Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

19 Genomics of Infectious Diseases and Private Industry

*Guy Vernet**

Fondation Mérieux, Lyon, France

19.1 Introduction

Genomics is the science of studying genomes of living organisms. The main driver of companies that specialize in genomics is human genome sequencing and most of the technology breakthroughs have been achieved to improve the speed and quality of human genome sequencing and to reduce its cost. Sanger sequencing was used for the 13-year-long Human Genome Project, which resulted in the first whole-genome sequence in 2003 for a budget of $2.7 billion. Five years later, the same result was obtained in 5 months for just $1.5 million (Voelkerding et al., 2009). The National Human Genome Research Institute (NHGRI) part of the NIH awarded in 2008 more than $20 million in grants (the $1,000 Genome grants) to develop innovative sequencing technologies inexpensive and efficient enough to sequence a person's DNA as a routine part of biomedical research and health care for less than $1000. The 1000 Genomes Project, a China, Germany, UK, and US collaboration, is currently sequencing the whole genome from 2000 individuals worldwide to identify genetic variations (Via et al., 2010). In the 1990s, microarray technologies have been developed to address mostly needs of human genomics. As of May 2010, 21,520 scientific papers using the Affymetrix (Santa Clara, CA, USA) proprietary microarray technology have been listed on the web site of the company (www.affymetrix.com). Microarrays have also been used for resequencing or gene expression monitoring of infectious agents. New sequencing technologies and instruments, referred to as Next-Generation Sequencing (NGS), have appeared during the last 5 years and interestingly were first used to sequence the whole genome of human pathogens *Mycoplasma genitalia* (Margulies et al., 2005) and *Escherichia coli* (Shendure et al., 2005).

Understanding hosts, vectors, and pathogens' genomes, as well as their transcriptomic and epigenomic modifications following infection, during the course of the disease and under treatment will ultimately lead to personalized medicine for which genome characteristics will be important to tailor treatments (Snyder et al., 2010).

Several biotech companies have been created in the last 15 years to develop and market systems for the analysis of genomes, many of them by scientists issued

* Email: guy.vernet@fondation-merieux.org

Genetics and Evolution of Infectious Diseases. DOI: 10.1016/B978-0-12-384890-1.00019-4

from universities. Since then, they have often been acquired by larger companies that are major players in the field of in vitro diagnostics.

19.2 Customers and Their Needs

19.2.1 Customers

Customers of companies involved in genomics are both public and private organizations. Research laboratories from universities or institutes have different needs, which were, up to now, impossible to address with a single technology or instrument. Due to high prices of equipment and maintenance and frequent needs for instrument upgrade, platforms serving the needs of several research laboratories and providing services to the scientific community have emerged, which provide a full set of equipments and dedicated human resources.

The Wellcome Trust Sanger Institute (Cambridge, UK) is an example where such a platform, linked to bioinformatics resources serves different research projects, including a pathogen genetic group exploring parasites especially malaria—and viruses genomics. The Broad Institute (Cambridge, MA, USA) is another example with a genome-sequencing platform that also considers fungi, bacteria, and virus genomes. The J. Craig Venter Institute, a not-for-profit private organization based in Rockville, Maryland (USA), has more than 500 scientists and staff, more than 250,000 square feet of laboratory and operates several resource centers (sequencing, genotyping, functional genomics, bioinformatics) for infectious disease genomics. In China, the Beijing Genomics Institute (BGI) from the Chinese Academy. of Sciences, based in Shenzhen and Hong Kong will have 3500 staff by the end of 2010 and has acquired 128 HiSeq 2000 NGS machines from Illumina (San Diego, CA, USA) for sequence service and its own research projects, some of them on infectious diseases like the severe acute respiratory syndrome (SARS)-Coronavirus (Cyranoski, 2010).

Such large institutes are not very numerous worldwide. However, they attract large international budgets that generate important sells for genomics companies in terms of instruments, maintenance, and reagents.

Genomics platforms often provide services to remain competitive through return on investment. Thus, they represent potential customers for genomics companies. A list of companies and public facilities that provide DNA-sequencing services in different parts of the world can be found at http://www.nucleics.com/DNA_sequencing_support/sequencing-service-reviews.html. It is quite complete, although it does not mention organizations like Illumina, Beckman Coulter, or the Wellcome Trust Sanger Institute. It is a very useful tool for scientists who need to choose a service company and includes, when available, prices, DNA-sequencing instruments used by each facility, specific DNA template and primer requirements, sample shipping and sequencing facility contact details, whether the DNA-sequencing facility is GLP/FDA certified, and the other DNA-sequencing-related services offered by the facility. There are more than 230 such facilities worldwide, including many universities in the USA and Canada. Many of them are compliant with good clinical

practices (GCP), good laboratory practices (GLP), and good manufacturing practices (GMP) or, for clinical diagnostic services to the Clinical Laboratory Improvement Amendments (CLIA) regulations.

Pharmaceutical companies also have needs for complex genomic data for R&D purposes—new molecules or vaccine developments, toxicity and pharmacokinetic studies—although there is a trend for outsourcing upstream research to public laboratories or biotechnology companies.

At the other extremity of data complexity, high-level public or private laboratories, such as those in hospitals, use technologies based on genomic information for diagnostic, forensic, or surveillance needs. Although currently very centralized due to the requirement for sophisticated equipments and relatively high skills, these techniques will probably reach more and more customers in a near future. Finally, pharmaceutical and agro-food companies also use low-complexity data for quality control purposes ensuring product biosafety.

19.2.2 Research Needs

Whole-Genome Sequencing, Comparative-Genome Sequencing, and Targeted Resequencing

The first need of scientists is to sequence microbe genomes. The whole-genome sequence of 1987 viruses, 916 bacteria, and 67 archeal species were deposited to GenBank release 2.2.6 (Wooley et al., 2010), and these numbers grow rapidly, the emphasis being on species that are pathogenic for animals or plants. Sequencing will identify single nucleotide polymorphisms (SNPs), insertions, and deletions (indels) and large chromosomal rearrangements (structure or copy number changes).

Once a reference sequence as been established for a given species, the need is for resequencing (i.e., comparison of the sequence of new isolates to this reference sequence to identify differences) (Herring and Palsson, 2007). This comparison concerns the whole genome or targets genes of interest where modifications are expected to have consequences on the phenotype. Scientists will use these data for fundamental research: to annotate all genes, understand genome organization, classify species, and study their evolution. The knowledge of genome organization, combined with functional genomics is the basis to understand pathobiological mechanisms of infectious agents. Among downstream applications are the development of new drugs, vaccines, and diagnostics including the establishment of resistance and escape profiles. Surveillance networks will also use genomic information to monitor pathogens evolution: resistance to treatment, crossing of host speciation barriers, increased transmissibility, and virulence.

Metagenomics

A new field of investigation is the study of microbiomes or metagenomics (Wooley et al., 2010), which consists in the systematic sequencing of all nucleic acids in a given ecological "niche"—gut, upper respiratory tracts, and skin—to identify all

microbes. This allows us to characterize the "normal" flora, for example at different ages in life, and interactions between this flora in pathogenic agents during infection, in particular the exchange of genetic material conferring resistance and virulence. Applications by pharmaceutical, diagnostic, and agro-food companies are very important. The study of microbiomes also have a potential application for forensics: sequencing the "bacteriome" in traces left on can be used to identify people as the flora present on the skin is a signature depending on food habits, environment, and diseases.

Functional Genomics

Transcriptome analysis is the characterization of all coding and noncoding transcriptional activity in any organism without a priori assumptions through annotation of SNPs and mapping to reference genomes, characterization of transcript isoforms, regulatory RNAs, or splice junctions and determination of the relative abundance of transcripts (gene expression analysis). Analysis of differential gene expression is important in hosts and pathogens as well as in vectors of transmissible diseases (mostly insects). Human, plant, or animal cells can be studied when they are confronted to infection to identify the mechanisms targeted by the pathogen and those by which they resist infections. Pathogens' gene expression during infection of the cell is an important area of investigation as it may reveal pathobiological mechanisms and targets for new drugs. Epigenome analysis is the study of chromatin structure and gene regulation by CpG methylation, histone modifications, or DNA−protein interactions. Besides fundamental research, functional genomics can find applications in pharmaceutical companies for new anti-infection drugs and diagnostic companies to identify new biomarkers for diagnostic or disease or treatment monitoring.

19.2.3 Diagnostic Needs

Genomics research generates tremendous amount of information. Among this are sequences that can be used to detect infection by a pathogen species by simple molecular techniques like polymerase chain reaction (PCR), transcription-mediated amplification (TMA), or nucleic acid sequence-based amplification (NASBA), which, if quantitative, can also help monitor treatment efficacy against, for example, human immunodeficiency virus (HIV) or hepatitis viruses. A syndrome-based approach (i.e., the detection of multiple pathogens responsible for a disease, such as pneumonia, fever, neurological diseases, diarrhea) may also be useful. Signature sequences can identify infection by a variant with a specific phenotype with given virulence, host specificity, or resistance to treatment and help patient care or epidemiological surveillance. The latter signatures can be entire genes (acquired by horizontal transmission, either plasmid or recombination), individual SNPs, groups of SNPs carried by a unique or different genes, indels, or even modified expression of a gene. A comprehensive assay associating pathogen identification, pathogen

typing, identification of virulence, or resistance markers may prove valuable in chronic infections like HIV, hepatitis virus infections, or tuberculosis.

19.3 Technologies and Companies

Microarrays, Sanger sequencing, and NGS generate large quantities of data necessary for whole-genome resequencing, targeted resequencing, gene copy number variations, gene expression analysis, chromosome structural changes, or protein−DNA interactions. However, microarray analysis, by definition, requires a priori knowledge of sequences and only sequencing allows determination of unknown sequences. Only NGS allows rapid and low-cost whole-genome sequencing and metagenomics, as well as massive parallel processing of multiple specimens. Less complex technologies like pyrosequencing or low- to medium-density microarrays can be used for targeted sequencing or resequencing of one or a few genome regions which is of interest for patients management or for molecular epidemiology. Molecular diagnostics using PCR, TMA, NASBA, loop-mediated isothermal amplification (LAMP) have many more applications for infectious diseases research, epidemiological surveillance and treatment. However, we will not address this domain of activity which exploits data issued from genomics research.

19.3.1 Microarray Companies

At least 36 companies providing microarrays have been identified in 2009 in the USA and Europe (North Shore LIJ Research Institute; http://www.nslij-genetics.org/microarray/). Most of them propose low- to medium-density custom arrays, which are glass plates or beads with DNA probes either spotted or synthesized in situ using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using premade masks, photolithography using dynamic micromirror devices, ink-jet printing, or electrochemistry on microelectrode arrays. For a recent review on the use of microarrays for clinical microbiology, see Miller and Tang (2009). We will focus on companies providing high-density microarrays.

The Affymetrix GeneChip® technology, based on photolithography to synthesize probes in situ on the array, was invented in the late 1980s. More than 1500 publications can now be retrieved from the Affymetrix database with the key words "virus," "bacteria," "parasite," and "fungi." The current company platform GCS 3000Dx v.2 is 510(k) cleared and CE marked for in vitro diagnostic use and consists of a scanner, a fluidics station, and the Affymetrix Molecular Diagnostics Software (AMDS) for data interpretation. The Human Genome U133 Plus 2.0 array analyzes over 47,000 transcripts from human genome and allows gene expression profiling of cells infected by various pathogens. Several commercial arrays address human pathogens. The *E. coli* array contains probe sets to detect transcripts from the K12 strain of *E. coli* and three pathogenic strains of *E. coli*. It includes approximately 10,000 probe sets for all 20,366 genes present in four strains of *E. coli* over the entire open reading frame (ORF) of *E. coli*, over 700 intergenic regions as well

as probe sets for various antibiotic resistance markers. The *Staphylococcus aureus* Genome Array allows the analysis of the expression of sequences in four strains of *S. aureus*. It contains probe sets to over 3300 *S. aureus* ORFs and to study both forward and reverse orientation of over 4800 intergenic regions. The *Pseudomonas aeruginosa* Genome Array represents the annotated genome of *P. aeruginosa* strain PA01 and includes 5549 protein-coding sequences, 18 tRNA genes, a representative of the ribosomal RNA cluster, and 117 genes present in strains other than PA01. In addition, 199 probe sets corresponding to all intergenic regions exceeding 600 base pairs have been included. The Plasmodium/Anopheles Genome Array includes probe sets to over 4300 *Plasmodium falciparum* transcripts and approximately 14,900 *Anopheles gambiae* transcripts.

The SARS Resequencing Array provides a standard assay for complete sequence analysis of the SARS coronavirus.

BioMérieux (Marcy-l'Etoile, France) has developed several resequencing microarrays covering pathogens genomes. A *Mycobacterium tuberculosis* array is based on two sequence databases: one for the species identification of mycobacteria (82 unique 16S rRNA sequences corresponding to 54 phenotypical species) and the other for detecting *M. tuberculosis* rifampin resistance in rpoB (Troesch et al., 1999; Sougakoff et al., 2004). An *S. aureus* array tiles 16S rDNA sequences to identify staphylococcus species (Couzinet, 2005a), grlA, gyrA, grlB, and gyrB genes for the presence of mutations involved in fluoroquinolone resistance (Couzinet, 2005b), and multilocus sequence typing (MLST) of *S. aureus* strains (van Leeuwen et al., 2003). An HIV microarray was designed to detect 204 antiretroviral resistance mutations simultaneously in Gag cleavage sites, protease, reverse transcriptase, integrase, and gp41 of HIV1 (Gonzalez et al., 2004). Similarly, a hepatitis B virus (HBV) microarray was designed to detect 245 mutations, 20 deletions, and 2 insertions at 151 positions and to determine the genotype of the HBV (Tran et al., 2006; Pas et al., 2008).

Roche NimbleGen (Madison, WI, USA) manufactures custom, high-density DNA arrays based on its proprietary Maskless Array Synthesizer (MAS) technology using a Digital Micromirror Device (DMD) combined with DNA synthesis chemistry allowing 385,000 to 2.1 million unique probe features in a single array. Arrays are synthesized on standard-sized glass microscope slides and are compatible with a range of hybridization, washing, and scanning instrumentation. With the new generation of HD2 arrays, oligos between 50 and 75 bases in length can be synthesized, increasing sensitivity, specificity, and reproducibility.

Recently, Roche NimbleGen introduced Sequence Capture arrays to produce targeted, sequencing-ready samples for use with NGS instruments. High density NimbleGen arrays with long oligonucleotides are used to hybridize either the whole human exome or human genomic regions of interest are hybridized. The purified human sequences are then eluted, amplified and sequenced. Although currently restricted to human the genome, and more recently, to wheat and rapeseed, this technology may prove valuable for infectious diseases genomics. Agilent Technologies (Santa Clara, CA, USA) and Febit (Heidelberg, Germany) also market capture microarrays.

19.3.2 NGS Companies

NGS is parallel sequencing of clonally amplified or single DNA molecules by iterative cycles of polymerase-based extension or oligonucleotide ligation that takes place in flow cells. These technologies have revolutionized all aspects of genomics: whole-genome sequencing, targeted resequencing, metagenomics, gene expression profiling, epigenomics, and DNA−protein interactions study (ChIP-Seq). For a recent review on NGS, see Holt and Jones (2008) and Voelkerding et al. (2009). Three companies have launched NGS platforms requiring clonal amplification: 454 Life Sciences, a Roche company (Branford, CT, USA), Illumina Inc. (San Diego, CA, USA), and Applied Biosystems, a division of Life Technologies (Carlsbad, CA, USA); one company has launched a system sequencing single DNA molecules: Helicos Biosciences Corporation (Cambridge, MA, USA). The main characteristics of these platforms are listed in Table 19.1.

454 Life Sciences was created in 2000 and released the first NGS platform, Genome Sequencer 20, in 2005. It was used to sequence the first human genome for less than US$1 million in 2006. 454 Life Sciences has been acquired by Roche Diagnostics in January 2007. The current 454 Life Sciences platform, the Genome Sequencer FLX system, is using the 454's sequencing-by-synthesis (SBS) technology for de novo sequencing, resequencing of whole genomes, and target DNA regions, metagenomics, and RNA analysis. The chemistry used for sequencing is described in Figure 19.1A. Automation using a magnetic beads technology simplifies emulsion-PCR and allows library preparation of genomics samples in hours in a single tube, eliminating cloning, and colony picking. The recognized advantage

Table 19.1 Characteristics of NGS Platforms

Platform	GS FLX	HiSeqTM2000	SOLiD™ 4	Genetic Analysis System
Company	Roche	Illumina	Applied Biosystems	Helicos
Throughput/run	1 million reads	1 billion reads	1.4 billion reads	1 billion reads
	0.4−0.6 Gb	Up to 200 Gb	Up to 180 Gb	Up to 35 Gb
Run duration	10 h	2.5 days	6 days	8 days
Multiplexing (samples/run)	8	200 (gene expression profiling)	48 RNA, 96 DNA	Up to 4800
Base call quality (manufacturer data)	99% of bases at QV20*	>90% at QV30*	80% at QV30*	0.2% error rate (SNPs)
Human genome coverage	>20	30	30	28
Read length	400 bp	100 bp	50 bp	35 bp
Paired end read	Yes	Yes	Yes	Yes

*Phred score quality value (QV): miscall probabilities of 10%, 1%, and 0.1% yield QV of 10, 20, and 30, respectively.

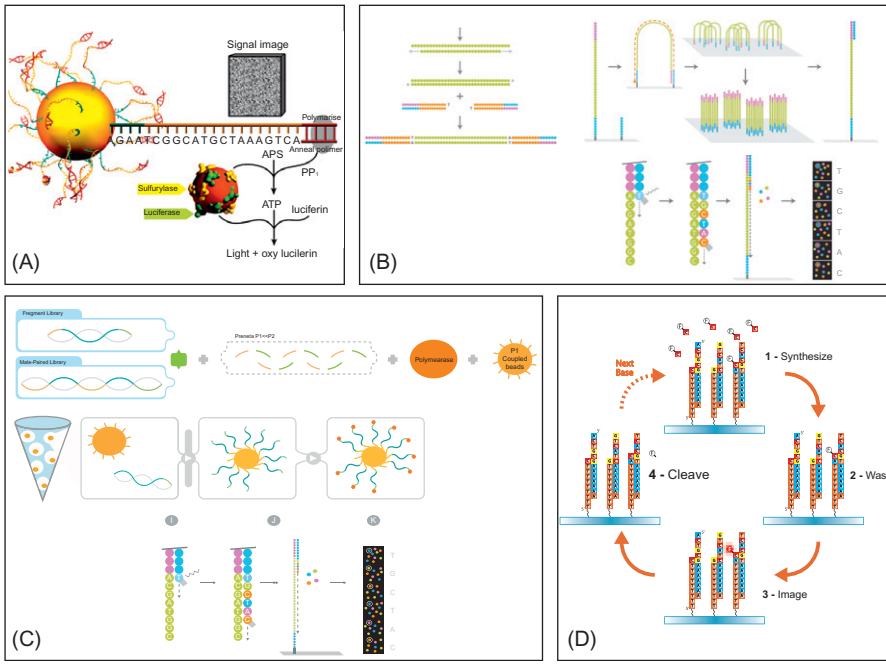

Figure 19.1 (A) 454 technology. Template DNA is fragmented with adapters added at both ends, and clonal amplification is done by emulsion-PCR using magnetic beads. Each single bead is added to a well of a picotiter plate and iterative pyrosequencing is used for sequencing. *Source:* 454 Sequencing. Copyright 2010. Roche Diagnostics. (B) Illumina HiSeq technology. Template DNA is fragmented, adapters are added at both ends, and DNA is attached to the flow cell. Bridge amplification generates clonal clusters and iterative SBS is performed. *Source:* Copyright 2010. Illumina Inc. (C) Applied Biosystems SOLiD technology. Template DNA is fragmented, adapters are added at both ends and clonal amplification is done by emulsion-PCR using magnetic beads. Sequencing is done by iterative ligation using a set of four fluorescently labeled di-base probes. *Source:* Copyright 2010. Life Technology Inc. (D) Helicos tSMS technology. Original DNA samples are first fragmented, the DNA double-helix is melted into single strands and a polyA tail is added to these DNA molecules. Billions of these single DNA molecules are captured on a proprietary surface within a flow cell and serve as templates for the SBS process. Genomic DNA is fragmented, polyA tail are added, and DNA molecules captured by oligo dT primers inside the flow cell. The tSMS process is a cyclical process involving multiple rounds of (1) synthesis using labeled nucleotides, (2) washing, (3) imaging, and (4) cleaving the fluorescent label until the desired read length is achieved. *Source:* Copyright 2010. Helicos Biosciences corp. All rights reserved.

of the FLX system compared to other NGS platforms is that it generates the longest single reads (400 bp) and long paired end reads of 20 kb, 8 kb, or 3 kb. In May 2010, the company introduced the GS Junior System which provides an integrated sequencing and bioinformatics solution, all in the size of a typical desktop laser printer. The company announces improvements of its sequencing chemistry to

increase reads length to 1000 bp. There have been 748 published studies using the 454 technology, among which 219 concern infectious diseases. The technology has found applications in metagenomics, such as identification of a new arenavirus in transplantation patients (Palacios et al., 2008) or the characterization of microflora in oral cavity or guts (Keijser et al., 2008; Turnbaugh et al., 2009).

Illumina was created in 1998 to exploit rights on the BeadArray technology developed at Tufts University. The HiSeq TM 2000 platform is based on the SBS chemistry, which generates reads of 100 bp and paired end reads allowing the assembly of long scaffolds (Figure 19.1B). After library preparation, cluster generation is done on the cBot automated cluster generation system, which significantly reduces hands-on time compared to emulsion-PCR. Sequencing a genome can be done on one flow cell while, simultaneously, the other flow cell can analyze its epigenome and transcriptome. More than 500 publications illustrate the versatility of Illumina NGS technology. Recent publications on infectious diseases include *Trichinella spiralis* (Webb and Rosenthal, 2010), *P. falciparum* (Jiang et al., 2010), virus discovery in Drosophila cells and adult mosquitoes (Wu et al., 2010), methicillin-resistant *Staphylococcus aureus* (MRSA) molecular epidemiology (Harris et al., 2010), *Burkholderia cenocepacia* therapeutic targets (Yoder-Himes et al., 2010), and *Pasteurella multocida* virulence factors (Steen et al., 2010). The Illumina platform is the most versatile of the three NGS platforms requiring clonal amplification because it associates gigabases outputs and read lengths of 100 bp.

Life Technologies was created by the combination of Invitrogen Corporation and Applied Biosystems Inc. in 2008. Applied Biosystems commercializes a NGS platform called SOLiD 4. The chemistry used in this platform is described in Figure 19.1C. It is based on emulsion-PCR and oligonucleotides ligation. The SOLiD 4 platform has two flow cells allowing two independent experiments at the same time and can multiplex up to 96 samples. Applied biosystems proposes automated solutions for reproducible templated bead preparation with less than 1 h of hands-on time (EZ Bead™ System). The SOLiD platform generates the shortest read lengths among the three platforms, which makes it less versatile for the various applications. However, future evolutions of the SOLiD platforms, SOLiD 4hq, and SOLiD PI will generate longer reads (75 bp).

Helicos Biosciences Corporation (Cambridge, MA, USA) is commercializing the first platform that allows sequencing from single DNA molecules, thereby avoiding biases due to amplification. As the three companies described earlier, Helicos is a recipient of the "$1000 Genome" grant. The chemistry used on the Helicos Genetic Analysis System is based on tSMS technology (Figure 19.1D) in which single-stranded DNA molecules generated from a library and tagged with a polyA tail are attached to a proprietary surface at a density of up to 100×10^6 molecules per square centimeter and sequenced by synthesis. This system is now installed in several research centers in the USA and Europe. The Helicos platform was used for the first single-molecule whole-genome sequencing in 2009 (Pushkarev et al., 2009). This study achieved 28× average coverage of the human genome and detected over 2.8 million SNPs, of which over 370,000 were novel. Validation with a genotyping array demonstrated 99.8% concordance. The unbiased

nature of the single-molecule sequencing approach also allowed the detection of 752 copy number variations in this genome.

The NGS technologies generate a much higher amount of data than the well-established ABI Sanger platform. The analysis of gigabases or terabases requires complex software solutions. A list of software used with NGS platforms can be found in Voelkerding et al. (2009).

However, performances claimed by manufacturers may be overestimated. Harismendy et al. (2009) have compared the platforms of 454 Life Sciences, Illumina, and Applied Biosystems on a 260 kb human genome sample. Although the Illumina and Applied Biosystems produce the largest amounts of data, only 43% and 34% of them, respectively, are usable after quality filtration. In contrast, 95% of the data generated by the 454 platform are usable. All three technologies have biases that induce heterogeneous coverage of bases along the sequence: the 454 platform shows the lowest variability among unique and repetitive sequences, whereas the ABI and Illumina platforms tend to be affected by high Adenine/Thymine contents. NGS platforms tend to better detect indels than ABI Sanger platform as they sequence single strands. As expected, ABI Sanger sequencing has an error rate of approximately 7% and careful comparison with NGS reveals false positive and false negative rates of 0.9% and 3.1%. The overall sequencing accuracy of NGS platforms was very high (>99.99%), but the ability to detect variant was 95% for the 454 platform (which has the lowest sensitivity), 100% for the Illumina platform, and 96% for the ABI platform, the last two technologies being less specific. Overall, NGS platforms need to improve their uniformity of per-base sequence coverage as accuracy is lower in poorly covered regions.

New technologies based on nanopores are currently being explored to develop platforms for single-DNA molecule sequencing (Branton et al., 2008). A nanopore-based device provides single-molecule detection and analytical capabilities that are achieved by electrophoretically driving molecules in solution through a nano-scale pore. The nanopore provides a highly confined space within which single nucleic acid polymers can be analyzed at high throughput by one of a variety of means, and the perfect processivity that can be enforced in a narrow pore ensures that the native order of the nucleobases in a polynucleotide is reflected in the sequence of signals that is detected. Kilobase length polymers (single-stranded genomic DNA or RNA) or small molecules (e.g., nucleosides) can be identified and characterized without amplification or labeling, a unique analytical capability that makes inexpensive, rapid DNA sequencing a possibility. Further research and development to overcome current challenges to nanopore identification of each successive nucleotide in a DNA strand offers the prospect of third-generation instruments that will sequence a diploid mammalian genome for approximately $1000 in approximately 24 h. Oxford Nanopore Technologies (http://www.nanoporetech.com; Oxford, UK), our first generation of DNA-sequencing technology, uses a protein nanopore combined with a processive enzyme, multiplexed on a silicon chip. This elegant and scalable system has unique potential to transform the speed and cost of DNA sequencing. Future generations may interrogate single strands of DNA and may use "solid-state" nanopores for further improvements in speed and cost. Lingvitae

(http://www.lingvitae.com; Oslo, Sweden) and Eid et al. (2009) from Pacific Biosciences (Menlo Park, CA, USA) have designed a prototype instrument able to sequence a single DNA molecule using DNA polymerase based on the observation of the temporal incorporation of labeled nucleotides, which takes place in a nano-photonic structure. This can drastically reduce the volume of observation. Currently, this prototype is able to multiplex 3000 such structures, allowing sequencing of small viral or bacterial genomes with high accuracy.

19.4 Conclusion and Perspectives

Microarrays and NGS have revolutionized genomics because they drastically increased scientists' access to tremendous amounts of information on genomes and gene expression and decreased time-to-result, hands-on time, and costs. Research needs for these technologies are well understood but still represent small (although rapidly growing) markets. The total market of NGS was evaluated to be $484 million in 2008 (http://www.researchandmarkets.com/reportinfo.asp?report_id=614823) and the market of functional genomics was about $2.2 billion in 2007 with an annual growth rate of 18% (http://www.researchandmarkets.com/reportinfo.asp?report_id=5545). The real development of sales of companies proposing technologies for genomics will mostly come from diagnostic applications in human or animal health. Genomics research can be translated into molecular diagnostics in the fields of human genetic, oncology, and infectious diseases. Several in vitro diagnostic companies propose instruments and reagents to detect sequence signatures for diagnosis or treatment monitoring applications. However, the current clinical needs mostly require low-complexity genetic information that can be covered by multiplex real-time amplification, reverse hybridization-based line probe assay, low-density microarrays or simple sequencing platforms like the PyroMark Q24 platform. The place of technologies generating more complex datasets in clinical applications is still to be defined. Developing such applications will require clinical validation of the value of complex sequence information and strong efforts for clinician education on its benefit. It will also require that costs of instruments and reagents further decrease and that the technologies become easier to use and more robust. All the major companies cited in this chapter make efforts to simplify equipments and reduce costs per analysis. Affymetrix recently launched a new, more affordable and smaller platform, GeneAtlas, along with a microarray strip format, which enables users to process up to two strips per day or eight strips per week. 454 Life Sciences announced the release of the GS Junior System scaled to suit the needs of individual laboratories for rapid sequencing of amplicons, targeted human resequencing studies, de novo sequencing of microbial and other small genomes, and for pathogen detection. Illumina recently launched the Genome Analyzer$_{IIe}$ with a lower cost, making the technology more accessible to laboratories of various sizes. Similarly, Applied Biosystems announces a less expensive, low-throughput benchtop platform (50 Gb per run) will allow shorter times-to-results. Qiagen (Hilden, Germany) is marketing a relatively simple, low footprint platform (PyroMark Q24) for real-time, sequence-based detection and

quantification of sequence variants and epigenetic methylation that uses pyrosequencing technology. The instrument can process 1−24 samples in 15 min. This platform, which is affordable to mid-sized laboratories, allows analysis of CpG methylation, SNPs, insertion/deletions, short tandem repeats (STRs), and variable gene copy number.

References

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., et al., 2008. The potential and challenges of nanopore sequencing. Nat. Biotechnol. 26, 1146−1153.

Couzinet., S., Yugueros, J., Barras, C., Visomblin, N., Francois, P., Lacroix, B., et al., 2005a. Evaluation of a high-density oligonucleotide array for characterization of grlA, grlB, gyrA and gyrB mutations in fluoroquinolone resistant *Staphylococcus aureus* isolates. J. Microbiol. Methods 60, 275−279.

Couzinet, S., Jay, C., Barras, C., Vachon, R., Vernet, G., Ninet, B., et al., 2005b. High-density DNA probe arrays for identification of staphylococci to the species level. J. Microbiol. Methods 61, 201−208.

Cyranoski, D., 2010. A primer on metagenomics. Nature 464, 22−24.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al., 2009. Real-time DNA sequencing from single polymerase molecules. Science 323, 133−138.

Gonzalez, R., Masquelier, B., Fleury, H., Lacroix, B., Troesch, A., Vernet, G., et al., 2004. Detection of human immunodeficiency virus type 1 antiretroviral resistance mutations by high-density DNA probe arrays. J. Clin. Microbiol. 42, 2907−2912.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., et al., 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol. 10, R32.

Harris, S.R., Feil, E.J., Holden, M.T., Quail, M.A., Nickerson, E.K., Chantratita, N., et al., 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science 327, 469−474.

Herring, C.D., Palsson., B.Ø., 2007. An evaluation of Comparative Genome Sequencing (CGS) by comparing two previously-sequenced bacterial genomes. BMC Genomics. 14, 274.

Holt, R.A., Jones., S.J., 2008. The new paradigm of flow cell sequencing. Genome Res. 18, 839−846.

Jiang, L., López-Barragán, M.J., Jiang, H., Mu, J., Gaur, D., Zhao, K., et al., 2010. Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. Proc. Natl. Acad. Sci. U.S.A. 107, 2224−2229.

Keijser, B.J., Zaura, E., Huse, S.M., van der Vossen, J.M., Schuren, F.H., Montijn, R.C., et al., 2008. Pyrosequencing analysis of the oral microflora of healthy adults. J. Dent. Res. 87, 1016−1020.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376−380.

Miller, M.B, Tang, Y.W., 2009. Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol. Rev. 4, 611−613.

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., et al., 2008. A new arenavirus in a cluster of fatal transplant-associated diseases. N. Engl. J. Med. 358, 991−998.

Pas, S.D., Tran, N., de Man, R.A., Burghoorn-Maas, C., Vernet, G., Niesters, H.G., 2008. Comparison of reverse hybridization, microarray, and sequence analysis for genotyping hepatitis B virus. J. Clin. Microbiol. 46, 1268−1273.

Pushkarev, D., Neff, N.F., Quake., S.R., 2009. Single-molecule sequencing of an individual human genome. Nat. Biotechnol. 27, 847−852.

Shendure, J., Porreca, GJ, Reppas, NB, Lin, X, McCutcheon, JP, Rosenbaum, AM, et al., 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728−1732.

Snyder, M., Du, J., Gerstein, M., 2010. Personal genome sequencing: current approaches and challenges. Genes Dev. 24, 423−431.

Sougakoff, W., Rodrigue, M., Truffot-Pernot, C., Renard, M., Durin, N., Szpytma, M., et al., 2004. Use of a high-density DNA probe array for detecting mutations involved in rifampicin resistance in *Mycobacterium tuberculosis*. Clin. Microbiol. Infect. 10, 289−294.

Steen, J.A., Harrison, P., Seemann, T., Wilkie, I., Harper, M., Adler, B., et al., 2010. Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. PLoS Pathog. 6, e1000750.

Tran, N., Berne, R., Chann, R., Gauthier, M., Martin, D., Armand, M.A., et al., 2006. European multicenter evaluation of high-density DNA probe arrays for detection of hepatitis B virus resistance mutations and identification of genotypes. J. Clin. Microbiol. 44, 2792−2800.

Troesch, A., Nguyen, H., Miyada, C.G., Desvarenne, S., Gingeras, T.R., Kaplan, P.M., et al., 1999. *Mycobacterium* species identification and rifampin resistance testing with high-density DNA probe arrays. J. Clin. Microbiol. 37, 49−55.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., et al., 2009. A core gut microbiome in obese and lean twins. Nat. Biotechnol. 27, 344−346.

van Leeuwen, W.B., Jay, C., Snijders, S., Durin, N., Lacroix, B., Verbrugh, H.A., et al., 2003. Multilocus sequence typing of *Staphylococcus aureus* with DNA array technology. J. Clin. Microbiol. 41, 3323−3326.

Via, M., Gignoux, C., Burchard, E.G., 2010. The 1000 Genomes Project: new opportunities for research and social challenges. Genome Med. 2, 3.

Voelkerding, K.V., Dames, S.A., Durtschi., J.D., 2009. Next-generation sequencing: from basic research to diagnostics. Clin. Chem. 55, 641−658.

Webb, K.M., Rosenthal., B.M., 2010. Deep resequencing of *Trichinella spiralis* reveals previously un-described single nucleotide polymorphisms and intra-isolate variation within the mitochondrial genome. Infect. Genet. Evol. 10, 304−310.

Wooley, J.C., Godzik, A., Friedberg, I., Miller, M.B., Tang, Y.W., 2010. Basic concepts of microarrays and potential applications in clinical microbiology. PLoS Comput. Biol. 26, e1000667.

Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E.C., Li, W.X., et al., 2010. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. Proc. Natl. Acad. Sci. U.S.A. 107, 1606−1611.

Yoder-Himes, D.R., Konstantinidis, K.T., Tiedje, J.M., 2010. Identification of potential therapeutic targets for *Burkholderia cenocepacia* by comparative transcriptomics. PLoS One. 5, 8724.