The journey from next-generation sequencing to personalized medicine?

Marion Vandeputte
(Illumina, UK)

Following the completion of the Human Genome Project in 2003, sequencing has become one of the most influential tools in biomedical research. Sequencing took off in earnest with the development of next-generation sequencing techniques in the early 2000s, making sequencing high throughput, faster, more affordable and commercially available to individual laboratories. With the improved understanding of the role of genetics in human disease, coupled with rapid advancement in sequencing technology, we are progressively unlocking the secrets of how our genes control the development of diseases. This has the potential to revolutionize medicine and healthcare, providing a significant step towards personalized medicine. How did we arrive here? What are the major achievements of sequencing technologies of the past two decades and how does it help us to piece the clues together towards personalized treatments and diagnosis?

What is next-generation sequencing?

Before next-generation sequencing, there was DNA sequencing. DNA sequencing is the ability to read the information contained in our genome, meaning to read the order of As, Cs, Gs and Ts that constitutes our DNA. It all started in 1977 (Figure 1) with the development of a method to determine the base sequences in nucleic acids by Frederik Sanger who won the Nobel Prize 3 years later for his contribution. Sanger sequencing dominated the research landscape until the early 21st century and led to exceptional achievements. It started as a low-efficient and high-cost method that was improved thanks to the work of many scientists. In the 1990s, the Human Genome Project was initiated which aimed to complete the map of the human genome. It was successfully completed in 2003, after the involvement of thousands of international scientists and 3bn dollars.

Since then, DNA sequencing has brought us a long way. The 21st century marked the era of next-generation sequencing, also known as high-throughput, or massively parallel, sequencing. The first next-generation sequencing methods were published in 2005 in two ground-breaking studies that described sequencing of a whole bacterial genome. Three years later, next-generation sequencing technologies were published for human genomes, opening the way for population-scale genome sequencing. Since its emergence, it has become one of the most influential tools in biomedical research and there have been exponential advances in our capacity to sequence a human genome. Various sequencing technologies have been developed by different commercial companies, making sequencing high throughput, affordable and highly sensitive in detecting genetic variants. In parallel, the development of new computational tools had rapidly peaked in 2009 and has since continued to expand. Advances in sequencing technology have enabled a genome to be sequenced within hours, at a fraction of the initial cost, resulting in widespread application for diagnosis and research.

Decoding the human genome of people with rare diseases and cancer

Sequencing the first human genome in 2003 was a major accomplishment of the Human Genome Project. However, discovering this sequence was only the beginning of a long road in understanding how the instructions coded in DNA lead to a functioning human being. The next stage of genomic research will improve our understanding of how alterations in our DNA are linked to certain diseases. This will open possibilities for new cures and prevention for genetic diseases, creating a new way of treating patients. After the completion of the Human Genome Project, the 1000 Genomes Project was launched in 2007 with the aim of forming a public reference database to catalogue human genetic variation across different human population groups. In 2013, scientists took a step further with the 100,000 Genomes Project, aimed to sequence 100,000 genomes from
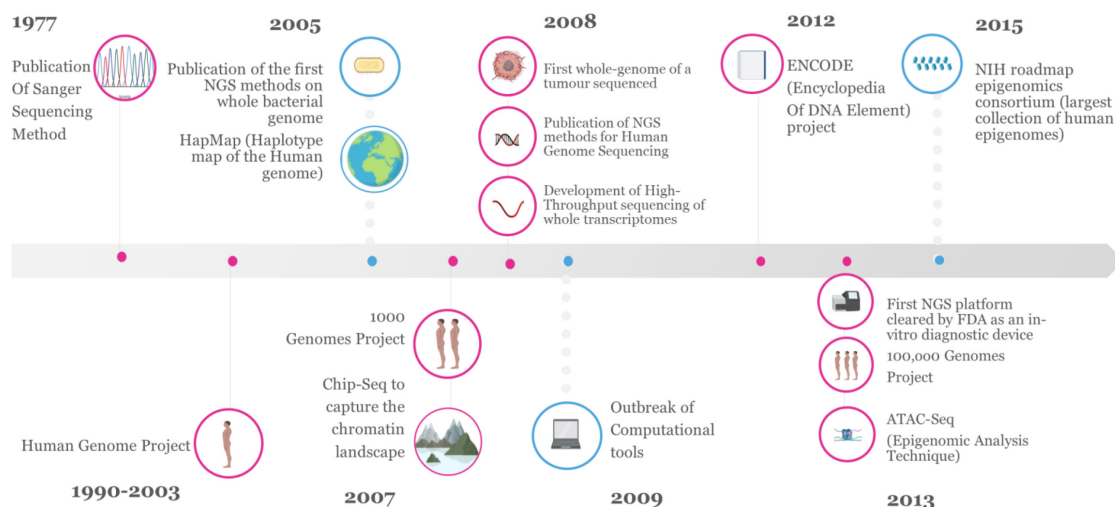
Figure 1. Milestones in next-generation sequencing over the last two decades.

around 85,000 National Health Service (NHS) patients affected by a rare disease, or cancer. This initiative aimed to help build up a database of genetic information to improve clinical care by using it for research and feeding it back into the NHS. Today, over 100,000 genomes have been sequenced from over 97,000 patients and their family members.

The era of large-scale sequencing projects

In addition to the 100,000 Genomes Project, the 21st century has marked the beginning of large-scale sequencing initiatives worldwide. These studies serve as a reference dataset for many genome-wide association studies in disease research. In 2005, the International HapMap Project set out to develop a haplotype map (HapMap) of the human genome that describes the common patterns of human DNA sequence variation. Other projects such as the British Cancer Genome Project in 2000 and the Cancer Genome Atlas in America in 2005 were initiated with the goal of understanding the role of genetic variations in cancer. In 2012, the Encyclopedia of DNA Elements (ENCODE) project was the first and largest international effort to characterize the functional side of the human genome. Currently, large international programmes are on-going and building up evidence to support the implementation of next-generation sequencing in personalized medicine.

What is personalized medicine?

What if we could be given personalized treatments based on personal effectiveness and minimal side effects? What if we could get a faster diagnosis based on our unique situation, or even have the ability to predict and prevent certain conditions from developing in the first place? This is the basis of personalized medicine.

At the core of this concept is the understanding that our DNA plays a role in our health, our susceptibility to develop disease and how we respond to treatments. Today, for most diseases, treatments are based on 'standard' care, involving applying the same treatment that is effective for most patients. We used to treat all people with a certain disease the same, despite all their differences. However, all patients do not respond in the same way to treatments, especially with cancer or rare diseases. This can lead to rounds of ineffective treatments, decreasing chances of success, waste of time, money and effort. Today, by combining and analysing information about our genome with other clinical information, we have become much better at measuring individual differences and using what makes us truly unique: our DNA. The concept of personalized medicine is not new, but using next-generation sequencing in medicine has the capacity to revolutionize the healthcare of an individual with a rare disease or cancer by offering personalized diagnosis, treatments, and prevention (Figure 2).

Genome, transcriptome and epigenome

Next-generation sequencing was first applied in genomics research mainly to study DNA. However, DNA is only one layer of the 'central dogma', the process by which the instructions in DNA are converted into a functional product (DNA codes for RNA, which codes for proteins). Understanding our genetics at the global level is essential for our comprehension of the development of disease in personalized medicine. The
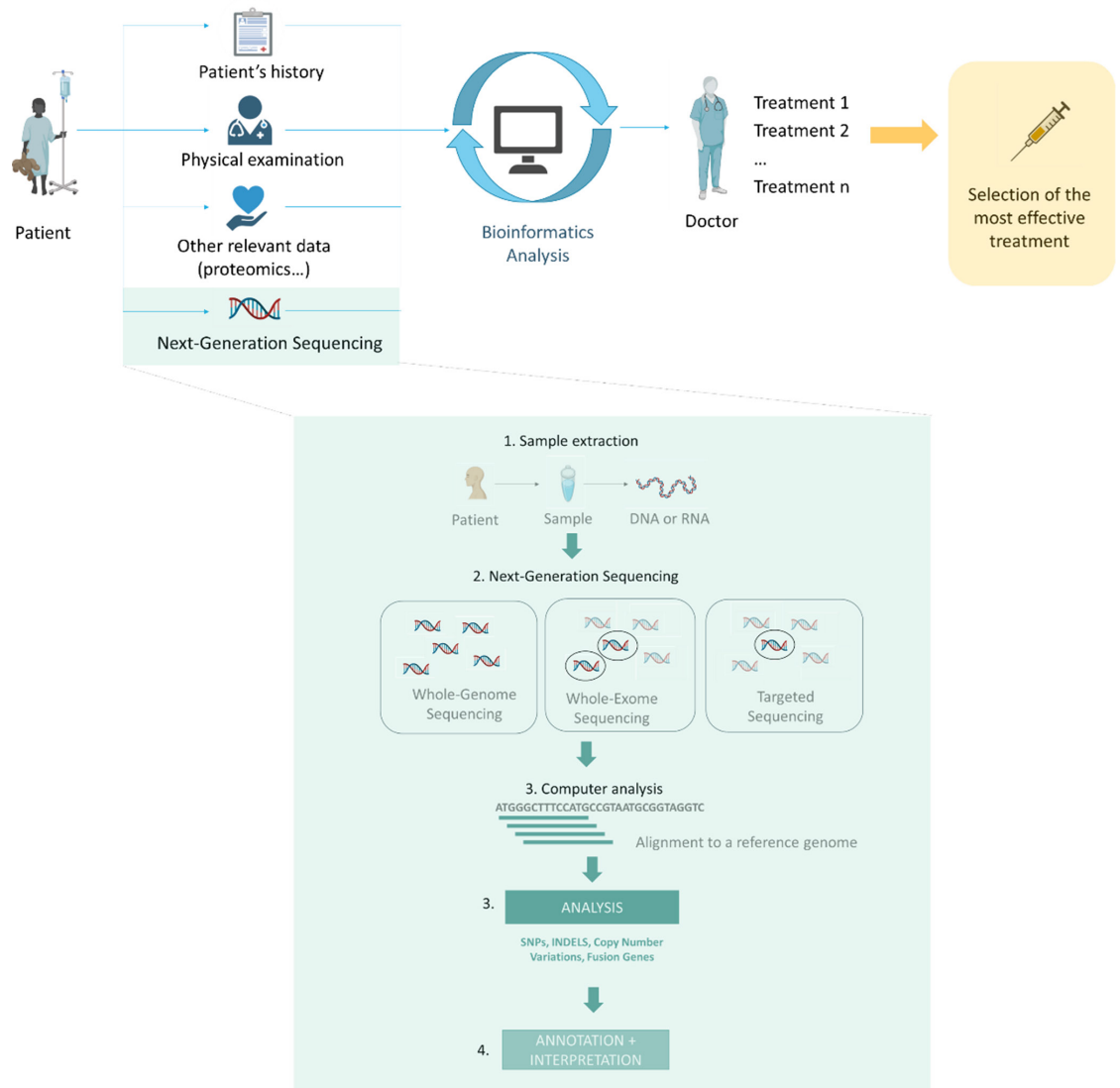
Figure 2. The road from next-generation sequencing to personalized medicine.

advances in next-generation sequencing over the last 20 years have been marked by some milestones in the analysis of our genome, transcriptome and epigenome, bringing us a step closer towards personalized medicine.

Genome: markers and drivers of some disease development, progression and treatment response can be detected at the DNA level. Alterations in genes such as single-nucleotide polymorphisms (what could be described as the equivalent to a 'typo' in the genome), INDELS (*in*sertion or *del*etion of bases in a sequence), copy number variations and fusion genes (when two individual genes form a hybrid gene), can be analysed and linked to cancers, rare and other diseases. For example, a deletion in the CTFR gene was associated with cases of cystic fibrosis, a rare disease. The fusion of the two individual genes TMPRSS2 and ERG was found in some cases of prostate cancers. Next-generation

sequencing methodologies applied to an entire genome is called 'whole-genome sequencing'. It enables the sequencing of both the coding and non-coding regions of the genome. Whole-exome sequencing is instead specifically designed to sequence the coding regions of the genomes. Targeted sequencing looks at specific sequences of interest of the genome, allowing a more selective analysis. These methods differ by the number of times a single region of the genome will be read and how fast and cheap the process will be. According to what we want to study and the applications, one of these three methods is selected (Figure 2).

Transcriptome: more recently, in 2008, transcriptome analysis or RNA sequencing has also emerged as an application of next-generation sequencing. A series of milestone publications reported the development of high-throughput sequencing of RNA across different

species. This also allows the discovery of additional modifications, similar to the genome modifications.

Epigenome: another emerging field of next-generation sequencing is epigenomics. It refers to the chemical modifications of DNA through internal and external factors, which can repress gene expression, leading to disease or treatment resistance. Over the last two decades, significant achievements in unlocking our epigenome have been made. In 2007, a method called Chip-Seq that captures the chromatic landscape was published, improving the understanding of how chromatin-bound proteins affect gene expression and, ultimately, cell behaviour. In 2013, a method called ATAC-seq was developed allowing an integrative epigenomic analysis. In 2015, the NIH Roadmap Epigenomics Consortium generated the largest collection of human epigenomes for primary cells and tissues. This work demonstrated how a cell's epigenome is complex and exquisitely arranged, revealing biologically relevant cell types for diverse human traits and providing a resource for interpreting the molecular basis of human disease.

The genome and cancer

Cancer is a genetic disease. It results from an accumulation of mutations in a particular tissue causing uncontrolled cell division and forming a tumour. The identification of specific DNA alterations allowed the development of targeted therapies, more effective and less toxic compared to standard chemotherapies. For example, two drugs (trastuzumab and imatinib) were the first ones to be approved and showed the potential of targeted therapy, in 1998 and 2001, respectively. This was followed by many molecules approved for different types of cancer. The next stage in cancer genetic has been to use next-generation sequencing to catalogue genetic mutations responsible for cancer. In 2008, the first whole genome of a tumour sample was sequenced, showing that cancer genome sequencing can identify disease-associated mutations. This marked the start of a sequencing revolution in cancer and led to several large-scale cancer sequencing projects being initiated worldwide. Sequencing DNA extracted from a tumour constructs a picture of the mutations responsible for the cancer. These genomic mutations can direct therapy based on likely tumour response.

The genome and rare diseases

At least 80% of rare diseases are genomic and caused by mutations in a single gene. They can include rare cancers such as childhood cancers and some other well-known conditions such as cystic fibrosis, Huntington's disease,

and neonatal diabetes mellitus. The completion of the Human Genome Project in 2003 had broad implications for how these diseases are diagnosed and managed. First, on the diagnosis side, rare diseases usually take a long time to be diagnosed, several physicians and several misdiagnoses. Next-generation sequencing offers the highest likelihood of rare disease diagnosis. Second, on the management side, genetic variants can respond differently to different treatments. For example, in the case of the neonatal diabetes mellitus, a genetic syndrome caused by severe β-cell dysfunction in the pancreas and presenting often before patients are 6 months old, five subtypes of genetic variants were identified. Each variant responds differently to the treatments, making next-generation sequencing a useful tool to give the most appropriate care to patients.

How unlocking the secrets of our genes can inform personalized medicine?

Next-generation sequencing is becoming routine aspects of medical care. It allows the diagnosis, treatment and prevention of rare genetic disorders and a variety of cancers. The choice of a particular treatment on the market can be made based on genetic data. One example is the work of Dr Stephen Kingsmore, who became the official title holder of the Guinness World Records® in 2016 (fastest genetic diagnosis) for his pioneering work in the field of new-born genome sequencing and analysis. In the USA, up to one-third of babies admitted to a neonatal intensive care have a genetic disease, and genetic illnesses cause more than 20% of infant deaths. Treatments are currently available for more than 500 genetic diseases; for about 70 of these, initiation of therapy in new-borns can help prevent disabilities and life-threatening illnesses. Genetic disorders can take decades to correctly diagnose. In this case, developing a rapid, complete sample-to-diagnosis pipeline has provided multiple life-changing diagnoses for sick babies. This allowed the use of genome data in newborns as a starting point for lifelong precision medicine.

In addition, the journey towards personalized medicine has opened some doors to the development of gene therapy. In recent years, there has been a rapid expansion of research into gene therapy using adeno-associated viruses (AAVs), a treatment consisting of administrating a neutralized virus as a vector to carry a functional gene. A few years ago, the Food and Drug Administration approved the first gene therapy to treat children with spinal muscular atrophy (SMA), a neurodegenerative disease caused by a deficiency in the survival motor neuron protein. However, despite knowing the genetic cause of a disease through next-generation sequencing and the progress of viral-vector gene therapies, we still are finding it difficult to treat genetically defined diseases. Also, in most developmental

disorders with a genetic cause, using AAVs as vectors has created an unprecedented opportunity for future treatment. Nonetheless, the timing of the intervention is crucial, making gene therapy challenging to implement.

The adoption of next-generation sequencing in clinical practice, what is next?

Next-generation sequencing has transformed the capacity, speed and cost of genomic sequencing and is slowly going from research to clinics. The next-generation sequencing platform, the Illumina MiSeq™Dx, was cleared in 2013 as an *in vitro* diagnostic device by the Food and Drug Administration. This was followed by the approvals of other assays for rare diseases detection and targeted therapies. Next-generation sequencing has also been involved in clinical trials including the development of early cancer detection assays or tests of new treatments for patients not responding to standard therapy. Recently, GRAIL, a healthcare company whose mission is to detect cancer early, announced the validation of its multi-cancer early detection test using the power of next-generation sequencing. Grail's cancer blood test tracks down over 50

types of early-stage disease and is currently being tested by the NHS.

Besides this great potential, the implementation of precision medicine in the real world has several limitations. Radical expansion of genomic medicine within clinical care requires new infrastructure, extended skills, education of the workforce and diligent engagement with the public. The Genomics England 100,000 Genomes Project was initiated in 2013 to establish the use of whole-genome sequencing in the NHS and drive change within NHS services to adopt this technology. Moreover, achieving diversity in genomic research has been slow with some under-represented populations. There is currently a global effort to apply genomic science and associated technologies to further the understanding of health and disease in diverse populations from regions such as the Middle East, Oceania and large proportions of Africa. Finally, to promote new discoveries and improvements, it is essential that information collected in a dedicated database can be accessible worldwide to rapidly process and interpret the growing volumes of data. This raises some concerns about privacy and data privacy, showing the need to continue to explore the ethical, legal and social issues raised by genomic research. ■

Further readings

- Milestones in Genomic Sequencing (2021), *Nature*, https://www.nature.com/immersive/d42859-020-00099-0/index.html [Accessed 26 November 2021]
- Ofman, J.J., Hall, M.P. and Aravanis, A.M. (2021) GRAIL and the quest for earlier multi-cancer detection. *Nature Portfolio*, https://www.nature.com/articles/d42473-020-00079-y
- Morganti, S., Tarantino, P., Ferraro, E., et al. (2019) Role of next-generation sequencing technologies in personalized medicine in *P5 eHealth: An Agenda for the Health Technologies of the Future*. Pravettoni G., Triberti S. (eds) Springer, Cham.
- Gonzalez-Garay, M.L. (2014). The road from next-generation sequencing to personalized medicine. *Per. Med*. **11**, 523–544. DOI: 10.2217/pme.14.34
- Lightbody, G., Haberland, V., Browne, F. et al. (2019) Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. **20**, 1795–1811. DOI: 10.1093/bib/bby051
- Nakagawa, H. and Fujita, M. (2018) Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci*. **109**, 513–522. DOI: 10.1111/cas.13505
- Morash, M., Mitchell, H., Beltran, H., et al. (2018) The role of next-generation sequencing in precision medicine: a review of outcomes in oncology. *J. Pers. Med*. **8**, 30. DOI: 10.3390/jpm8030030
- Dr. Stephen Kingsmore sets Guinness World Records title for fastest genetic diagnosis (2016) https://www.rchsd.org/about-us/newsroom/press-releases/dr-stephen-kingsmore-sets-guinness-world-records-title-for-fastest-genetic-diagnosis/ [Accessed 29 November 2021]

Marion Vandeputte is a scientist at Illumina. She completed a master's degree in biochemistry and biotechnology at Strasbourg University, France. During her final year, she spent 8 months in Boston, USA, working on DNA epigenetic modifications as part of her master's thesis. Upon graduating, she moved to Cambridge, UK, to work in the field of next-generation sequencing. Since 2016, she has been working at Illumina. She started working in R&D and is now working as an applications scientist trainer.