Published in final edited form as:
Expert Rev Vaccines. 2009 August ; 8(8): 963–967. doi:10.1586/erv.09.67.

Next-generation sequencing: a transformative tool for vaccinology

Neelam Dhiman, PhD,
Mayo Vaccine Research Group, Mayo Clinic, Rochester, MN 55905, USA and Program in Translational Immunovirology and Biodefense, Mayo Clinic, Rochester, MN 55905, USA, Tel.: +1 507 266 3065, Fax: +1 507 266 4716, dhiman.neelam@mayo.edu

David I Smith, and
Laboratory of Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA, Tel.: +1 507 266 0309, Fax: +1 507 266 4716, smith.david@mayo.edu

Gregory A Poland, MD
Director, Mayo Vaccine Research Group, Mayo Clinic, Guggenheim 611C, 200 First Street SW, Rochester, MN 55905, USA, Tel.: +1 507 284 4968, Fax: +1 507 266 4716, poland.gregory@mayo.edu and Program in Translational Immunovirology and Biodefense, Mayo Clinic, Rochester, MN 55905, USA

Next-generation sequencing (NGS) has revolutionized the fields of basic and clinical research in the 5 years since its introduction to the scientific community. Current NGS platforms on the market such as Roche 454 (Roche, CT, USA), Genome Analyzer (Illumina/Solexa, CA, USA), SOLiD™ (Applied Biosystems [ABI], CA, USA) and Heliscope™ (Helicos Biosciences, MA, USA) share one common attribute: highly parallel sequencing of DNA molecules to generate sequence outputs on an order of magnitude of 0.5 to greater than 40 Gb per sequencing run, thus enabling researchers to design experiments that were previously beyond the realms of feasibility and affordability [1-5]. The genotypic information obtained using these high-throughput systems allows the in-depth understanding of genotype–phenotype associations, critical to the development of the field of personalized medicine.

Next-generation sequencing technologies are based on repetitive polymerase-based nucleotide extensions or oligonucleotide ligation in a massively parallel fashion to increase efficiency and throughput. The Roche 454 sequencing technology combines the principles of emulsion PCR and pyrosequencing. The steps involve fragmentation of the template DNA, ligation to adaptors and clonal amplification of DNA using emulsion PCR. The emulsion beads are then deposited in picotiter plate wells containing smaller beads with sequencing enzyme and buffers required to perform iterative pyrosequencing – translating each nucleotide incorporation event into a well-specific pyrophosphate-tagged luminescence. The newer and robust ‘titanium’ chemistry

Correspondence to: Gregory A Poland.

Financial & competing interests disclosure
Copyright Mayo Clinic (2009). This work was supported by NIH grants AI 48793, AI 33144, N01-AI 40065 and 1 UL1 RR024150-01 from the National Center for Research Resources (NCRR), a component of the NIH and the NIH Roadmap for Medical Research. Its content is solely the responsibility of the authors and does not necessarily represent the official view of the NCRR or NIH. Gregory Poland is the chair of a Safety Evaluation Committee for novel non-rubella vaccines undergoing clinical studies by Merck Research Laboratories. David Smith has a nonfunded research collaboration with Applied Biosystems. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.
No writing assistance was utilized in the production of this manuscript.

can generate 1 × 10^6 sequence reads of longer read length (≥400 bp), yielding 500 million bp of sequence per run.

The Illumina/Solexa Genome Analyzer was the first ‘short read’ sequencing platform commercially available that involves sequencing-by-synthesis using reversible terminators. Fragmented ssDNA is hybridized to oligonucleotide anchors on a solid surface referred to as a ‘flow cell’. Solid-phase bridge amplification of DNA templates is conducted to generate amplified clusters. Massively parallel sequencing of cleaved products from amplified clusters is carried out using DNA polymerase and a set of four base-specific color-coded reversible terminators that result in growing oligonucleotide chains. This platform originally produced 35-bp reads to yield 1 Gb of sequence output per 2–3-day run. Subsequent upgrades on this platform have increased both the density of clusters and read lengths so that this machine can currently yield 4 Gb of sequence output in a 2–3-day run.

The third commercially available platform is the ABI SOLiD platform, which uses hybridization–ligation methodologies for massively parallel sequencing. The initial emulsion PCR step is the same as that in the Roche 454 platform, except that the beads are only 1 μm in size. The amplified product on the beads is then covalently linked to a glass surface, and sequencing is carried out using hybridization–ligation with an octamer interrogation probe consisting of two probe-specific, three degenerate and three promiscuous bases. Each nucleoide position is ascertained using a four-dye encoding schema and each position is interrogated twice to distinguish sequencing errors from single nucleotide polymorphisms (SNPs). When first available (early 2007), this platform had an output in short reads of 35 bp and produced 1–3 Gb of sequence data per 8-day run. The current upgrade on this system (now the SOLiD 3) is capable of a much higher density of beads and has an output of 20–40 Gb per 8–10-day sequence run.

Heliscope is the first single-molecule sequencing platform that does not require clonal amplification of DNA. The technology involves polyadenylation of fragmented 3′ DNA followed by hybridization to high-density (100 million strands/cm^2) poly(dT) oligonucleotides immobilized on a glass surface. Cy5-labelled dNTPs called ‘virtual terminators’ are added sequentially in ‘quads’ to achieve read lengths of 45–50 bp, with an estimated output of 1 Gb per day.

While NGS has been successfully applied to numerous areas of basic and clinical research, genomic research has particularly advanced as a result of the efficiency and high-throughput sequencing of entire genomes of varying complexity, ranging from microbial to human [6-13]. In addition to whole-genome sequencing, NGS allows focused fine mapping of narrow genomic intervals of interest by ‘targeted resequencing’ to identify structural variations such as SNPs, insertions and deletions (indels), duplications, mutations, chromosomal rearrangements and copy-number variations (CNVs). ‘Metagenomics’ is another upcoming field tremendously accelerated by NGS, whereby the technology allows qualitative (identifying known and previously uncharacterized species by *de novo* assembly of overlapping sequence reads) and quantitative (determining relative abundance of sequence reads) analyses of biologically diverse microbes in complex clinical and environmental samples [14-16]. Another application of NGS, ChIP-seq, has been useful in generating genome-wide protein-binding maps, which can be used to confirm previously identified genomic methylation sites and to delineate novel ‘protein-binding genomic motifs’ [17,18].

Next-generation sequencing can also be used to characterize the transcriptional output of cells of interest. This can be done with either a 3′-tagging procedure that generates short (21 bp) reads from each cDNA molecule interrogated (in a manner analogous to serial analysis of gene expression), or by fragmenting the RNA and completely sequencing it (RNA-seq). One of the advantages of RNA-seq over the 3′ tagging procedure is that the resulting qualitative and quantitative output data can simultaneously provide insight into gene-expression profiling, genomic annotation such as 5′/3′ and exon/intron boundaries, and structural sequence variation such as SNPs, indels, splice variants and mutations [6,8,12,13]. Owing to its versatility, efficiency and affordability, RNA-seq has rapidly evolved into a promising technology to gather comprehensive transcriptional level information on a variety of organisms and biological processes [6,8,9,12,13].

Applications for NGS in the medical field are rapidly emerging. NGS has recently been applied to monitor emerging drug resistance to HIV-1 via the detection and quantification of HIV quasispecies [19]. ABI SOLiD has been used to investigate the transcriptional complexity of key signaling pathways controlling cell proliferation and differentiation in animal models [8]. Roche 454 deep-sequencing transcriptome profiling has been used to identify ‘molecular signatures’ for breast carcinomas [20]. The application of whole-transcriptome digital gene-expression profiling at single-cell resolution demonstrates the power of this technology, allowing unprecedented accuracy and depth when the source sample is limited [21]. In this article, we focus on the potential of this cutting-edge technology to transform the field of vaccinology.

Heritability studies on twins, including those carried out by our group, have demonstrated significant interindividual variations in antibody response to viral vaccines attributable to the distinct genetic constitution of these individuals [22]. Both HLA and non-HLA gene families have been identified as contributory candidates towards the overall variation in immune phenotypes [23]. Based on our comprehensive data on genetic associations between candidate immune-response gene families (HLA, cytokines, cytokine receptors, Toll-like receptors, viral entry and recognition receptors, antiviral proteins, and innate and vitamin receptors) and immune phenotypes using various viral vaccine models, we have developed an ‘immune-response network theory’ [23]. The major thrust of the theory is that the overall response to vaccines is the cumulative result of host genetic variations and multigenic interactions that result in complex, yet unique, signatures referred to as an ‘immunogenetic profile’, which can be used to predict the immune response [23]. Currently, our model integrates information from various techniques such as immunoassays, candidate and whole-genome genotyping, microarrays, and sequencing. However, the approach is limited to what we know about the contributing components of the immune network. Keeping in mind the complexity of the immune system, an ‘all-in-one’ integrated, high-throughput but inexpensive tool is required to understand, synthesize and decipher the cumulative effects of individual genetic variations and their multigenic interactions to predict immune response to vaccination. In addition, since many of these genes are expressed at very low-abundance levels within cells, their expression is not detectable using classical microarrays. Thus, NGS offers greater sensitivity, breadth and accuracy in detecting low-abundance transcripts important to vaccine immunology.

As a result, NGS is a technology that has transformative potential to the field of vaccinology by virtue of its multifaceted applications. For vaccinology, a critically important application of transcriptome analysis has been comprehensive gene-expression profiling of populations to identify and quantitate differences in mRNA species, and to characterize these differences at the genetic level. The ‘sequence census’ application allows the use of short reads or ‘tags’ to identify the presence and abundance of the gene(s) of interest simultaneously at a massively high-throughput and genome-wide scale, essentially giving a digital overview of a transcribed product in response to vaccination. Furthermore, it is also possible that differences in immune response may just be due to alterations in the expression of the coding genes, but also to the noncoding portion of the genome. RNA-seq can be used to examine this previously ‘unseen’ portion of the genome. In addition, immune-response variations could be due to the abundance of specific transcript isoforms, and RNA-seq can detect these isoforms. Comparative analysis of these digital readouts between a ‘good’ and ‘poor’ responder to a vaccine could help construct signature immune-response profiles, which may be used to predict an immune response or a possible adverse response to a vaccine, as well as providing a starting place for understanding genotype–phenotype associations.

Another area of investigation for vaccinologists is an in-depth understanding of structural variations in genes that regulate differential mRNA expression in response to vaccines. These variations can be in the form of SNPs, indels, CNVs, repeats and other genome rearrangements. NGS can be applied to whole genome, exome or, more recently, targeted sequencing of specific candidate genes, or fine-mapping genetic intervals of interest in population-wide gene-association studies. NGS has been successfully applied to deep profiling of highly polymorphic major histocompatibility complex genes in animal models [11]. This application validates the potential of this technology in understanding the role of variations in multiallelic genes, such as HLA, in vaccine response on the population level.

Another area of investigation is analyzing the transcriptome for profiling, and discovery of known and novel RNA molecules that are not translated into downstream protein products, such as tRNA, rRNA, small nuclear RNA, microRNA (miRNA) and small interfering RNA, often collectively referred to as small noncoding RNA (ncRNA). Amongst these, miRNAs have surfaced as important post-transcriptional regulators in humans [24]; however, their role in immune regulation of vaccines is unexplored. RNA-seq not only allows deep and high-throughput interrogation of known miRNAs, but also allows novel miRNA discovery, which is a major limitation of the traditional microarray approach. These novel ncRNA genes may be critical to determining gene expression and therefore may play an important role in regulating gene products. There have also been a number of considerably larger ncRNAs (such as metastasis-associated lung adenocarcinoma transcript-1 and trophoblast-derived noncoding RNA) that have recently been identified, which can be analyzed with RNA-seq. The high single-base resolution of RNA-seq allows greater precision in isoform distinction and allelic expression at the same time, thereby allowing a precise understanding of both structural and functional variants that may be involved in vaccine immune response.

Copy-number variation has recently been defined as another major type of genetic variation covering approximately 12% of the total genome, and a major contributor to genetic diversity in humans [25]. CNV-Seq offers a high-throughput sequencing method that uses the number of short-reads for CNV analysis with high resolution [26]. NGS also provides an efficient tool to combine and correlate information about genetic diversity and analysis to dissect genetic complexity in regulating immune response to vaccines. In addition to genetic variations, several epigenetic regulatory mechanisms, such as histone modifications (methylation, acetylation, phosphorylation and ADP-ribosylation), nucleosome positioning, and regulatory protein binding are important and reversible variations that can influence the activation or silencing of immune-response genes, and hence modulate host–pathogen and host–antigen interactions [27]. A new field of patho–epigenetics has emerged that involves elucidating the epigenetic consequences of host–pathogen interactions. Histones constitute key chromatin proteins that play an important role in DNA packaging, replication and gene expression, as well as exhibiting frequent post-translational modifications. DNA methylation is the most common epigenetic modification, and DNA methylation markers could be of great diagnostic and therapeutic potential in infectious disease and cancer treatments. The epigenetic regulatory mechanisms target viral and retroviral genomes, and hence may have an important phenotypic impact in response to viral vaccines. NGS can catalog genome-wide DNA epigenetic variations in a cost-effective and comprehensive manner, compared with currently used immunoprecipitation techniques. Our immune-response network theory suggests that phenotype response to a vaccine or disease is the summative result of intra-and epigenetic diversity, thus warranting their simultaneous evaluation. If investigators can profile many or all of the genetic modifications that may be responsible for altered chromatin structure, and therefore responsible for variations in immune response to a vaccine into specific ‘epigenetic barcodes’, then these barcodes could conceivably be used to predetermine and predict vaccine response. Barcoding epigenetic modifications and their upstream pathways also offers promising targets for both drug and vaccine development by identifying key interactive elements of the genome beyond the sequence variations that may affect the function of a specific gene or pathway in response to drug administration, or vaccination.

Next-generation sequencing has taken medicine and biology into a new era of ‘personalized genomics’ where individual genome-wide characterization, including RNA transcript profiling, CNV differences at the genomic level, chromatin structure, and DNA methylation patterns, is feasible. In our view, this cutting-edge technology has a transformative potential in the field of vaccinology. We have previously reviewed the developing field of ‘personalized vaccinology’ and why this is the future standard of care, as well as how ‘predictive vaccinology’ will guide this field [23,28]. Significant interindividual variations exist in vaccine-induced immune responses, and this calls for individualized vaccination approaches that can be tailored and guided by the genetic uniqueness of each individual. A vaccinologist’s dream is ‘predictive vaccine-response profiling’ on a chip that will answer important clinical questions regarding disease susceptibility, vaccine efficacy, number of doses needed, potency of dose, route of administration and probable adverse events in one laboratory experiment at an affordable cost. A comprehensive vaccine-response profile that can be developed using NGS offers the promise of tailoring vaccination approaches to individual requirements. In addition, this information may also direct future vaccine development in several ways. Identification of key HLA alleles, haplotypes and supertypes associated with immune response can guide isolation of the corresponding immunogenic peptides that could be used for peptide-based vaccine development. Identification of specific genetic variants and/or pathways that relate to ‘poor’ response could be boosted by the incorporation of cytokines/adjuvants or other technology within vaccine formulations.

Some examples of the application of NGS in vaccinomics have begun to appear in the literature. Reif *et al.* conducted a highly parallel genotyping and subsequent validation study using two independent clinical trials in healthy vaccinia virus-naive individuals to determine associations between host genetics and adverse events (AEs; fever, lymphadenopathy or generalized rash) in response to smallpox vaccination [29]. The authors identified and validated three SNPs, a single SNP in the 5,10-methylenetetrahydrofolate reductase gene and two SNPs belonging to an immunological transcription factor interferon regulatory factor-1 gene, associated with AEs to smallpox vaccine. The study demonstrated how common genetic variants can be related to a complex clinical phenotype, and prescreening for these markers before smallpox vaccine administration may predict AEs. This can help identify ‘high-risk individuals’ for vaccine-associated AEs and guide important decisions about vaccine administration, which can be of substantial public-health importance. Our own studies have identified genetic variations in HLA and non-HLA genes associated with non-or hyperimmune phenotypes after measles, mumps, rubella and smallpox vaccination [23,30], which could be used as genetic blueprints to guide ‘personalized vaccination regimens’.

Among the major challenges that face NGS are cost, data storage and analysis. This technology generates an unprecedented volume of data, which poses challenges in terms of processing, storage, management, data mining and analysis, and requires a resource-intensive data pipeline and infrastructure of trained statisticians and bioinformaticians to draw meaningful conclusions from the data. Current analysis tools available to the research community are complex and time consuming and there are trade-offs between analysis speed and accuracy. However, the technology is rapidly moving forward and an international effort has been initiated to sequence 1000 human genomes in order to characterize and catalog all human genetic variation [101].

Additional funding has also been allocated to the Revolutionary Genome Sequencing Technologies program with the overall aim of reducing the costs of human-genome sequencing to US$1000 or less [102].

New single-molecule sequencing technologies on the horizon are being developed by several companies. These technologies are expected to streamline sample preparation and to further reduce sequencing time and cost. Nanopore-based sequencing technology is also in the pipeline and holds the promise of bringing the sequencing costs per genome to significantly less than US$1000. This technology not only generates very long sequence reads but can also differentiate between five different bases, A, C, G, T and 5′-methyl-C (hence analyzing both the sequence of the DNA fragments and their methylation status). The integration of NGS technologies into clinical practice may take years; however, the tremendous potential and applications of NGS in transforming medicine and biology are evident.

In conclusion, NGS offers a giant leap forward for the future of vaccinology. With advances in the current technology and new single-molecule sequencing technologies on the horizon, it will become cost and time efficient. Advances in the analysis of such data may well revolutionize how we practice vaccinology. This technology is likely to take vaccinology into an era where a prevaccination genome-wide screen may be used to generate a ‘personalized vaccination regimen’, which will guide the physician in decisions about the type, timing, dosage and likely AEs associated with each vaccine. The public-health impact of an individualized vaccination approach will be tremendous, as it will reduce vaccine failure and adverse events rates. In addition, by enhancing our understanding of vaccine-immunogenetic regulation, new insights into vaccine development and adjuvant strategies are likely to result. Additional research is required to explore and integrate the enormous potential of NGS in directing the field of predictive and personalized vaccination, but the benefits of individual and population-level health are likely to be enormous.

References

1. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem 2009;55(4):641–658. [PubMed: 19246620]
2. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. Genomics 2008;92:255–264. [PubMed: 18703132]
3. Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 2009;10:R32. [PubMed: 19327155]
4. Lo YM, Chiu RW. Next-generation sequencing of plasma/serum DNA: an emerging research and molecular diagnostic tool. Clin Chem 2009;55:607–608. [PubMed: 19233905]
5. Ansorge WJ. Next-generation DNA sequencing techniques. Nat Biotechnol 2009;25:195–203.
6. Sultan M, Schulz MH, Richard H, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 2008;321:956–960. [PubMed: 18599741]
7. Gorringe KL, Campbell IG. Large-scale genomic analysis of ovarian carcinomas. Mol Oncol 2009;3:157–164. [PubMed: 19383377]
8. Cloonan N, Forrest AR, Kolle G, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 2008;5:613–619. [PubMed: 18516046]
9. Rozera G, Abbate I, Bruselles A, et al. Massively parallel pyrosequencing highlights minority variants in the HIV-1 *env* quasispecies deriving from lymphomonocyte sub-populations. Retrovirology 2009;6:15. [PubMed: 19216757]
10. Clark AG. Genome sequences from extinct relatives. Cell 2008;134:388–389. [PubMed: 18692462]
11. Wegner KM. Massive parallel MHC genotyping: titanium that shines. Mol Ecol 2009;18:1818–1820. [PubMed: 19317846]
12. Nagalakshmi U, Wang Z, Waern K, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 2008;320:1344–1349. [PubMed: 18451266]

13. Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63. [PubMed: 19015660]

14. Huber JA, Mark Welch DB, Morrison HG, et al. Microbial population structures in the deep marine biosphere. Science 2007;318:97–100. [PubMed: 17916733]

15. Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS ONE 2008;3:e2527. [PubMed: 18575584]

16. Keijser BJ, Zaura E, Huse SM, et al. Pyrosequencing analysis of the oral microflora of healthy adults. J Dent Res 2008;87:1016–1020. [PubMed: 18946007]

17. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein–DNA interactions. Science 2007;316:1497–1502. [PubMed: 17540862]

18. Ball MP, Li JB, Gao Y, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 2009;27:361–368. [PubMed: 19329998]

19. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res 2007;17:1195–1201. [PubMed: 17600086]

20. Guffanti A, Iacono M, Pelucchi P, et al. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC Genomics 2009;10:163. [PubMed: 19379481]

21. Tang F, Barbacioru C, Wang Y, et al. mRNA-seq whole-transcriptome analysis of a single cell. Nat Methods 2009;6:377–382. [PubMed: 19349980]

22. Jacobson RM, Ovsyannikova IG, Targonski PV, Poland GA. Studies of twins in vaccinology. Vaccine 2007;25:3160–3164. [PubMed: 17284336]

23. Poland GA, Ovsyannikova IG, Jacobson RM, Smith DI. Heterogeneity in vaccine immune response: the role of immunogenetics and the emerging field of vaccinomics. Clin Pharmacol Ther 2007;82:653–664. [PubMed: 17971814]

24. Berezikov E, Cuppen E, Plasterk RH. Approaches to microRNA discovery. Nat Genet 2006;38 (Suppl):S2–S7. [PubMed: 16736019]

25. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature 2006;444(7118):444–454. [PubMed: 17122850]

26. Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 2009;10:80. [PubMed: 19267900]

27. Minarovits J. Microbe-induced epigenetic alterations in host cells: the coming era of patho-epigenetics of microbial infections. A review. Acta Microbiol Immunol Hung 2009;56:1–19. [PubMed: 19388554]

28. Poland G. Personalized vaccines: the emerging field of vaccinomics. Exp Opin Biol Ther 2008;8:1659–1667.

29. Reif DM, McKinney BA, Motsinger AA, et al. Genetic basis for adverse events after smallpox vaccination. J Infect Dis 2008;198:1–7. [PubMed: 18544010]

30. Poland GA, Ovsyannikova IG, Jacobson RM. Vaccine immunogenetics: bedside to bench to population. Vaccine 2008;26:6183–6188. [PubMed: 18598732]

Websites

101. 1000 genomes. A deep catalog of human genetic variation. www.1000genomes.org

102. National Human Genome Research Institute. NHGRI seeks DNA sequencing technologies fit for routine laboratory and medical use. New grants drive development of rapid, cost-effective sequencing technologies. www.genome.gov/27527585

Biographies

Neelam Dhiman

David I Smith

Gregory A Poland