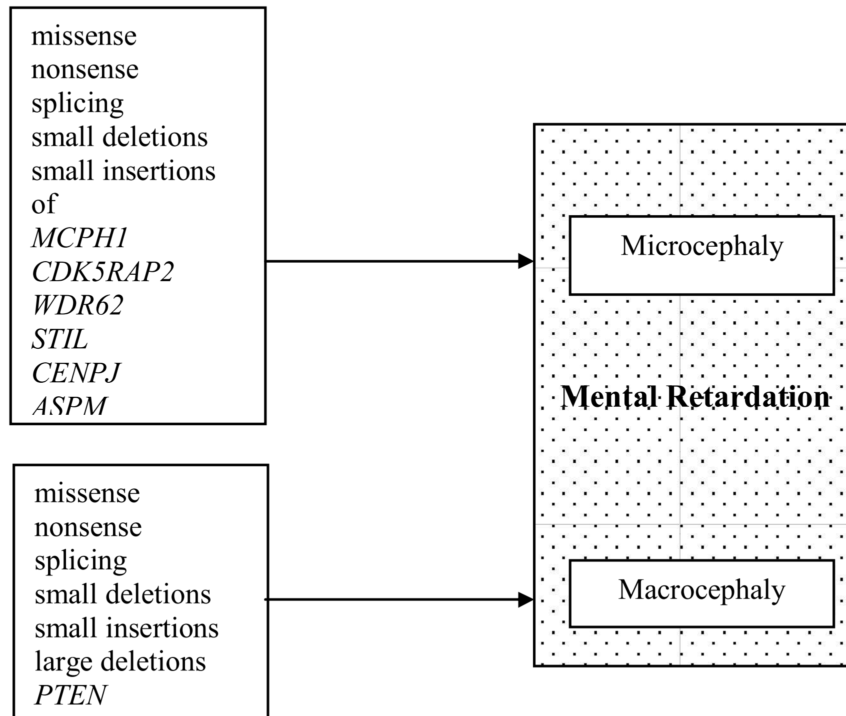

Figure 9. Conceptual Framework

2.7. Hypothesis

The hypotheses of this study are the followings:

1. Pathogenic mutations which include missense, nonsense, splicing, small deletions or small insertions can be found in the *ASPM, WDR62, CENPJ, CDK5RAP2, MCPH1* and *STIL* genes in the Indonesian mentally retarded individuals with microcephaly.
2. Pathogenic mutations which include missense, nonsense, splicing, small deletions, small insertions, or large deletions can be found in the *PTEN* gene in the Indonesian mentally retarded individuals with macrocephaly.