REVIEW ARTICLE

The impact of next-generation sequencing on the DNA methylation–based translational cancer research

JAVIER SOTO, CARLOS RODRIGUEZ-ANTOLIN, ELENA VALLESPÍN, JAVIER DE CASTRO CARPEÑO, and INMACULADA IBANEZ DE CACERES

MADRID, SPAIN

Epigenetics is currently in an exponential phase of growth, constituting one of the most promising fields in science, particularly in cancer research. Impaired epigenetic processes can lead to abnormal gene activity or inactivity, causing cellular disorders that are closely associated with tumor initiation and progression. Thus, there is a pivotal role of massive sequencing techniques for epigenetics, which aim to find novel biomarkers, factors of prognosis and prediction, and targets for achieving personalized treatments. We present a brief description of the evolution of next-generation sequencing technologies and its coupling with DNA methylation analysis techniques, highlighting its future in translational medicine and presenting significant findings in several malignancies. We also expose critical topics related to the implementation of these approaches, which is expected to be affordable for most research centers in the near future. (Translational Research 2015;■:1–18)

Abbreviations: BS = bisulfite sequencing; DMR = differentially methylated region; HGOSC = high-grade ovarian serous carcinoma; MBD = methyl-binding domain; 5-mC = 5-methylcytosine; MeCP2 = methyl CpG binding protein 2; NGS = next-generation sequencing; NSCLC = non-small cell lung cancer; PGM = personalized genome machine; qMSP = quantitative methylation–specific PCR; TCGA = the cancer genome atlas; TSG = tumor suppressor gene; WGBS = whole-genome bisulfite sequencing

INTRODUCTION

For more than 30 years, the method for DNA sequencing devised by Frederick Sanger[1] has been used as the gold standard by which to sequence the human genome (Fig 1), supporting the Human Genome Project, published in 2001 at a cost of US $3 billion ($1/nt).[2] For more than the last 6 years, technological improvements have made the development of novel techniques possible, offering scientific advantages and at the same reducing costs. Thanks to Sanger's contribution (considered "first-generation" sequencing technology), genomic research can now count on an arsenal of cheaper, more straightforward, and less time-consuming "second-" and "third-generation" technologies, known as next-generation sequencing (NGS). Their primary goal is sequencing the whole genome of a

From the Cancer Epigenetics Laboratory, INGEMM, La Paz University Hospital, Madrid, Spain; Biomarkers and Experimental Therapeutics in Cancer, IdiPAZ, Madrid, Spain; Structural and Functional Genomics, INGEMM-IdiPAZ-CIBERER, La Paz University Hospital, Madrid, Spain.

Submitted for publication September 2, 2015; revision submitted October 29, 2015; accepted for publication November 14, 2015.

Reprint requests: Inmaculada Ibanez de Caceres, Cancer Epigenetics Laboratory, INGEMM, Biomarkers and Experimental Therapeutics in Cancer IdiPAZ, Paseo de la Castellana 261, Madrid 28046, Spain; e-mail: inma.ibanezca@salud.madrid.org.

1931-5244/$ - see front matter

http://dx.doi.org/10.1016/j.trsl.2015.11.003

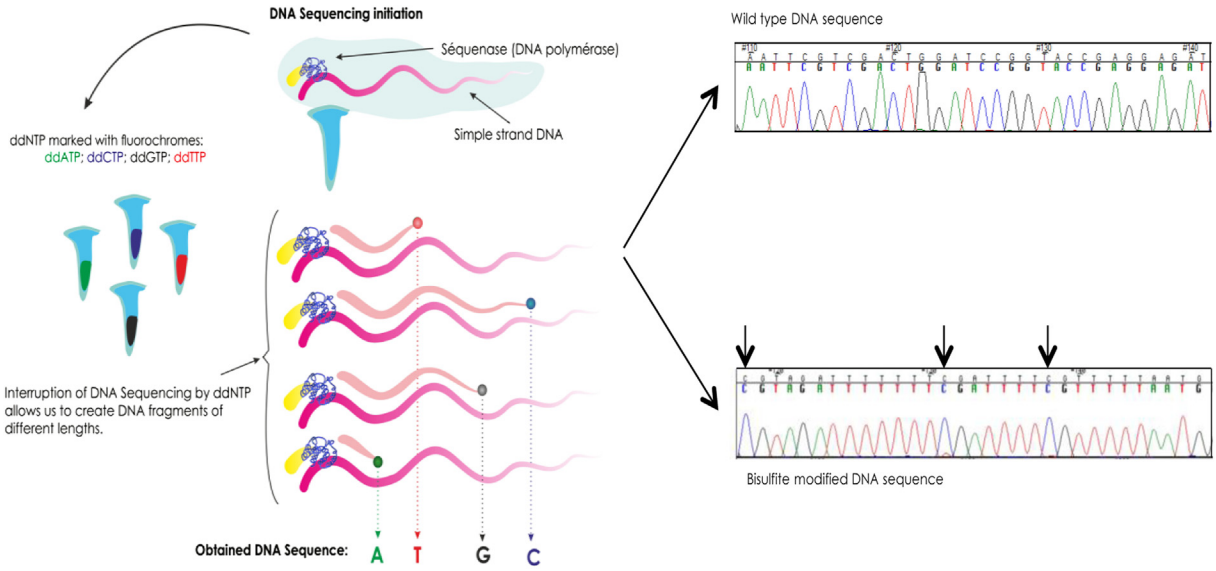

Fig 1. Sanger sequencing. Schema representative of the technology and examples of wild-type DNA and sodium bisulfite–converted DNA. Arrows in the sequence indicate the presence of methylated CpG sites. Abbreviations: ddNTPs, Dideoxynucleotide triphosphates; ddATP, dideoxyadenosine triphosphate; ddCTP, dideoxycytidine triphosphate; ddGTP, dideoxyguanosine triphosphate; ddTTP, dideoxythymidine triphosphate.

determined genetic source, enabling the detailed understanding of genomes, transcriptomes, and methylomes and accelerating the acquisition of knowledge in biological and biomedical research.

NGS development has facilitated the discovery of new transcripts, the identification of polymorphisms and their association with Mendelian and complex genetic diseases, and a deeper analysis and understanding of RNA structure, transcriptomes, and miRNomes. One of the primary goals of clinical management is to find new single-nucleotide variants and structural variants so that they can be associated with phenotypical differences, achieving genomic personalized profiles with medical purposes. As a first step, in 2004, the International Human Genome Sequencing Consortium published the first and only finished-grade human reference genome.[3] Another example is the cancer genome atlas (TCGA) database, a collaborative project that catalogs genomic, transcriptomic, epigenomic, and proteomic data for more than 30 tumor types—an excellent source of data for the identification of robust prognostic and predictive markers as well as therapeutic targets. It is a practical in silico method to validate research results in independent patient cohorts.

NGS embraced epigenetics quickly, with important profits in recent years, thus promoting the knowledge and positioning of epigenetics in science. Epigenetics is the study of heritable gene regulation that does not involve the DNA sequence. The 3 major types of epigenetic regulators are DNA methylation by covalent modification of cytosine-5′, post-translational modifications of histone tails, and microRNA (miRNA) gene expression regulation.[4] Their analysis is a key to understanding the heterogeneity and complexity of human beings, whose various cell types express their genes in hundreds of different ways (epigenome), despite having an identical genome. In 1974, Riggs, Holliday, and Pugh proposed cytosine DNA methylation as a primary factor in gene regulation and cellular proliferation[5,6]; nearly 50 years later, it remains the best-known epigenetic event in human beings and the most frequent in human cancer. DNA methylation analysis has become more widely used, and thanks to NGS it is possible to understand methylation status on a large scale and at a single-base resolution. High sensitivity, specificity, and scalability make NGS a powerful tool in the search for new genetic and epigenetic biomarkers in clinical cancer research, which is the theme of this study.

NGS PLATFORMS AND TEMPLATES

NGS is focused on the implementation of specific protocols regarding library template preparation, sequencing, detection, alignment of the read, and subsequent data analysis by specialized software provided by the various sequencing platforms. The need for robust methods that produce a representative source of nucleic acids from the genome under investigation should be of particular interest. Current methods typically involve randomly breaking genomic DNA (gDNA) into smaller sizes, from which either fragment templates or mate-pair templates are produced. A primary feature of