Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing

Sang Tae Park[1], Jayoung Kim[2,3]

[1]Macrogen Corporation, Rockville, MD, [2]Departments of Surgery and Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA,
[3]Department of Medicine, University of California Los Angeles, Los Angeles, CA, USA

This article is a mini-review that provides a general overview for next-generation sequencing (NGS) and introduces one of the most popular NGS applications, whole genome sequencing (WGS), developed from the expansion of human genomics. NGS technology has brought massively high throughput sequencing data to bear on research questions, enabling a new era of genomic research. Development of bioinformatic software for NGS has provided more opportunities for researchers to use various applications in genomic fields. *De novo* genome assembly and large scale DNA resequencing to understand genomic variations are popular genomic research tools for processing a tremendous amount of data at low cost. Studies on transcriptomes are now available, from previous-hybridization based microarray methods. Epigenetic studies are also available with NGS applications such as whole genome methylation sequencing and chromatin immunoprecipitation followed by sequencing. Human genetics has faced a new paradigm of research and medical genomics by sequencing technologies since the Human Genome Project. The trend of NGS technologies in human genomics has brought a new era of WGS by enabling the building of human genomes databases and providing appropriate human reference genomes, which is a necessary component of personalized medicine and precision medicine.

Keywords: High-Throughput Nucleotide Sequencing; Sequence Analysis, RNA; Genomics; Epigenomics; Human Genome Project

• **Conflict of Interest:** No potential conflict of interest relevant to this article was reported.

NGS TECHNOLOGIES AND PLATFORMS

Next-generation sequencing (NGS) technologies are methods that sequence nucleotides faster and cheaper than Sanger sequencing. These massively parallel DNA sequencing methods have opened a new era of genomics and molecular biology. Compared to the traditional Sanger capillary electrophoresis sequencing method [1,2], which is considered a first-generation sequencing technology, NGS technologies provide higher throughput data with lower cost and enable population-scale genome research. NGS technologies have three major improvements compared to first-generation sequencing [3]. First, NGS methods do not require a bacterial cloning procedure and prepare libraries for sequencing in a cell free system. Second, NGS technologies process millions of sequencing reactions in parallel and at the same time. Third, detection of bases is performed cyclically and in parallel.

These major improvements allow scientists to process sequencing of entire genomes with low cost and in a very short period of time. Fig. 1 shows the overall workflows of conven-

Corresponding author: Sang Tae Park (ID) http://orcid.org/0000-0003-0790-7302
Macrogen Clinical Laboratory, Macrogen Corporation, 1330 Piccard Drive, Suite 103, Rockville, MD 20850, USA
E-mail: stpark@macrogenlab.com / Tel: +1-301-251-1007 / Fax: +1-301-251-4006
Submitted: October 12, 2016 / **Accepted after revision:** October 18, 2016
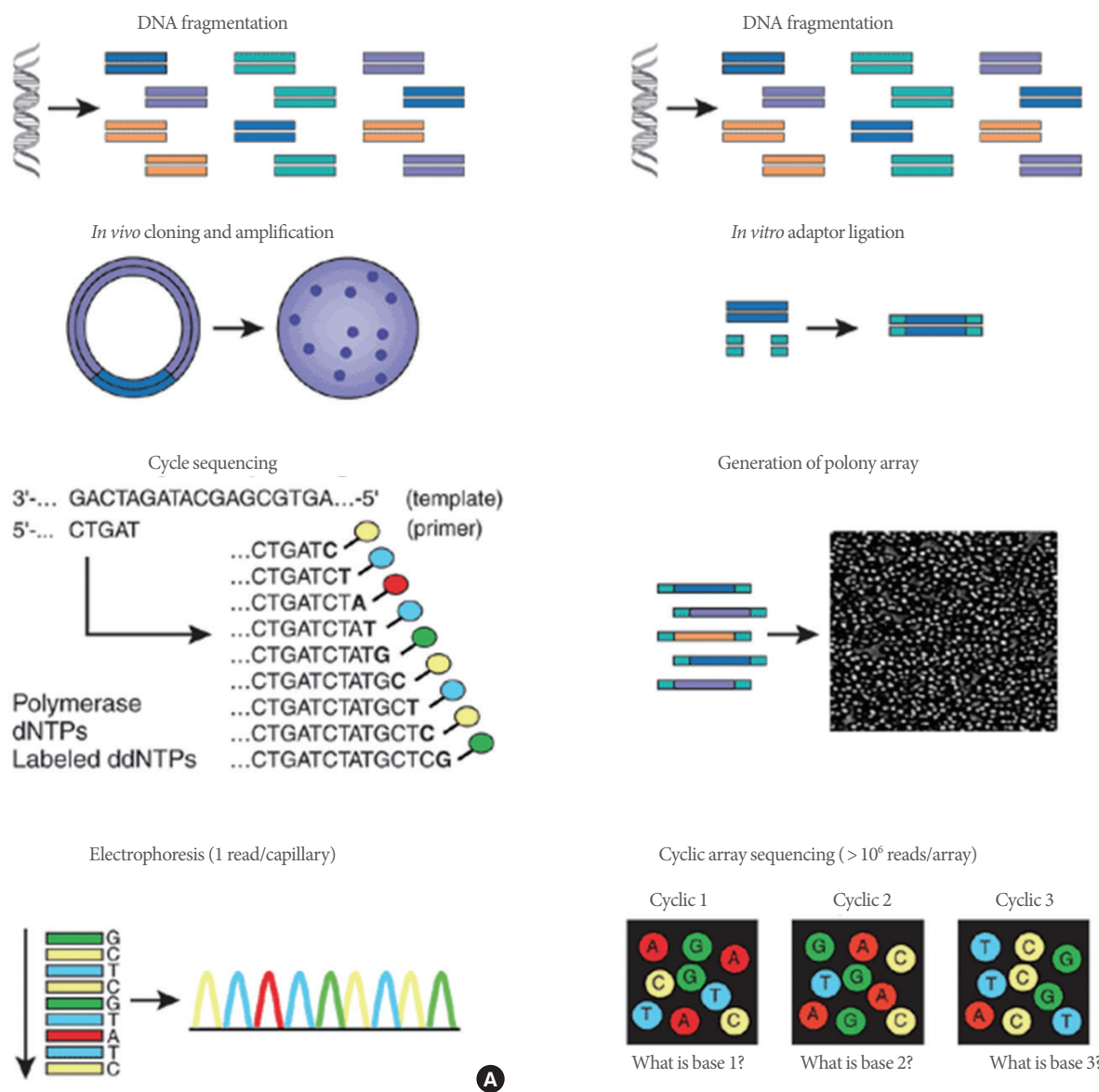
Fig. 1. Comparison of Sanger sequencing and next-generation sequencing (NGS). (A) Shotgun Sanger sequencing workflow. (B) Shotgun based NGS workflow with cyclic-array method. This figure shows the basic process of the 2 technologies and also shows 3 major improvements in NGS from Sanger sequencing described in that review. Adapted from Shendure J, et al. Nat Biotechnol 2008;26:1135-45 [4].

tional sequencing and NGS [4]. However, NGS technologies needed the development of novel alignment algorithms to assemble and map the genome from the relatively short reads [5]. As of 2016, there are 3 major platforms of NGS technologies. Roche (formerly 454) has the most recent 454-based sequencer, GS FLX+ (Roche Diagnostics Co., Branford, CT, USA), which generates about 1,000,000 reads of 700 base pairs (bp) [6]. Life Technologies, part of Thermo Fisher (Waltham, MA, USA), has the Personal Genome Machine and Proton with their Ion tor-

rent technology. Proton uses semiconductor-sequencing technology with solid-state pH meters detecting a hydrogen ion released from DNA on chip. Proton generates 60–80 million reads of up to 200-bp fragments, which delivers up to 10 Gb of sequence per run with an Ion P1 Chip [7]. Illumina (San Diego, CA, USA) acquired Solexa in 2007 and is supplying most NGS platforms in the world such as HiSeq2500, HiSeq4000, and HiSeq X. Illumina's the most recently available high-throughput sequencer is the HiSeq X TEN system consisting of 10 HiSeq X.
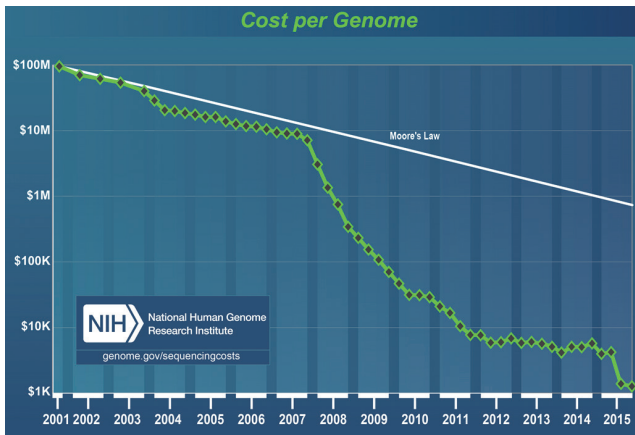
Fig. 2. Graph of "Cost per Genome." This graph illustrates the nature of the reductions in sequencing costs and also hypothetical data reflecting Moore's Law. Adapted from National Human Genome Research Institute (NHGRI) [9].

The HiSeq X delivers 1.8 Tb of sequence per run in 3 days from ~6 billion reads of 150 bp and is especially designed for whole genome sequencing (WGS) that requires ultrahigh throughput and multiparallel sequencing at the same time [8]. This system may be able to break US$1,000 per human WGS. This is almost a 10,000-fold reduction in the cost of sequencing a human genome since 2004 (Fig. 2) [9].

Other NGS technologies have been developed by companies like Qiagen, Helicos Biosciences, and Pacific Biosciences (now part of Roche). These platforms sequence directly from template DNA, while the previously discussed NGS technologies amplify the DNA template during the library preparation steps. Thus, these technologies are called third-generation sequencing technologies (to distinguish from NGS technologies). The leader of the third-generation sequencing field is Roche. Roche has released PacBio RS II from Pacific Biosciences technology (Menlo Park, CA, USA), and the PacBio RS II generates several thousands of long reads with up to 20,000 bp [10]. This long-reads sequencing technology has an advantage in *de novo* genome assemblies because the contig and scaffold N50 values are substantially higher than *de novo* genome assemblies by short-reads sequencing [11,12].

APPLICATIONS OF NGS

There are many applications for NGS and new methods being developed continuously. There are several classifications for NGS applications. In this section, we classify NGS applications

according to the experimental purpose.

(1) To build a new genome from unknown organisms, researchers use *de novo* sequencing with assembly. This *de novo* genome assembly requires a tool called an "assembler." Assemblers put fragmented reads of DNA together like a jigsaw puzzle by aligning regions with overlap to build a genome sequence [13].

(2) To measure genetic variation from an organism with an existing reference genome, researchers can do DNA-sequencing, RNA-sequencing, and epigenome sequencing. In the case of DNA-sequencing, whole genome, whole exome (for eukaryotes), and targeted sequencing are available with NGS technologies. By comparing sequencing results to reference genomes, researchers can see the genetic variation such as single nucleotide polymorphisms (SNPs), structural variations, copy number variations, and other variations using various software programs [14].

(3) To analyze transcriptome results with sequencing, researchers synthesize complementary DNA from RNA for sequencing (There are RNA preparation library kits on the market for NGS platforms.) RNA sequencing allows researchers to examine splicing of RNA, gene fusion, mutation, and differential gene expression. Compared to the hybridization-based microarrays for gene expression studies, microarrays show artifacts of hybridizations, a narrow range of expression quantitation, low resolution from several to 100 bp, and limitation of coverage based on probes [15]. The technical advantages of RNA-Seq has led to a transition in transcriptomics from microarrays to sequencing-based methods.

(4) For epigenome studies and regulatory mechanisms of the genome, researchers can use DNA methylation sequencing and chromatin immunoprecipitation followed by sequencing (ChIP-Seq). To determine methylation of CpG dinucleotides, the bisulfite sequencing method is applied. Bisulfite treatment converts cytosine residue to uracil so only the methylated cytosine residues are detected. ChIP-Seq is a method for analyzing protein-DNA interactions, such as the binding sites of transcription factors. ChIP-Seq requires antibodies for proteins of interest to enrich the DNA regions bound by proteins in living cells. Several research publications have used ChIP-Seq to demonstrate and predict genome-wide networks of regulation [16,17].

(5) NGS technologies allow microbial ecology scientists to investigate genetic materials from environmental samples on a tremendous scale. Scientists can use extracted DNA from environmental samples without cloning [18].

As described above, NGS technologies provide opportunities for better quality and quantity to scientists in many fields. Many other new methods with NGS and preparation technologies are being introduced now.

HUMAN GENETICS AND GENOMICS

Human genetics is the study of inheritance in human beings and encompasses various fields including classical genetics, cytogenetics, genomics, population genetics, and clinical genetics. To study human genetics for many purposes, researchers wanted to create a fully mapped sequence of the human genome and initiated the human genome project (HGP) in 1990. The HGP aimed to map the nucleotides in a human haploid reference genome. The HGP was completed in April 2003 and finally published on May 27, 2004 [19]. The sequence is now stored in databases available to anyone on the internet. The National Center for Biotechnology Information (NCBI) built a database known as GenBank [20] and other organizations including the University of California, Santa Cruz and Ensembl [21] present additional data with powerful tools for search and visualization in the UCSC Genome browser [22] and Ensembl Browser [21], respectively. The human reference genomes are maintained by the Genome Reference Consortium (GRC), which is an inter-
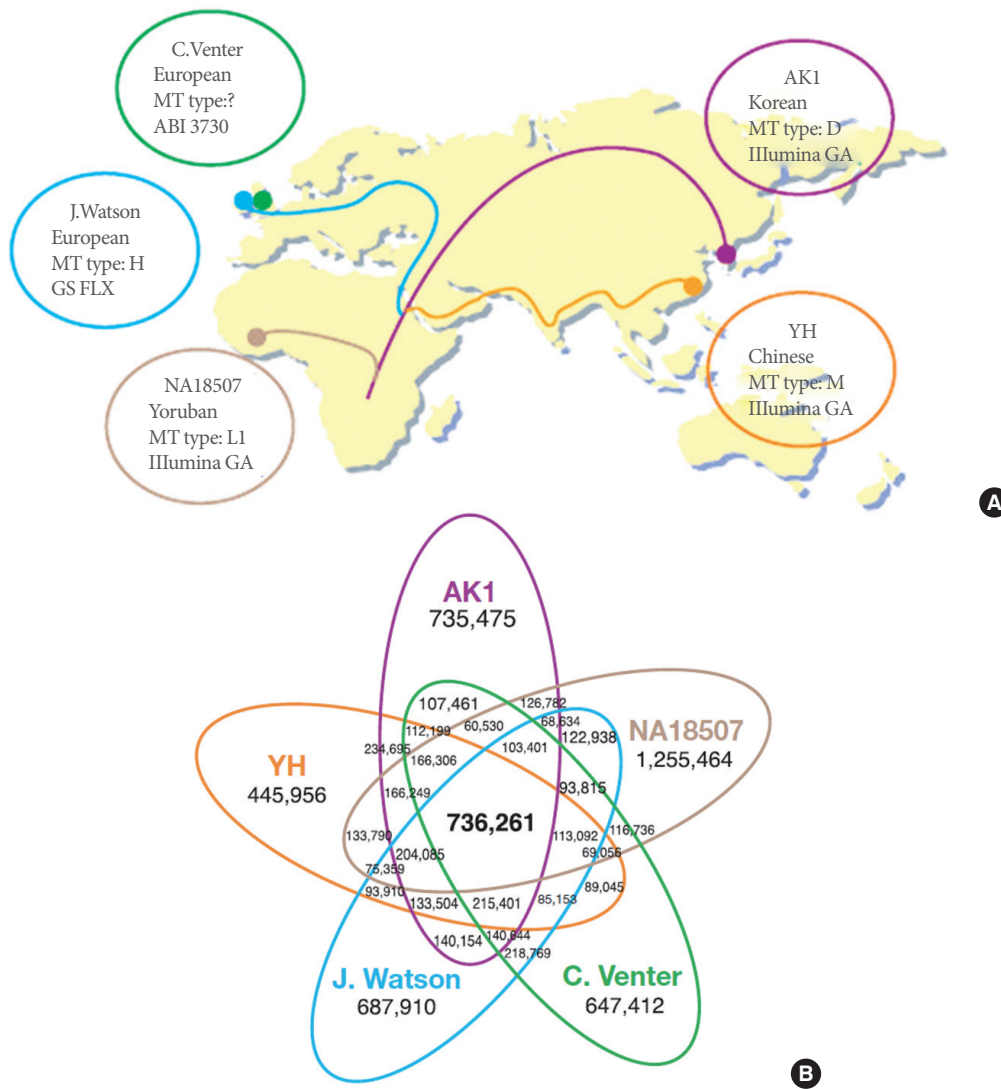

Fig. 3. (A) Geographic map of 5 sequenced genomes. MT type represents the mitochondrial haplogroup. Illumina GA, ABI 3730, and GS FLX represent the sequencing platform used for these sequencing projects. (B) Diagram of single nucleotide polymorphism overlapping results between 5 genomes. Adapted from Kim JI, et al. Nature 2009;460:1011-5 [28].

national collective of academic and research institutes from the HGP. The current reference genome is GRCh38.p8, which was released on June 30, 2016. This is an updated release from GRCh38, released on December 24, 2013 (new human genome assembly [GRCh38] released) [23].

The introduction and low cost of new technologies in sequencing is leading to an era of personal genome sequences. Many of the new human genome sequencing projects from individuals have reported additional human genome sequences. Comparison studies with the original human reference genome have shown diversity in human genetic variation by ethnicity and regions of ancestry [24-28]. Some examples follow.

(1) One of the comparison studies showed an indel distribution pattern among 5 different sequenced genomes [28]. Five sequenced genomes from 3 different distinct geographic regions including the newly sequenced AK1 from a Korean individual (Fig. 3A). The possibility of differences in technical procedures or interindividual variability was explained. Bioinformatic analysis for SNP detection from aligned sequences showed that 21% of AK1's SNPs were unique and 8% were identical in all genome sequences (Fig. 3B). Many other individual genome sequences are being reported continuously, and one report noted that South Asians have a higher risk of type-2 diabetes and cardiovascular disease compared to Europeans, which is the current human reference genome [29]. These reports explained that the genome of each individual is unique and emphasized the necessity of personal genome sequencing and population genomics to identify genome sequences possibly related with diseases in medical genetics.

(2) Population genomics is the large-scale comparison of DNA sequences of populations. This is a neologism and a new paradigm in population genetics by combining genomics concepts and technologies [30]. Population genomics uses genome-wide sampling to identify the phenotypic variation such as gene flow and inbreeding and to improve understanding on microevolution [31,32]. In human population genomics, there was a revolution with recent advancements in sequencing and data analysis. These advancements allowed scientists to study hundreds and thousands of loci from populations and enable genome wide effects and/or focus with genome-scale data.

(3) Medical genetics is the branch of medicine involving the diagnosis and management of hereditary disorders. Medical genetics considers the diagnosis, management, and counseling people with genetic disorder as a form of medical care, while research-oriented human genetics focuses on the causes and inheritance of genetic disorders. Clinical genetics and cancer genetics are the practices of clinical medicine for hereditary disorders and cancers, respectively. Also, clinical and cancer genomics are new fields with genome sequencing to inform patient diagnosis and care by diagnosing genetic diseases, categorizing patients for appropriate treatment, and providing information about an individual's response to treatment. However, clinical and cancer genomics require a more comprehensive view of genomics information and associated biological implications [32]. It is believed that the convergence of research-based genomic research and clinical/cancer genomics for medical care will become increasingly important.

NEW ERA OF WGS

WGS follows the current trends in the convergence of fundamental genomic research and clinical implications of the presence or absence of certain genes. Now, WGS is becoming one of the most widely used applications and is providing tremendous quantities of genome sequences relative to the past through public and private human genome sequencing projects throughout the world. As of 2015, the genomes of 2,504 people from 26 different populations have been reconstructed according to reports [33,34] from the 1,000 Genomes Project [35].

The release of HiSeq X Ten systems with a capacity of 1.8-Tb sequences per run brought a new era of WGS explosively by the reduction of costs. These HiSeq X Ten systems can deliver 18,000 Genomes a year. The National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH) planned and processed 20,000 Genomes for their TOPMed (Trans-Omics for Precision Medicine) program by 2015 and has expanded the WGS project with 62,000 individual genomes. Besides NHLBI/NIH, many nonprofit and profit organizations are initiating and expanding their WGS project to build and understand individuals' genomes. Researchers are expecting to discover molecular biomarkers, identify potential drug targets, enable clinical trials, and accelerate systems medicine and emerging precision medicine for predicting, preventing, diagnosing, and treating diseases [36]. Fig. 4 shows a WGS workflow using the HiSeq X and analysis pipeline.

There are several interesting human genome sequencing projects that continue this research throughout the world.

(1) The Human Genome Project–Write (HGP-Write) is a ten year extension of the HGP to synthesize the human genome (Science June 2, 2016 and The Scientist, June 2, 2016). HGP re-
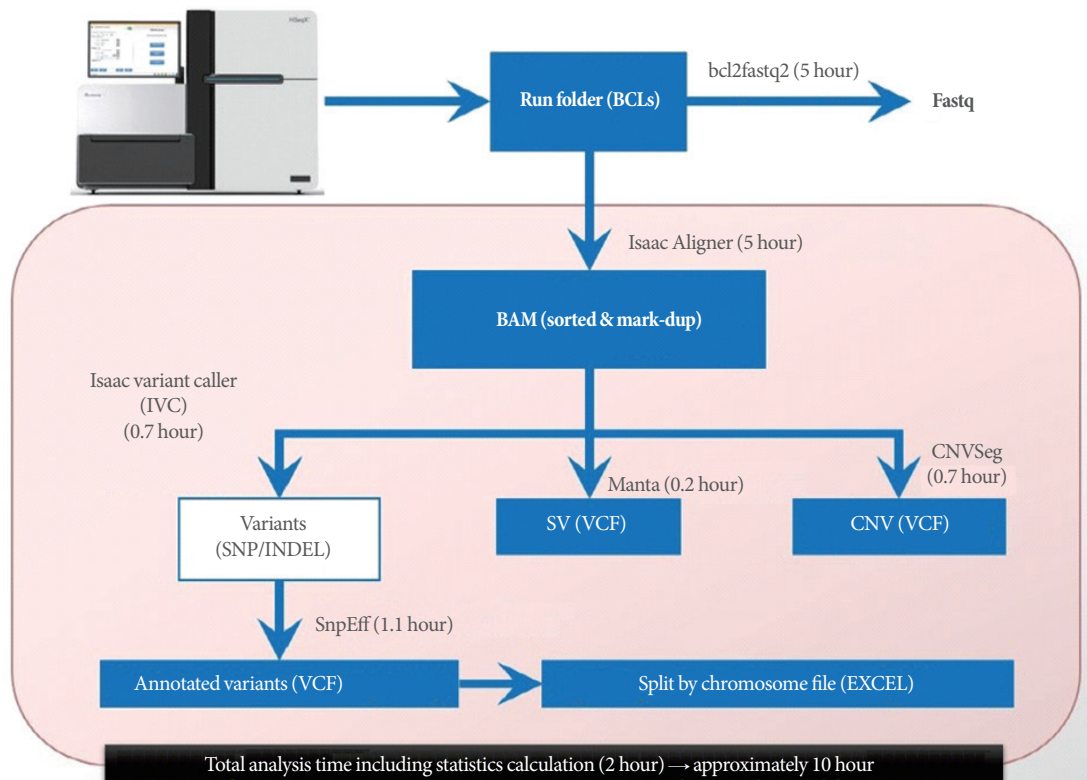
Fig. 4. Whole genome sequencing application workflow based on HiSeq X system. This figure is provided from a genomic sequencing service provider with HiSeq X system. Actual workflow will vary by institutions and companies. BCL, basecall file; BAM, the binary version of a SAM (sequence alignment/map); mark-dup, marks duplicate reads; SNP, single nucleotide polymorphism; INDEL, insertion or deletion of bases; SV, structural variation; VCF, variant call format; CNV, copy number variation.

vealed that the human genome consists of 3 billion DNA nucleotides and HGP-Write will try to synthesize large portions of the human genome for scientific and medical advances [37,38]. This project will be managed by the Center of Excellence for Engineering Biology, a new nonprofit organization.

(2) The 100,000 Genomes Project is continuing the sequencing project from the 1,000 Genome Project. Genomics England, a government-owned company, has introduced this new 100,000 Genome Project to sequence 100,000 whole genomes from National Health Service (NHS) patients and their families. The aim of this project is to create a new genomic medicine service for the NHS and to enable new medical research that combines genomic sequence data with medical records. Researchers will study the best way to use and interpret genomic data for healthcare and investigate the cause, diagnosis, and treatment of diseases [39].

(3) The most recent human genome project is GenomeAsia 100K (GA100K). GA100K is a mission-driven nonprofit con-

sortium consisting of Nanyang Technological University (NTU), Macrogen, and MedGenome. NTU is a research-intensive public university and has a new medical school, the Lee Kong Chian School of Medicine, set up jointly with Imperial College London. Macrogen is a world leading genetic service provider with global locations including Korea, the United States of America, Japan, and The Netherlands, and a spinoff company from the Genomic Medicine Institute in Seoul National University. MedGenome is a genomic-driven research and diagnostic company and the market leader of genetic diagnostic testing in India. This consortium will collaborate to sequence and analyze 100,000 Asian individuals' genomes to help accelerate population specific medical advances and precision medicine. Asians are significantly underrepresented in current genomic studies with Caucasian based human genome reference and databases even though there are unique genetic differences between South and East Asians. GA100K plans to create reference genomes for Asian populations as well as identify

population-specific alleles. With this project, GA100K expects to understand biology of diseases and identify new possible therapeutic drugs [40].

CLOSING REMARKS AND PROSPECTIVE

In this short review article, we summarized and discussed the current trends of genomics and the reference genomes being built with new NGS technologies that can be applied to personalized and precision medicine. NGS technologies have enabled scientists to interrogate biological systems with population-scale genomics and have been increasingly popular for biological and clinical research. The low cost and high throughput data of NGS technologies provide more opportunities in human genetics and genomics along with newly developed data algorithms. Sanger sequencing initiated HGP, which provided a human reference genome with huge amounts of data. The 1,000 Genome Project generated genome data from 2,504 individuals in 26 different populations for 5 years. The most recent NGS instrument, the HiSeq X Ten, allows researchers to have 20,000 genomes available from the NHLBI TOPMed program in 2015. This trend in human genome research initiated the Precision Medicine Initiative by the U.S. government [41] and other, similar projects/programs by countries and profit/nonprofit organizations have also been progressing. The accumulation of human genome data and have demonstrated the importance of appropriate human reference genomes as an aspect of personalized medicine. The 100,000 Genome Project and GenomeAsia 100K are major projects for accelerating specific population-scale medical advances and precision medicine.

Third-generation sequencing technologies have been introduced and used mostly at the research level. There are some studies that compare NGS and third-generation sequencing technologies [42]. So far, research is still more dependent on NGS technologies, and third-generation sequencing technologies are aiding and/or supplementing NGS methods. However, we may expect another paradigm shift in genomics in the near future.

REFERENCES

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74:5463-7.
2. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A 1977;74:560-4.
3. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet 2014;30:418-26.
4. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008;26:1135-45.
5. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. Genomics 2010;95:315-27.
6. GS FLX+ System [Internet]. Branford (CT): Roche Diagnostics Co.; c1996-2016 [cited 2016 Oct 1]. Available from: http://454.com/products/gs-flx-system/.
7. Ion Proton System Specifications [Internet]. Waltham (MA): Thermo Fisher Scientific Inc.; c2016 [cited 2016 Oct 1]. Available from: https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-proton-system-for-next-generation-sequencing/ion-proton-system-specifications.html.
8. HiSeq X System Specifications [Internet]. San Diego (CA): Illumina Inc.; c2016 [cited 2016 Oct 1]. Available from: http://www.illumina.com/systems/hiseq-x-sequencing-system/performance-specifications.html.
9. DNA Sequencing Costs: Data [Internet]. Bethesda (MD): National Human Genome Research Institute; 2016 [cited 2016 Oct 1]. Available from: https://www.genome.gov/sequencingcostsdata/.
10. The original long-read sequencer [Internet]. Menlo Park (CA): Pacific Biosciences; c2015-2016 [cited 2016 Oct 1]. Available from: http://www.pacb.com/products-and-services/pacbio-systems/rsii/.
11. Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods 2015;12:780-6.
12. Shi L, Guo Y, Dong C, Huddleston J, Yang H, Han X, et al. Long-read sequencing and de novo assembly of a Chinese genome. Nat Commun 2016;7:12065.
13. Baker M. De novo genome assembly: what every biologist should know. Nat Methods 2012;9:333-7.
14. Ng PC, Kirkness EF. Whole genome sequencing. Methods Mol Biol 2010;628:215-26.
15. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57-63.
16. Niu W, Lu ZJ, Zhong M, Sarov M, Murray JI, Brdlik CM, et al. Diverse transcription factor binding features revealed by genome-wide ChIP-seq in C. elegans. Genome Res 2011;21:245-54.
17. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, et al. The *Mycobacterium tuberculosis* regulatory network and hypoxia. Nature 2013;499:178-83.

18. Settings M. Metagenomics versus Moore's law. Nat Methods 2009; 6:623.

19. Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, et al. Quality assessment of the human genome sequence. Nature 2004;429:365-8.

20. GenBank Overview [Internet]. Bethesda (MD): National Center for Biotechnology Information, U.S. National Library of Medicine; 2016 [updated 2016 Mar 11; cited 2016 Oct 1]. Available from: http://www.ncbi.nlm.nih.gov/genbank/.

21. e!Ensembl [Internet]. Cambridgeshire (UK): Ensembl, EMBL-EBI; 2016 [cited 2016 Oct 1]. Available from: http://useast.ensembl.org/index.html.

22. Genome Browser Gateway [Internet]. Cambridgeshire (UK): Ensembl, EMBL-EBI; 2016 [cited 2016 Oct 1]. Available from: http://useast.ensembl.org/index.html.

23. Human Genome Overview [Internet]. Bethesda (MD): National Center for Biotechnology Information, U.S. National Library of Medicine; 2016 [cited 2016 Oct 1]. Available from: http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/.

24. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol 2007;5:e254.

25. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452:872-6.

26. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53-9.

27. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, et al. The diploid genome sequence of an Asian individual. Nature 2008;456:60-5.

28. Kim JI, Ju YS, Park H, Kim S, Lee S, Yi JH, et al. A highly annotated whole-genome sequence of a Korean individual. Nature 2009;460:1011-5.

29. Chambers JC, Abbott J, Zhang W, Turro E, Scott WR, Tan ST, et al. The South Asian genome. PLoS One 2014;9:e102645.

30. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 2003;4:981-94.

31. Black WC 4th, Baer CF, Antolin MF, DuTeau NM. Population genomics: genome-wide sampling of insect populations. Annu Rev Entomol 2001;46:441-69.

32. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 2010;11:415-25.

33. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015;526:68-74.

34. UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. The UK10K project identifies rare variants in health and disease. Nature 2015;526:82-90.

35. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061-73.

36. Trans-Omics for Precision Medicine (TOPMed) Program [Internet]. Bethesda (MD): NHLBI Health Information Center; [updated 2015 Oct; cited 2016 Oct 1]. Available from: https://www.nhlbi.nih.gov/research/resources/nhlbi-precision-medicine-initiative/topmed.

37. Boeke JD, Church G, Hessel A, Kelley NJ, Arkin A, Cai Y, et al. GENOME ENGINEERING. The Genome Project-Write. Science 2016;353:126-7.

38. Akst J. "Human Genome Project-Write" Unveiled [Internet]. The Scientist. 2016 Jun 2 [cited 2016 Oct 1]. Available from: http://www.the-scientist.com/?articles.view/articleNo/46237/title/-Human-Genome-Project-Write--Unveiled/.

39. The 100,000 Genomes Project [Internet]. London (UK): Genomics England; [cited 2016 Oct 1]. Available from: https://www.genomicsengland.co.uk/the-100000-genomes-project/.

40. GenomeAisa 100K [Internet]. GenomeAisa 100K [cited 2016 Oct 1]. Available from: http://www.genomeasia100k.com/.

41. White House announces efforts to accelerate precision medicine initiative [Internet]. New York (NY): GenomeWeb; c2016 [cited 2016 Oct 1]. Available from: https://www.genomeweb.com/molecular-diagnostics/white-house-announces-efforts-accelerate-precision-medicine-initiative.

42. Case study: Assembling high-quality human genomes–Beyond the $1,000 genome. PacBio literature [Internet]. Menlo Park (CA): Pacific Biosciences; Pacific Biosciences of California Inc.; c2015-2016 [cited 2016 Oct 1]. Available from: http://www.pacb.com/literature/case-study-assembling-high-quality-human-genomes-beyond-the-%C2%911000-genome%C2%92/.