□종 설□

대 한 주 산 회 지 제25권 제3호, 2014
Korean J Perinatol Vol.25, No.3, Sep, 2014
http://dx.doi.org/10.14734/kjp.2014.25.3.133

Next Generation DNA Sequencing and Its Application in Clinical Medicine

Sechin Cho, M.D., FAAP, Founding Fellow ACMG, Professor Emeritus

University of Kansas School of Medicine, Wichita, Kansas, USA

Watson and Crick published a paper on the double helical structure of DNA in Nature in April 25, 1953. The human genome is contained in the 23 pairs of chromosomes and in the mitochondrial DNA of each cell. The Human Genome Project was launched in 1990 under the direction of Watson and concluded in 2003, on the 50th anniversary of Watson and Crick paper. Over 6 billion of nucleotides of genetic codes are in single cells. There are 23,000 protein coding genes and the remainder are non-coding DNA, regulatory DNA. Since the completion of Human Genome Project, these huge genomic information has been translated into clinically usable medical information. With the advent of massively parallel DNA sequencing, known as next generation DNA sequencing, the cost and turn-around time were significantly reduced so that the era of Whole Genome Sequencing entered into hospitals and medical clinics. On June 16, 2014 American Society of Human Genetics revised its mission statement as follows. "Our mission is to advance human genetics in science, health and society through research, education and advocacy". Finally medical genetics nestled its roots in the midst of genetics and genomics.

Key Words : New generation DNA sequencing, Whole genome/exome sequencing, Human genetic disease

The editor in chief of the Korean Journal of Perinatology, Hee-Sup Kim, M.D., Ph.D., invited me to write a review article titled "Next Generation DNA Sequencing and Its Application in Clinical Medicine". In order to write clinically useful review paper for actively practicing physicians in perinatal medicine area, Medline search under the keyword of next generation DNA sequencing was confined to Journal of the American Medical Association, New England Journal of Medicine, Obstetrics and Gynecology, Pediatrics, Journal of Pediatrics, Journal of Obstetrics and Gynecology, Obstetrics and Gynecology, Genetics in Medicine, and Journal of Perinatology which were published between January 01, 2004 and June 30, 2014. Upon careful review the reprints of 20 articles were obtained to be used for the references for this review paper.

In the beginning of the 20th century, Mendel's law of inheritance could explain the recurrence of certain diseases in families and it was recognized by Sir Garrod et al. From this point and on, medical genetics grew from a tiny subspecialty dealt with rare hereditary diseases to a fully recognized medical specialty. Medical genetics' concept and approaches are important components of the diagnosis and management of nearly all disorders. Medical genetics was a front line specialty which believed in evidence based medicine. The diagnosis of genetic disorders is usually proved or supported by genetic testing. Hemoglobin electrophoresis, lipid profiles, karyotype and gene sequences are examples.

Biochemical genetics, cytogenetics, and molecular genetics are the 3 recognized laboratory subspecialties by American College of Medical

Received : 22 September 2014, Revised: 25 September 2014
Accepted : 28 September 2014
Correspondence : Sechin Cho, M.D.
Genetic Consultant, 2140 Turnberry Way, Woodstock, Maryland 21163, USA
Tel : +1-410-750-9188, E-mail : cho3gene@gmail.com

Genetics.[1] Biochemical genetic testing has been available for many decades. The most frequent form of genetic testing is the direct analysis of the genetic material itself such as chromosome, genes, and the genome, which falls under the cytogenetics and molecular genetics.

Chromosomes were first observed in plant cell as early as in 1842. Lejeune discovered patients with Down syndrome had an extra copy of chromosome 21 in 1959. The other abnormalities of chromosome number and structure were continuously discovered and our understanding on function of chromosome and gene mapping information were enlarged. Technical advances such as high resolution banding, G banding, Q banding, fluorescent in situ hybridization (FISH) banding connect cytogenetics to molecular genetics. Future molecular cytogenetics include automated system for counting the results of standard FISH preparations and technique for virtual karyotyping such as comparative genomic hybridization arrays (CGH) and single nucleotide polymorphism (SNP). The new technology, new generation DNA sequencing, has placed within reach of the ultimate, most fundamental, and highest resolution genetic analysis: the precise identification and ordering of all 6 billion nucleotides of human genome.[2]

DNA sequencing technique

Over 40 year history of molecular diagnostics, DNA sequencing technique has been changed and evolved. Maxam-Gilbert sequencing, Sanger sequencing and multicapillary sequencing were invented in 1970, 1972, and 1992 respectively. For the entire duration of the Human Genome Project, DNA sequencing was performed on semi-automated capillary electrophoresis instruments by using biochemical method known as dideoxy chain-termination (ddA, ddC, ddG, and ddT) or Sanger sequencing. Along with Gilbert, Sanger received the Nobel Prize in 1980 for developing DNA sequencing.

This highly accurate but relatively slow method can provide the DNA sequencing of targeted and limited region of 150 to 200 nucleotides in a single run of one to two days. It took 13 years and 3 billion dollars to sequence entire 6 billion human genome. The reason for slow pace of Sanger sequencing is that it is based on daughter strand synthesis of a small stretch of DNA that is selected by using the hybridization of specific oligonucleotide primers to just that regions. The primers serve as start sites for DNA polymerase to make complementary strands that terminate whenever a particular dideoxynucleotide derivative is incorporated into elongating strands, and resulting sequence is deduced by measuring the sizes of the terminated fragments on capillary electrophoresis instrument.

However, new generation, or massively parallel, sequencing breaks up the whole genome into > 3 million small fragments, which are universally primed for DNA polymerase copying by copying 4-color fluorescently labeled nucleotides and are analyzed on instruments that take instantaneous snapshots of the added nucleotides after each round of synthesis across the entire genome. Each fragment is copied between 10 and more than 100 times. And the aggregated fluorescent photos of all of the synthesis products comprise "cluster arrays" representing 4 to 5 terabytes of computer data. The alignment software reconstructs the entire genome from these of hundreds of millions of DNA fragments. Depending on the instruments used, the

method can provide the sequence 3 to 10 gigabases of DNA in a single run easily enough to cover the entire haploid or diploid individual genome. Currently the most advanced instrument, Hi-Xeq 10, decreased the net cost per sample per run down close to $1,000., which is the cost of a single gene sequencing test such as BRCA1 and BRCA2. Now genome sequencing is a legitimate test for certain clinical indications.[2]

Terms of Genomics

The term "genetics" refers to the study of single genes in isolation while the term "genomics" refers to the study all genes in genome and interaction among them and their environments.

SNPs are the most common of all polymorphisms. SNPs have only two alleles corresponding to the two different bases occupying a particular location in the genome. SNPs are common and occur on average of once every 1,000 base pairs, which means that there is an average of 3,000,000 differences between any two human genomes. Any clinical effects of common SNPs are subtle altering of disease susceptibility rather than a direct cause of serious illness.

Exomes which comprise about 250,000 exons are small subset (about 1.5%) of the total human genome that codes for protein. Exome sequencing involves an additional step: an exon-capture step by which coding regions are selected from the total genome DNA by means of hybridization, either to microarray or in solution. Because most inherited disorders are believed to be due to mutations in the coding regions, this approach will allow laboratories to focus exclusively on those regions and eliminate the tremendous mass of non-coding DNA in the genome. However, downside is that exon-capture technique is not 100% efficient. Three to five percent of the exons will not be captured and sequenced. If missing exons are deemed necessary for ruling out a particular diagnosis, they must be specifically targeted and sequenced by using any of a variety of work around techniques based on polymerase chain reaction. Some laboratories further restrict the exons being queried to just a panel of genes known to be associated with disorder suspected. Example of commonly ordered gene panels, consisting of 20 to 100 genes per disorder, are shown on Table 1.[2]

Discoveries and Diagnosis by DNA Sequencing

The primary application of new generation DNA sequencing, which applied to children who may have a syndrome but are not diagnosed has produced some dramatic results; some are the new gene discovery and others are the detection of mutations in known genes that were not suspected to be associated with the particular patient's phenotype. Between 2010 and 2013 more than 150 heritable disorders have been discovered this way. The biggest challenge and dilemma in performing whole genome sequencing or whole exome se-

Table 1. Disease gene panels that use next-generation sequencing

Hypertrophic cardiomyopathy
Dilated cardiomyopathy
Hereditary arrhythmias (channelopathies)
Retinitis pigmentosa
Albinism
Intellectual disability
DNA repair defects
Skeletal dysplasia
Disorders of sexual development
Hearing loss

quencing on a clinical basis is dealing with the huge quantity of unexpected sequence of variants are detected.

From an ethical perspective, more concerning fact is the incidental discovery of "off-target mutations". These are deleterious changes in known genes that are irrelevant to the patient's current phenotype and the reason for ordering the test but which may predict future diseases, for example, the finding of BRCA mutation in a 2 year old girl undergoing whole exome sequencing for intellectual disability. The BRCA mutation has nothing to do with intellectual disability, and we would never do such predictive testing for adult onset disease in 2 year old girl. However, what is the liability of failing to disclose this incidentally detected risk. Table 2 is the collection of incidentally found inherited diseases so far.[1,2]

In high-risk pregnant women, noninvasive prenatal testing with the use of massively parallel sequencing of maternal plasma cell-free DNA (cfDNA) testing accurately detects fetal autosomal aneuploidy. However, its performance in low risk women is unclear. At 21 centers in the United States, blood samples were collected from women with singleton pregnancies who were undergoing standard aneuploidy screening (serum biochemical screening with or without nuchal translucency measurement). Massively parallel sequencing was performed in a blinded fashion to determine the chromosome dosage for each sample. The primary end point was a comparison of the false positive rates of detection of fetal trisomies 21 and 18 with the use of standard screening and cfDNA testing. Birth outcomes or karyotypes were the reference standard. For trisomies 21 and 18, the false positive rates with cfDNA testing were significantly lower than those with standard screening (0.3% vs. 3.6% for trisomy 21, $P<0.001$: and 0.2% vs. 0.6% for trisomy 18, $P=0.03$). The use of cfDNA detected all cases of aneuploidy. The positive value for cfDNA testing versus standard screening was 45.5% versus 4.2% for trisomy 21 and 40.0% versus 8.3% for trisomy 13. The cfDNA testing may have significant impact on standard aneuploidy screening.[3-6]

Newborn screening program was introduced 50 years ago and changed genetic-metabolic world and greatly expanded the concept of preventive medicine. This expansion has been marked by two major milestones: the first, pre-tandem mass spectrometry, included phenylketonuria, galactosemia, homocystinuria, maple syrup urine disease, congenital hypothyroidism, congenital adrenal hyperplasia, sickle cell disease and biotinidase deficiency, the second, tandem mass spectrometry-based, has seen an explosive increase in information, often instrumental for diagnosis, prevention, and appropriate management of many additional metabolic disorders including the organic acidemias and fatty acid oxidation defects. Incomplete knowledge of natural history of disease and inconclusive diagnosis are shortcomings and pitfalls. Currently genotyping in newborn screening is confined to confirming the disorder suggested by newborn genetic screen. As new generation DNA sequencing technology emerged, serious consideration of potential application to newborn genetic screening itself has been emerged.[7] Not only medical aspects but also ethical, legal and social issues have to be thoroughly examined.[8]

An FDA advisory committee voted unanimously to recommend that the Pap smear be replaced with a human papillomavirus test as first-time standard of care for cancer screening on March 12, 2014.

Table 2. Conditions, genes, and variants recommended for return of incidental findings in clinical sequencing

Phenotype	MIM-disorder	PMID-Gene Reviews entry	Typical age of onset	Gene	MIM-gene	Inheri-tance[a]	Variants to report[b]
Hereditary breast and ovarian cancer	604370 612555	20301425	Adult	BRCA1 BRCA2	113705 600185	AD	KP and EP
Li-Fraumeni syndrome	151623	20301488	Child/adult	TP53	191170	AD	KP and EP
Peutz-Jeghers syndrome	175200	20301443	Child/adult	STK11	602216	AD	KP and EP
Lynch syndrome	120435	20301390	Adult	MLH1 MSH2 MSH6 PMS2	120436 609309 600678 600259	AD	KP and EP
Familial adenomatous polyposis	175100	20301519	Child/adult	APC	611731	AD	KP and EP
MYH-associated polyposis; adeno-mas,multiple colorectal, FA type 2; colorectal adenomatous Polyposis,autosomal recessive, with pilomatricomas	608456 132600	23035301	Adult	MUTYH	604933	ARC	KP and EP
Von Hippel-Lindau syndrome	193300	20301636	Child/adult	VHL	608537	AD	KP and EP
Multiple endocrine neoplasia type1	131100	20301710	Child/adult	MEN1	613733	AD	KP and EP
Multiple endocrine neoplasia type2	171400 162300	20301434	Child/adult	RET	164761	AD	KP
Familial medullary thyroid cancerd	1552401	20301434	Child/adult	RET	164761	AD	KP
PTENhamartoma tumor syndrome	153480	20301661	Child/adult	PTEN	601728	AD	KP and EP
Retinoblastoma	180200	20301625	Child	RB1	614041	AD	KP and EP
Hereditary paraganglioma-pheo-chromocytoma syndrome	168000 (PGL1) 601650 (PGL2) 605373 (PGL3) 115310 (PGL4)	20301715	Child/adult	SDHD SDHAF2 SDHC SDHB	602690 613019 602413 185470	AD	KP and EP KP KP and EP
Tuberous sclerosis complex	191100 613254	20301399	Child	TSC1 TSC2	605284 191092	AD	KP and EP
WT1-related Wilms tumor	194070	20301471	Child	WT1	607102	AD	KP and EP
Neurofibromatosis type2	101100	20301380	Child/adult	NF2	607379	AD	KP and EP
Ehlers-Danlos syndrome, vascular type	130050	20301667	Child/adult	COL3A1	120180	AD	KP and EP
Marfan syndrome, Loeys-Dietz syn-dromes, and familial thoracic Aortic aneurysms and dissections	154700 609192 608967 610168	20301510 20301312 20301299	Child/adult	FBN1 TGFBR1 TGFBR1 SMAD3 ACTA2 MYLK MYH11	134797 190181 190182 603109 102620 600922 160745	AD	KP and EP

[a]Some conditions that may demonstrate semidominant inheritance (SD) have been indicated as autosomal dominant (AD) for the sake of simplicity. Others have been labeled as X-linked (XL); [b]KP: known pathogenic, sequence variation is previously reported and is a recognized cause of the disorder; EP: expected pathogenic, sequence variation is previously unreported and is of the type that is expected to cause the disorder. Note: The recommendation to not repot expected pathogenic variants for some genes is due to the recognition that truncating variants, the primary type of expected pathogenic variants, are not an established cause of some diseases on the list. [c]Although carriers may have modestly increased risk, we recommend searching only for individuals with biallelic mutations; [d]On the basis of evidence presented to the Working Group after the online posting of these recommendations, the decision was made to remove one gene, NTRK1, from the recommended list.

Abbreviations: MIM, Mendelion Inheritance in Man; PMID, PuMed identifier; MYH, mutY homolog; FAP, familial adenomatous polyposis; WT1, Wilms tumor1

Table 2. Continued

Phenotype	MIM-disorder	PMID-Gene Reviewsentry	Typical age of onset	Gene	MIM-gene	Inheri-tance[a]	Variants to report[b]
Hypertrophic cardiomyopathy, dilated cardiomyopathy	115197	20301725	Child/adult	MYBPC3	600958	AD	KP and EP
	192600			MYH7	160760		KP
	601494			TNNT2	191045		KP and EP
	613690			TNNI3	191044		KP
	115196			TPM1	191010		
	608751			MYL3	160790		
	612098			ACTC1	102540		
	600858			PRKAG2	602743		
	301500			GLA	300644	XL	KP and EP (hemi, het, hom)
	608758						
	115200						
				MYL2	160781	AD	KP
				LMNA	150330		KP and EP
Catecholaminergic polymorphic ventricular tachycardia	604772			RYR2	180902	AD	KP
Arrhythmogenic right-ventricular cardiomyopathy	609040	20301310	Child/adult	PKP2	602861	AD	KP and EP
	604400			DSP	125647		
	610476			DSC2	125645		
	607450			TMEM43	612048		KP
	610193			DSG2	125671		KP and EP
Romano-Ward long QT syndrome Types 1, 2, and 3, Brugada syndrome	192500	20301308	Child/adult	KCNQ1	607542	AD	KP and EP
	613688			KCNH2	152427		
	603830			SCN5A	600163		
	601144						
Familial hypercholesterolemia	143890	No Gene reviews Entry	Child/adult	LDLR	606945	SD	KP and EP
	603776			APOB	107730	SD	KP
				PCSK9	607786	AD	
Malignant hyperthermia susceptibility	145600	20301325	Child/adult	RYR1	180901	AD	KP
				CACNA1S	114208		

[a]Some conditions that may demonstrate semidominant inheritance (SD) have been indicated as autosomal dominant (AD) for the sake of simplicity. Others have been labeled as X-linked (XL); [b]KP: known pathogenic, sequence variation is previously reported and is a recognized cause of the disorder; EP: expected pathogenic, sequence variation is previously unreported and is of the type that is expected to cause the disorder. Note: The recommendation to not report expected pathogenic variants for some genes is due to the recognition that truncating variants, the primary type of expected pathogenic variants, are not an established cause of some diseases on the list. [c]Although carriers may have modestly increased risk, we recommend searching only for individuals with biallelic mutations; [d]On the basis of evidence presented to the Working Group after the online posting of these recommendations, the decision was made to remove one gene, NTRK1, from the recommended list.

New England Journal of Medicine reported in April, 2014 that next generation DNA sequencing allowed presumptive diagnosis and successful targeted treatment for neuroleptospirosis in an immunocompromised adolescent boy who has adenosine deaminase deficiency after months of fruitless conventional testing and declining clinical status.

Next generation DNA sequencing screen of the pooled eight women in the case group generated 15 million reads. Two of the samples contained a total of seven unique viral sequences, assembling into three contigs that matched three isolates of torque-teno minivirus. Using polymerase chain reaction, these isolates were found in 5 of 55 cases (including the two identified in the eight - sample screen and none of 25 control samples). In conclusion, torque-teno minivirus was found in three women in case group and no women in the control group. Next

generation sequencing may prove a useful tool to perform unbiased screens for pathogens in maternal, fetal, and placental samples.[9]

Conclusion

The importance of molecular genetics in the diagnosis and treatment of inherited disorders has been widely recognized. New generation DNA sequencing technique further reduced the cost of whole genome/exome sequencing and its turnaround time. Now it becomes affordable and practical test in clinical medicine. With time and concentrated research efforts, background database and new discovery of mutations and variations will be rapidly accumulated enough that main missions of medical genetics are newborn genetic screening, pre-symptomatic detection of diseases, preventive treatment of diseases by lifestyle modification, nutritional intervention, environmental control, and pharmacotherapeutic treatment.[10-12]

References

1) Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013;15:565-74.
2) Grody WW, Thompson BH, Hudgins L. Whole-exome/genome sequencing and genomics. Pediatrics 2013;132 Suppl 3:211-5.
3) Bianchi DW, Parker RW, Wentworth J, Madankumar R, Saffer C, Das AF, et al. DNA sequencing versus prenatal aneuploidy screening. N Eng J Med 2014;370:799-808.
4) Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Obstet Gynecol 2012;119:890-901.
5) Greene MF, Phimister EG. Screening for trisomies in circulating DNA. N Engl J Med 2014;370:874-5.
6) Simpson JL. Is cell-free fetal DNA from maternal blood finally ready for prime time? Obstet Gynecol 2012;119:883-5.
7) Landau YE, Lichter-Konecki U, Levy HL. Genomics in newborn screening. J Pediatr 2014;164:14-9.
8) Greeley SA, Msall ME, Acharya K. Genomic sequencing in newborn screening programs. JAMA 2012;307:2146-7.
9) Shah AA, Wang D, Hirsch E. Next-generation sequencing of maternal serum to detect viruses in women with labor or premature rupture of membranes. Obstet Gynecol 2014;123 Suppl 1:35-6.
10) Korf BR, Rehm HL. New approach to molecular diagnosis. JAMA 2013;309:1511-21.
11) Jacob HJ. Next-generation sequencing for clinical diagnostics. N Eng J Med 2013;369:1557-8.
12) Abdul-Karim R, Berkman BE, Wendler D, Rid A, Khan J, Badgett T, et al. Disclosure of incidental findings from next-generation sequencing in pediatric genomic research. Pediatrics 2013;131:564-71.