



Diagnostic approach for genetic causes of intellectual disability

Shin-Young Yim*

The Children's Rehabilitation Clinic, Department of Physical Medicine and Rehabilitation, Department of Medical Genetics, Ajou University School of Medicine, Suwon, Korea

Intellectual disability (ID) is the most common disability among people under the age of 20 years. In the absence of obvious non-genetic causes of ID, the majority of cases of severe ID are thought to have a genetic cause. The advent of technologies such as array comparative genomic hybridization, single nucleotide polymorphism genotyping arrays, and massively parallel sequencing has shown that *de novo* copy number variations and single nucleotide variations affecting coding regions are major causes of severe ID. This article reviews the genetic causes of ID along with diagnostic approaches for this disability.

Key words: Intellectual disability, Diagnosis, Karyotyping, Comparative genomic hybridization, High-throughput nucleotide sequencing.

Introduction

Intellectual disability (ID) is a disability characterized by significant limitations in both intellectual functioning and adaptive behavior, where adaptive behavior covers social and daily practical skills. This disability originates before the age of 18 years [1]. Generally, an intelligent quotient test score of around 70-75 indicates a limitation in intellectual functioning [1]. ID has a prevalence of 1-3% [2,3] and is the most common disability among people under the age of 20 years. The Metropolitan Atlanta Developmental Disabilities Surveillance Program data compiled during 1991-1994 reported estimated prevalences of 12.0, 3.0, 1.2, and 1.1 per 1,000 children aged 5-10 years for ID, cerebral palsy, hearing loss, and visual impairment, respectively [4]. In South Korea, registration of persons with disabilities is subject to Article 19, Chapter 2, of the Welfare Law for Persons

with Disabilities, Ministry of Health and Welfare, Republic of Korea. According to the Employment and Development Institute, as of December 31, 2012, there were 92,810 registered disabled people under the age of 20 years in South Korea, and 44,977 (48%) of them had ID.

Establishing an etiological diagnosis of ID is necessary to provide information on the prognosis and the risk of recurrence, as well as to avoid unnecessary invasive testing. Since an etiological diagnosis may provide an in-depth understanding of the disease mechanisms and the eventual development of therapeutic interventions, an etiological diagnosis of ID will be a prerequisite for personalized therapy when it becomes available in the near future.

ID has both genetic and non-genetic causes; non-genetic causes of ID include alcohol, other teratogenic causes, infections, and other adverse events, such as head injury or perinatal

Received: 1 December 2014, Revised: 28 February 2015, Accepted: 28 February 2015, Published: 30 June 2015

*Corresponding author: Shin-Young Yim, M.D., Ph.D.

The Children's Rehabilitation Clinic, Department of Physical Medicine and Rehabilitation, Department of Medical Genetics, Ajou University School of Medicine, 164 Worldcup-ro, Yeongtong-gu, Suwon 443-380, Korea.

Tel: +82-31-219-5284, Fax: +82-31-219-5209, E-mail: syyim@ajou.ac.kr

Conflict of interest: I declare that we do not have any conflicts of interests.

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© Copyright 2015 by the Korean Society of Medical Genetics

www.e-kjgm.org

Table 1. Causes of intellectual disability

Cause	Intellectual disability	
Genetic cause		
Chromosomal aberrations	Down syndrome (trisomy 21)	
	Edwards syndrome (trisomy 18)	
	Patau syndrome (trisomy 13)	
	Turner syndrome (45, X)	
	Klinefelter syndrome (47, XXY)	
	Segmental deletions and duplications of more than 5-10 Mb, such as Cri du chat syndrome	
	Others	
Copy number variations	Williams syndrome	
	Sotos syndrome	
	Prader-Willi syndrome	
	Angelman syndrome	
	Beckwith-Wiedemann syndrome	
Single gene disorders	Others	
	Fragile X syndrome	
	Rett syndrome	
	Kabuki syndrome	
	Phenylketonuria	
Non-genetic cause	Others	
	Prenatal causes	
	Teratogens, such as radiation, maternal infections, chemicals, drugs	
	Maternal disorders, such as hypothyroidism, anemia	
	Placental dysfunction	
	Others	
	Perinatal causes	
	Prematurity	
	Hypoxic ischemic encephalopathy	
	Others	
	Postnatal causes	
Head trauma		
Infection, such as meningoencephalitis		
Toxins, such as lead		
Malnutrition		
Child abuse		
Sensory deprivation		
Others		

asphyxia (Table 1). In the absence of obvious non-genetic causes of ID, the majority of severe ID cases are thought to have a genetic cause. Mutations of more than 400 genes are linked to ID, and these mutations often have low prevalences and indistinguishable phenotypes [5]. In the absence of diagnostic clues from the clinical phenotype other than ID, unbiased genome-wide approaches are required to detect genetic variants causing ID. During the past decade, the advent of technologies such as array comparative genomic hybridization (array CGH), single nucleotide polymorphism (SNP) genotyping arrays, and massively parallel sequencing has provided genome-

wide approaches to the identification of etiological genetic causes in ID diagnostics. This article reviews the genetic causes of ID along with diagnostic approaches to this disability.

Chromosomal Aberrations in Intellectual Disability

ID can be caused by chromosomal aberrations. Chromosomal aberrations can be either numeric or structural. Numeric aberration means the loss or gain of an entire chromosome, leading to monosomy or trisomy. Triploidy and tetraploidy refer to numeric aberrations with an extra set of whole chromosomes.

The absence of a chromosome is known to have more severe consequences than its excess. While loss of an entire autosomal chromosome is not compatible with life, the monosomy of the X chromosome seen in Turner syndrome (45, X) may result in a live-born female. Structural aberration refers to all types of genomic rearrangements, including deletions, duplications, insertions, inversions, translocations, loss of heterozygosity, and more complex alterations. Structural abnormalities can result from damages to double-stranded DNA which are not properly repaired [6]. Several well-known ID-related syndromes are associated with chromosomal aberrations and have large structural variations leading to additional distinct phenotypes along with ID. Down syndrome (trisomy 21; OMIM 190685) is the most common single cause of ID, with an estimated prevalence of 1 : 750-1 : 800. Edwards syndrome (trisomy 18), Patau syndrome (trisomy 13), Turner syndrome (45, X), and Klinefelter syndrome (47, XXY), as well as segmental deletions and duplications of more than 5-10 Mb, such as in Cri du chat syndrome (5p-syndrome; OMIM 123450), are detected by conventional karyotyping. Karyotyping provides a genome-wide assessment of chromosomal aberrations in terms of structure and number. The human genome is known to have about 42,000 bands at the molecular level. A 2,000-band chromosomal resolution may contain 1.5 Mb of DNA, while a 300-band resolution contains 7-10 Mb of DNA [7]. Conventional metaphase Giemsa-banding (G-banding) provides a 550-band resolution with a detection limit of 5-10 Mb, depending on the location. About 800 bands are observed in high-resolution banding, which requires less condensed prometaphase chromosomes. Giemsa stain is specific for the phosphate groups of DNA and attaches itself to regions of DNA where there are large amounts of adenine-thymine bonding. Heterochromatic regions, which tend to be rich in adenine and thymine and relatively gene-poor, stain more darkly in G-banding. In contrast, less condensed chromatin, which tends to be rich in guanine and cytosine and shows more transcription, incorporates less Giemsa stain, and these regions appear as light bands in G-banding [8].

When phenotypic investigation enables targeted analyses, fluorescence *in situ* hybridization (FISH), which usually provides a resolution of 2-5 Mb, can be used. Microdeletion syndromes, such as Williams syndrome and Sotos syndrome, which show relatively specific phenotypes in addition to ID, may be diagnosed by using a targeted FISH method. Methods to detect cytogenetically visible chromosomal alterations, such as metaphase G-banding and FISH, show about 18% diagnostic yield for ID or unexplained developmental delay, whereas the

diagnostic yield for diseases other than trisomy 21 is about 9% [9].

Copy Number Variations in Intellectual Disability

Copy number variations (CNVs) are the most common form of structural variation in the human genome, corresponding to relatively large regions of the genome that are deleted or duplicated on certain chromosomes [10]. Individuals with ID are known to have enrichment of CNVs [11]. Array CGH and SNP genotyping arrays enable genome-wide detection of CNVs or the loss of heterogeneity [12]. CGH was introduced for molecular cytogenetic analysis of solid tumors in 1992 [13]. Using DNA microarrays in conjunction with CGH techniques, a more specific form, i.e., array CGH, was developed. In array CGH, the patient's DNA (test DNA) and a reference DNA are labeled with different fluorescent dyes and hybridized to DNA probes that can represent either regions of or the entire genome. As the reference and test samples are labeled with different fluorescent dyes, hybridization with genomic probes on a platform results in distinct fluorescence intensities. The resulting ratio of the fluorescence intensities is proportional to the ratio of the copy numbers of DNA sequences in the test and reference genomes. If the intensities of the fluorophores on a probe are equal, this region of the patient's genome is interpreted as having a quantity of DNA equal to that of the reference sample; if there is an altered ratio of fluorophores, this indicates a loss or gain of the patient DNA at that specific genomic region [14]. There are usually two types of DNA probes used in array CGH. The probes for bacterial artificial clone-based arrays vary from 150 to 200 kb and require large amounts of DNA for hybridization. The probes for oligonucleotide-based arrays range from 25 to 85 bp, which allows for a much higher resolution than that of bacterial artificial clone-based arrays [15]. Higher-resolution platforms are now used in routine procedures for etiological diagnosis of ID [16]. With the growing body of knowledge on CNVs in diseases, databases on CNVs have been updated regarding pathogenic and non-pathogenic variations. In a SNP genotyping array, only the test (patient) DNA is labeled and hybridized to allele-specific, SNP oligonucleotide probes, which represent the two alleles. By comparing the differential amount of hybridization of the test DNA to that of each of the SNP probes, it is possible to determine specific homozygous and heterozygous alleles. A copy-neutral region with a loss of heterozygosity detected by a SNP array indicates uniparental disomy or identity by descent. Imprinting disorders associated with uniparental disomy, such

as Prader-Willi syndrome, Angelman syndrome, and Beckwith-Wiedemann syndrome, can be detected by SNP arrays. In the case of identity by descent, such regions may contribute to autosomal recessive ID [15].

Microarray platforms for copy number analysis, with both unique non-polymorphic probes and SNP probes with a higher resolution, have been used in recent years [17]. About 14% of children with ID have CNVs of >400 kb, and children with ID show more CNV enrichment when they have a concurrent craniofacial abnormality, autism spectrum disorder, or cardiovascular abnormality [18,19]. Array CGH gives a higher diagnostic yield for genetic testing of individuals with unexplained ID than a G-banded karyotype, due to its higher sensitivity for detection of submicroscopic deletions and duplications. Truly balanced rearrangements and low-level mosaicism are generally not detectable by arrays, but these are relatively infrequent causes of abnormal phenotypes [20]. Therefore, array CGH is now recommended as the first-tier cytogenetic diagnostic test for patients with ID. G-banding karyotype analysis should be reserved for patients with obvious chromosomal syndromes, a family history of chromosomal rearrangement, or a history of multiple miscarriages [20]. CNVs contribute to gene functions by a variety of mechanisms, including 1) dosage changes of the products of the associated genes, such as over-production or under-production of proteins encoded by the associated genes with CNVs in the form of either duplication or deletion; 2) gene disruption by the breakpoints of a CNV; 3) duplicated segments arranged in a tandem or inverted way that may result in gene disruption; 4) rearrangement-mediated fusions between different genes, resulting in the production of a chimeric gene; and 5) *cis*-effects on the expression of flanking genes [18].

Single-gene Disorders in Intellectual Disability

Fragile X syndrome is the most common form of monogenic ID. Fragile X syndrome is caused by the expansion of the CGG trinucleotide repeat at the *FMR1* gene in the X chromosome, leading to the absence of the fragile X mental retardation protein (FMRP). FMRP is an RNA binding protein that shuttles between the nucleus and the cytoplasm. It is thought that FMRP plays an important role in synaptic plasticity through the regulation of mRNA transport and the translational inhibition of local protein synthesis at synapses [21]. Rett syndrome is a devastating genetic disorder that represents the most common genetic cause of severe ID in females. Most cases are caused

by mutations in the X-linked *MECP2* gene, with an estimated incidence of 1 : 10,000-1 : 15,000 female births [22]. Metabolic disorders or inborn errors of metabolism designate a wide group of diseases caused by genetic defects leading to alterations of metabolism. These are estimated to be responsible for 1-5% of cases of non-syndromic ID. The majority of these arise from mutations in single genes that code for enzymes, resulting in abnormal or reduced enzymatic activity. As a consequence, some undegraded substrates may accumulate to toxic levels. Another scenario is that some compounds that are required for normal cellular metabolism cannot be produced [6]. The most common inborn error is phenylketonuria, with an average worldwide estimated prevalence of 1 : 10,000. Phenylketonuria results from deficient activity of the enzyme phenylalanine hydroxylase, which converts phenylalanine to tyrosine [6].

Sanger sequencing-based analysis has been the standard method for DNA sequencing of single genes or for step-wise gene testing for a clearly defined disorder. However, Sanger sequencing is not adequate for fast and complete sequencing of one or multiple genomes. The recent application of massively parallel sequencing methods has complemented microarray-based methods and has led to an exponential increase in the discovery of smaller structural variations [23]. Terms such as massively parallel sequencing, next-generation sequencing (NGS), and high-throughput sequencing are used interchangeably, and they all provide rapid sequencing at a lower cost than that required for Sanger sequencing. NGS technologies have fundamental differences from conventional Sanger sequencing, relying on different technical approaches and usually requiring a previous enrichment step. NGS is based on the sequencing of clonally amplified single molecules of genomic DNA and requires 10-50 reads of the same base in order to reliably identify heterozygous sequence variants, since every read shows only one of the possible two alleles in the sequence.

Sequencing of candidate genes and the use of NGS techniques showed that a large proportion of sporadic cases of ID might be caused by *de novo* point mutations and small insertions or deletions that cannot be detected by array CGH. There are 3 kinds of NGS-based sequencing in clinical diagnosis: 1) NGS-based target gene panel sequencing; 2) whole-exome sequencing; and 3) whole-genome sequencing. NGS-based target gene panel sequencing enables the simultaneous analysis of a group of genes related to a particular phenotype, such as X-linked ID, autism spectrum disorder, or retinitis pigmentosa. NGS-based target gene panel sequencing provides almost 100% coverage

for the included genes. However, since genes with a less obvious relevance to a particular phenotype are not included in the panel, mutations of unknown genes cannot be found. Whole-exome sequencing provides simultaneous analysis of nearly all coding exons of the human genome for undiagnosed patients. The exome consists of 1% of the human genome. The basic workflow of whole-exome sequencing is as follows: 1) isolation of genomic DNA from cells; 2) assessment of the purity of nucleic acids; 3) construction of pre-enriched DNA libraries in which the exome is enriched by capture-based or amplicon-based methods; 4) cluster generation of the enriched exome libraries; 5) exome sequencing; 6) alignment of sequences to a reference genome; and 7) variant calling with an adequate bioinformatics pipeline [24]. Diagnostic exome sequencing for persons with severe ID gives a diagnostic yield of about 16% [5]. Factors associated with negative results of whole-exome sequencing include the following: 1) genes encoding RNAs and about 10% of protein coding genes are not examined for mutations; 2) non-coding sequences, such as enhancers, promoters, and introns, are not examined; and 3) not all protein-coding genes have been annotated adequately enough to be associated with a phenotypic trait. Recently, the long-range control of gene expression by an enhancer located 500 kb away was reported [25]. Moreover, transposable elements, which occupy as much as 45% of the human genome, have been reported to be associated with several diseases [26]. However, enhancers and repetitive DNA, such as transposable elements, are not included in whole-exome sequencing. In this context, whole-genome sequencing will likely be used more for the diagnosis of rare diseases such as ID.

Conclusion

ID is the most common disability among people under the age of 20 years. Establishing an etiological diagnosis of ID is necessary to provide information on the prognosis and the risk of recurrence as well as to avoid unnecessary invasive testing. In the absence of obvious non-genetic causes of ID, the majority of cases of severe ID are thought to have a genetic cause. The advent of technologies such as array CGH, SNP genotyping arrays, and massively parallel sequencing has shown that *de novo* CNVs and single nucleotide variations affecting the coding region are major causes of severe ID. Differential diagnosis of genetic causes of ID can be made through a targeted approach or a genome-wide approach, depending on the phenotypes accompanying ID. Diagnostic methods with various resolutions

can be employed, depending on the size of the genetic variation. Since an etiological diagnosis may provide an in-depth understanding of the disease mechanisms and the eventual development of therapeutic interventions, an etiological diagnosis of ID is a prerequisite for personalized therapy when it becomes available in the near future.

References

1. American Association on Intellectual and Developmental Disabilities. Intellectual disability: definition, classification, and systems of supports. 11th ed. Washington, DC: AAIDD, 2010.
2. Yim SY, Yu HH, Lee IY. The prevalence of mental retardation among third grade elementary school children in the Suwon area, Korea. *J Korean Med Sci* 2002;17:86-90.
3. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. *Res Dev Disabil* 2011;32:419-36.
4. Centers for Disease Control and Prevention (CDC). Economic costs associated with mental retardation, cerebral palsy, hearing loss, and vision impairment--United States, 2003. *MMWR Morb Mortal Wkly Rep* 2004;53:57-9.
5. de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 2012;367:1921-9.
6. Bessa C, Lopes F, Maciel P. Molecular genetics of intellectual disability. In: Tan U, ed. Latest findings in intellectual and developmental disabilities research. Rijeka Croatia: InTech, 2012.
7. Rajcan-Separovic E. Chromosome microarrays in human reproduction. *Hum Reprod Update* 2012;18:555-67.
8. Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology. *Nat Rev Genet* 2005;6:782-92.
9. Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, et al. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A* 2006;140:2063-74.
10. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med* 2010;61:437-55.
11. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;466:368-72.
12. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014;511:344-7.
13. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, et al. Comparative genomic hybridization for molecular cytogenetic

- analysis of solid tumors. *Science* 1992;258:818-21.
14. Mefford HC, Batshaw ML, Hoffman EP. Genomics, intellectual disability, and autism. *N Engl J Med* 2012;366:733-43.
 15. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet* 2011;12:363-76.
 16. Manning M, Hudgins L; Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med* 2010;12:742-5.
 17. Brady PD, Vermeesch JR. Genomic microarrays: a technology overview. *Prenat Diagn* 2012;32:336-43.
 18. Hochstenbach R, Buizer-Voskamp JE, Vorstman JA, Ophoff RA. Genome arrays for the detection of copy number variations in idiopathic mental retardation, idiopathic generalized epilepsy and neuropsychiatric disorders: lessons for diagnostic workflow and research. *Cytogenet Genome Res* 2011;135:174-202.
 19. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. *Nat Genet* 2011;43:838-46.
 20. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 2010;86:749-64.
 21. Yim SY, Jeon BH, Yang JA, Kim HJ. Fragile X syndrome in Korea: a case series and a review of the literature. *J Korean Med Sci* 2008;23:470-6.
 22. Bedogni F, Rossi RL, Galli F, Cobolli Gigli C, Gandaglia A, Kilstrup-Nielsen C, et al. Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *Neurosci Biobehav Rev* 2014;46 Pt 2:187-201.
 23. Ng SB, Nickerson DA, Bamshad MJ, Shendure J. Massively parallel sequencing and rare disease. *Hum Mol Genet* 2010;19:R119-24.
 24. Jin HS, Lee JB, Kim K, Lee KY, Choi VN, Kim JS, et al. Identification of the rare compound heterozygous variants in the NEB gene in a Korean family with intellectual disability, epilepsy and early-childhood-onset generalized muscle weakness. *J Hum Genet* 2014;59:643-7.
 25. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 2014;507:371-5.
 26. Ayarpadikannan S, Kim HS. The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics Inform* 2014;12:98-104.