The Molecular Revolution in Cutaneous Biology: Era of Next-Generation Sequencing

Ofer Sarig¹ and Eli Sprecher¹,²

Like any true conceptual revolution, next-generation sequencing (NGS) has not only radically changed research and clinical practice, it has also modified scientific culture. With the possibility to investigate DNA contents of any organism and in any context, including in somatic disorders or in tissues carrying complex microbial populations, it initially seemed as if the genetic underpinning of any biological phenomenon could now be deciphered in an almost streamlined fashion. However, over the past recent years, we have once again come to understand that there is no such a thing as great opportunities without great challenges. The steadily expanding use of NGS and related applications is now facing biologists and physicians with novel technological obstacles, analytical hurdles and increasingly pressing ethical questions.

Abbreviations: NGS, next-generation sequencing; WES, whole-exome sequencing

WHY?

Have you ever resisted the temptation to have a quick look at the end of a book you just started and then take a swift glimpse at one of its chapters bearing an intriguing title? This is what next-generation sequencing (NGS), also known as massive parallel sequencing, is all about: the simultaneous reading of the human genome at multiple points in an attempt to make sense of an often quite complex story.

The first generation of sequencing techniques, developed in the 1970s by Sanger and Maxam and Gilbert, led to major advances in our understanding of human biology, culminating with the deciphering of the human genome in 2004 (International Human Genome Sequencing, 2004). Paradoxically, it is this lengthy and quite costly technological and scientific tour de force which led to the recognition of the need for alternative, more rapid, and cheaper sequencing solutions. It then took only a few years for the first next-generation sequencer (the 454 pyrosequencing-based Genome Sequencer; Roche Diagnostics, Indianapolis, IN) to appear on the market (Margulies et al., 2005). This was followed by the development of a number of more advanced technologies (Reuter et al., 2015), which can still widely vary in terms of efficiency of costs (Bodi et al., 2013; Liu et al., 2012).

HOW?

Despite the existence of various high-throughput sequencing approaches, these different technologies share common principles. Genomic DNA is initially randomly sheared and used to construct a working library. The library fragments are flanked by adaptors and barcodes, which are used for subsequent DNA capture and NGS. At this stage, libraries can be enriched for genomic regions of interest using hybridization to specific probes. For example, one of the most common ways to use NGS is to enrich genomic DNA for coding regions (exons), resulting in whole-exome sequencing (WES) (Teer and Mullikin, 2010). Alternatively, for phenotypes characterized by extensive genetic heterogeneity, specific genes can be enriched for using designated panels and sequenced with high coverage (Baudhuin et al., 2015). Captured DNA is then recovered and sequenced using one of the various NGS technologies, and the resulting sequencing data are then processed and analyzed.

Advances in the development of analytical tools have been less rapid than progress in sequencing technologies, mostly reflecting, as will be described, the need to adapt these tools to the transforming insights resulting from the very use of these new technologies. The analysis of NGS data usually involves four distinct and successive steps. First, the data are aligned to a reference genome, with possible technical errors identified (quality control). Second, the variants are called: namely, each sequence alteration is characterized and localized relative to the reference sequence. Third, the variants are filtered on the basis of a number of criteria, including their frequency in the general population, their conservation scores, and their predicted effects on protein translation and/or function or RNA splicing and/or degradation. Finally, additional steps can be taken to confirm the pathogenicity of a given variant, such as cross-referring the NGS data to expression data or mapping data and modeling the predicted effects of the mutation in vivo or in vitro, to obtain direct evidence for a pathogenic effect of the mutation using human tissues (Erzurumluoglu et al., 2015).

WHEN?

The initial successes of NGS in gene discovery for Mendelian disorders (Ng et al., 2010) established its reputation. Since then, WES has become a major tool not only for the
identification of the cause of single-gene disorders (Boycott et al., 2013), but also for identifying causal variants underlying genome-wide associations (Cirulli and Goldstein, 2010), for detecting somatic mutations in congenital (Campbell et al., 2015) and neoplastic (Nakagawa et al., 2015) disorders (with obvious clinical implications [Shen et al., 2015]), and for other DNA-based applications in fields as diverse as forensic medicine (Zascavage et al., 2013) and microbiology (Firth and Lipkin, 2013). Recent technical improvements, cost reduction, and the advent of more powerful analytical algorithms suggest the possibility that whole-genome sequencing (rather than WES) may soon become widely available (Royer-Bertrand and Rivolta, 2015). NGS-based approaches (e.g., RNA sequencing) have also recently replaced microarrays as a means of quantitatively ascertaining gene expression (Ozsolak and Milos, 2011) and are used for genome-wide mapping of protein-DNA interactions (Mundade et al., 2014) and of epigenetic modifications (Ku et al., 2011). Although in many cases the availability of NGS renders previous approaches to gene identification obsolete, NGS can actually be combined with those techniques. For example, linkage analysis of NGS data can assist in disease-causing gene identification (Hu et al., 2014).

NGS has become one of the foundations of precision medicine, and it is increasingly used today for clinical purposes, although many regulatory and ethical questions remain open (Evans et al., 2015). To diagnose inherited disorders associated with mutations in a large number of genes or to identify mutations in cancer tissues, two approaches are being used: (i) WES is performed, and the resulting data are scrutinized for mutations in a defined list of candidate genes, or (ii) gene-specific capturing panels are being used to selectively sequence genomic areas harboring genes of interest (Hall et al., 2014; Xue et al., 2015). More recently, WES has been successfully used to assess patients with diseases suspected to be caused by genetic mutations but without a definite diagnosis, with a 25% positive yield (Yang et al., 2013). Finally, NGS may also play an important role in prenatal diagnosis (Talkowski et al., 2012), including in the context of advanced techniques such as prenatal screening for genetic diseases using maternal plasma-derived DNA (Chitty and Lo, 2015).

**AND WHAT ABOUT THE SKIN?**

NGS has played a major role in advancing investigative dermatology over the past 5 years. The genetic bases of numerous monogenic skin diseases have been uncovered using NGS-based approaches, leading to novel insights into the molecular basis of pivotal aspects of skin biology such as cornification (Boyden et al., 2015; Duchatelet and Hovnanian, 2015; Wang et al., 2015), cell-cell adhesion (Samuelov et al., 2013), hair cycling (Higgins et al., 2014), and organization of the extracellular matrix (Fischer-Zirnsak et al., 2015), as well as illuminating the complex developmental and functional relationships between skin and extracutaneous tissues (Petrof et al., 2014; Riviere et al., 2012). RNA sequencing studies are also shedding new light on the pathogenesis of disorders as central to our discipline as atopic dermatitis (Cole et al., 2014; Suarez-Farinas et al., 2015) and psoriasis (Li et al., 2014; Zhang et al., 2015). Of particular interest is the fact that NGS has played a pivotal role in the deciphering of the cause of many mosaic skin disorders, including Proteus syndrome (Lindhurst et al., 2011), sebaceous nevus syndrome (Groesser et al., 2012), and epidermal and pigmented nevi (Hafner et al., 2006; Kinsler et al., 2014).

NGS is now increasingly considered for the routine diagnosis of simple and complex dermatological conditions (Takeichi et al., 2013), especially genetically heterogeneous skin disorders such as epidermolysis bullosa (Tenedini et al., 2015), ichthyoses (Suga et al., 2015), and connective tissue diseases (Hosen et al., 2015; South et al., 2015). Gene panels are not only available for the clinical diagnosis of genodermatoses but are also used as an adjunct to the molecular characterization of skin tumors to predict drug responses (Siroy et al., 2015).

**A WORD OF CAUTION**

NGS has been instrumental to the dramatic progress achieved in the past few years in our understanding of the genetic basis of human diseases. However, the conceptual changes in the way genetic changes are addressed using this new approach have generated a large number of scientific, medical, and bioethical challenges, most of which have yet to be met. The first and most immediate question that has emerged from the large amount of data accumulated over the past few years regards the interpretation of genetic alterations and their relationship to clinical phenotypes. Whether a mutation is pathogenic and affects the physiological function of an organism may be dependent upon the context in which it appears. The remarkably large load of deleterious mutations present in every genome may in fact be tolerated in the absence of environmental triggers or in individuals over-expressing a molecule capable of compensating for the lack of expression of another protein (Henn et al., 2015). Conversely, NGS-based studies have repeatedly shown unusual phenotypes to be due to inheritance of more than one mutation (Cullinan et al., 2011; Lai-Cheong and McGrath, 2011). In addition, the fact that some of these genetic alterations may be associated with late-onset phenotypes or with partial penetrance further adds to the sense of uncertainty regarding the way NGS data should be handled and reported. Thus, NGS is forcing us to re-evaluate the way we view and understand the meaning of mutations, which in turn may bear major implications for genetic counseling and future therapeutics (Cutting, 2014). Much effort is now invested by policymakers to generate regulatory structures to guide care providers in this context (for example, see the recent American College of Medical Genetics guidelines [Rehm, 2013]).

Second, the technology, despite continuous improvements, is still far from perfect. Some regions of the genome are poorly covered by standard NGS, and coverage can vary between different sequencing platforms. Sequencing errors are also common, and therefore NGS data, mostly WES and whole-genome sequencing, are still very often validated using Sanger sequencing (Frebourg, 2014). Much effort is therefore currently invested in trying to improve our computational and predictive capabilities (Bolouri, 2014) and to better understand, account for and eventually remedy to these errors (O’Rawe et al., 2015).
A third and major challenge in the implementation of NGS for clinical purposes concerns the storage of the huge and steadily growing amount of sequencing data generated in each NGS study. Hardware limitations, confidentiality issues, and data accessibility and manageability are all pressing but still unresolved issues (Johansen Taber et al., 2014).

Despite these many and complex hurdles, NGS has opened a new era for researchers and clinicians alike, which is likely to change existing paradigms and to transfigure our daily practice in the immediate future.

CONFLICT OF INTEREST
The authors state no conflict of interest.

ACKNOWLEDGMENTS
This work was supported by a generous donation of the Ram family.

TO CITE THIS ARTICLE

REFERENCES


Cutting GR. Annotating DNA variants is the next major goal for human genomics. Am J Hum Genet 2014;94:5–10.


Mundade R, Ozer HG, Wei H, Prabhu L, Lu T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation and data accessibility and manageability are all pressing but still unresolved issues (Johansen Taber et al., 2014).

Mundade R, Ozer HG, Wei H, Prabhu L, Lu T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation and manageability are all pressing but still unresolved issues (Johansen Taber et al., 2014).


O Sarig and E Sprecher Era of Next-Generation Sequencing


Parceau CL, Mardis ER, Jorde LB. A third and major challenge in the implementation of NGS for clinical purposes concerns the storage of the huge and steadily growing amount of sequencing data generated in each NGS study. Hardware limitations, confidentiality issues, and data accessibility and manageability are all pressing but still unresolved issues (Johansen Taber et al., 2014).


www.jidonline.org e81


