Rev. Sci. Tech. Off. Int. Epiz., 2016, **35** (1), 53-65

High-throughput sequencing and vaccine design

F. Luciani

Systems Medicine, School of Medical Sciences, University of New South Wales, New South Wales, Australia
E-mail: Luciani@unsw.edu.au

Summary

Next-generation sequencing (NGS) technologies have reshaped genome research. The resulting increase in sequencing depth and resolution has led to an unprecedented level of genomic detail and thus an increasing awareness of the complexity of animal, human and pathogen genomes. This has resulted in new approaches to vaccine research. On the one hand, the increase in genome complexity challenges our ability to study and understand pathogen biology and pathogen–host interactions. On the other hand, the increase in genomic data also provides key information for developing and designing improved vaccines against pathogens that were previously extremely difficult to deal with, such as rapidly mutating RNA viruses or bacteria that have complex interactions with the host immune system. This review describes how the broad application of NGS technologies to genome research is affecting vaccine research. It focuses on implications for the field of viral genomics, and includes recent animal and human studies.

Keywords

Immune response – Next-generation sequencing – Quasispecies – Vaccine design – Viral evolution – Viral genome.

Introduction

Immunisation is one of the most effective and long-lasting ways of preventing infectious diseases, as well as some cancers associated with infection. In humans, vaccines against diphtheria, tetanus, poliomyelitis, influenza, hepatitis B, measles, mumps and rubella, as well as pneumococcal, meningococcal and *Haemophilus influenzae* type B infections, have reduced the incidence of these infectious diseases and their associated mortality by more than 97–99% (1). In animals, recent vaccines against viruses (e.g. foot and mouth disease virus [FMDV], porcine reproductive and respiratory syndrome virus, porcine circovirus, Newcastle disease virus, avian influenza virus, rabies virus, canine parvovirus, canine coronavirus) and bacterial infections (*Yersinia ruckeri* in fish, *Mycoplasma synoviae* in poultry, *Brucella abortus* in cattle) have significantly improved animal health, leading to economic benefits (2). However, there are still many infectious diseases for which no vaccine exists. A key constraint in the development of vaccines to protect against many prevalent human and animal pathogens is variability in pathogen genomes across epidemics or even within a single infection episode in a single host. There is also a lack of understanding about how these microorganisms evolve to escape host immune responses and insufficient insight into the characteristics of protective immunity.

Since 2005, the development of high-throughput sequencing (so-called next-generation sequencing [NGS]) has led to a massive increase in the capacity to sequence genomes at a relatively low cost and over a short time period (3, 4). Current NGS technologies encompassing the well-established 'second-generation' sequencing technologies such as the Illumina and 454 Roche platforms set a new standard in sequencing (previously represented by 'first-generation' Sanger sequencing), and third-generation technologies (such as PacBio RS from Pacific Biosciences and MinIon from Oxford Nanopore Technologies) are based on single molecule sequencing (5). The latter can be used to sequence genome segments up to 30 kilobases (kb) in length and thus provide an unprecedented capability to analyse complex genomic rearrangements.

This review considers the current NGS applications relevant to vaccine research and provides a brief critical analysis of how these technologies are influencing current vaccine research. It draws on examples from both human and animal research and provides examples of the use of NGS in genomics, transcriptomics and epigenetics.

doi: http://dx.doi.org/10.20506/rst.35.1.2417

RNA viruses

Although RNA viruses are the most common pathogens of humans and animals, no effective vaccines are available. These viruses include high-profile epidemic pathogens, such as human immunodeficiency virus 1 (HIV-1), hepatitis C virus (HCV), vector-borne viruses (e.g. dengue virus) and FMDV. In addition, emergent RNA viral pathogens are a major concern. These are typified by swine flu (influenza A H1N1), which arose via genetic recombination and then crossed species barriers to become pandemic in 2009 (6). A major change in the analysis of RNA viruses and other rapidly mutating pathogens has been the application of NGS as a deep-sequencing technology, which has allowed the detection of 'quasispecies', i.e. a population of related genomic variants co-evolving at the same time and within the same host but at different relative frequencies, ranging from the very common to the very rare (e.g. 1 in 10 million) (7). NGS provides a suitable method of analysing viruses with rapid evolutionary rates, but the large amount of data generated means that there is also a need for complex bioinformatics analyses (8).

Bacteria

For several bacterial pathogens of both humans and animals, no vaccine is yet available. Current therapies have limited success rates, along with a risk of developing drug resistance. NGS is rapidly replacing older technologies that focused on specific regions of complex bacteria genomes (such as variable-number tandem repeat or multi-locus variable-number tandem repeat analysis, multi-locus sequencing and single nucleotide polymorphism [SNP] analysis) and cannot detect new mutations in new clades. NGS technologies can be used to sequence complete bacterial genomes across a large number of samples and with limited cost. Hence, for the first time, it has become possible to investigate the heterogeneity and complexity of bacterial genomes on a large scale. NGS can also be used for deep sequencing to detect rare variants of bacterial genomes that may represent intermediate forms in the evolution of drug resistance (see [9] for a recent review). For instance, NGS has been used to sequence bacterial genomes from a number of highly virulent human and animal pathogens, such as *Brucella* spp., which are classified as category B potential biological warfare agents (10).

Reverse vaccinology

In less than a century, vaccines developed using Pasteur's original rules of 'isolate, inactivate and inject the microorganism' have eliminated some of the most devastating infectious diseases. Most existing vaccines were developed using conventional methods, i.e. by attenuation of the pathogen via serial passage *in vitro* to generate live attenuated strains which retain immunogenicity but are no longer pathogenic (such as the case for the most successful vaccine against smallpox) or by identifying protective single antigens for use in non-living subunit vaccines (11, 12). In the latter case, time-consuming and costly biochemical methods have traditionally been used to purify antigenic factors from organisms grown in culture, resulting in immunogenicity testing of only a few antigens, and limited consideration of the available evidence for naturally occurring protection against these antigens (13).

Reverse vaccinology is a more recent research discipline that uses pathogen genome analysis to identify or predict antigenic domains exposed on the pathogen's surface, and tests these experimentally (Fig. 1). Despite its relatively limited use, reverse vaccinology allows immunogenic regions to be identified from coding sequences of any pathogen, including those that cannot be cultured *in vitro* (14). The advent of new technologies has enabled vaccine research to move forward and develop vaccines against pathogens for which previous methods have failed.

A breakthrough in this field was the high-throughput sequencing of whole bacterial genomes, which has opened a new avenue into discovering highly complex genomic rearrangements and thus improved our understanding of bacterial evolution. One of the most successful applications was the whole-genome sequencing of *H. influenzae* (15). These data allowed researchers to move beyond Pasteur's approach to develop a more precise methodology involving the analysis of pathogen genomes to critically inform vaccine design. Genome-wide sequencing has been used to detect potential antigenic sites in serogroup B *Neisseria meningitides* (16), which is responsible for 50% of cases of meningococcal sepsis and meningitis worldwide This bacterium had been refractory to vaccine development because its capsular polysaccharide (polysialic acid) is non-immunogenic: it is expressed by a number of human tissues and, hence, can be considered a self-antigen to which the human immune system is tolerant. Using the reverse vaccinology approach, 600 putative antigens were discovered, of which 29 were shown to induce antibodies that kill the bacterium *in vitro* via complement-mediated mechanisms. Following this seminal application, many other pathogens for which previous technology has failed are currently being targeted with the reverse vaccinology approach, such as the group A *Streptococcus*, *Streptococcus pyogenes*, *Staphylococcus aureus*, *Streptococcus pneumoniae* and *Chlamydia pneumoniae* (17).

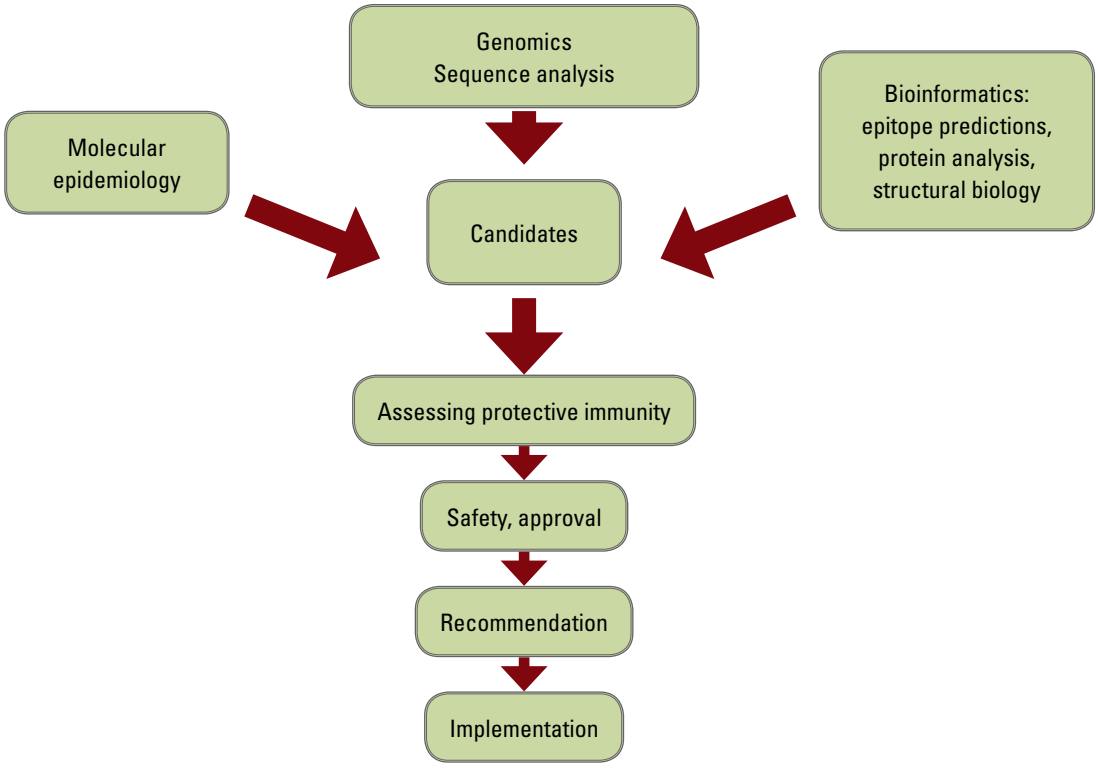


Fig. 1
Schematic diagram showing vaccine development using reverse vaccinology

Current applications of high-throughput sequencing technologies

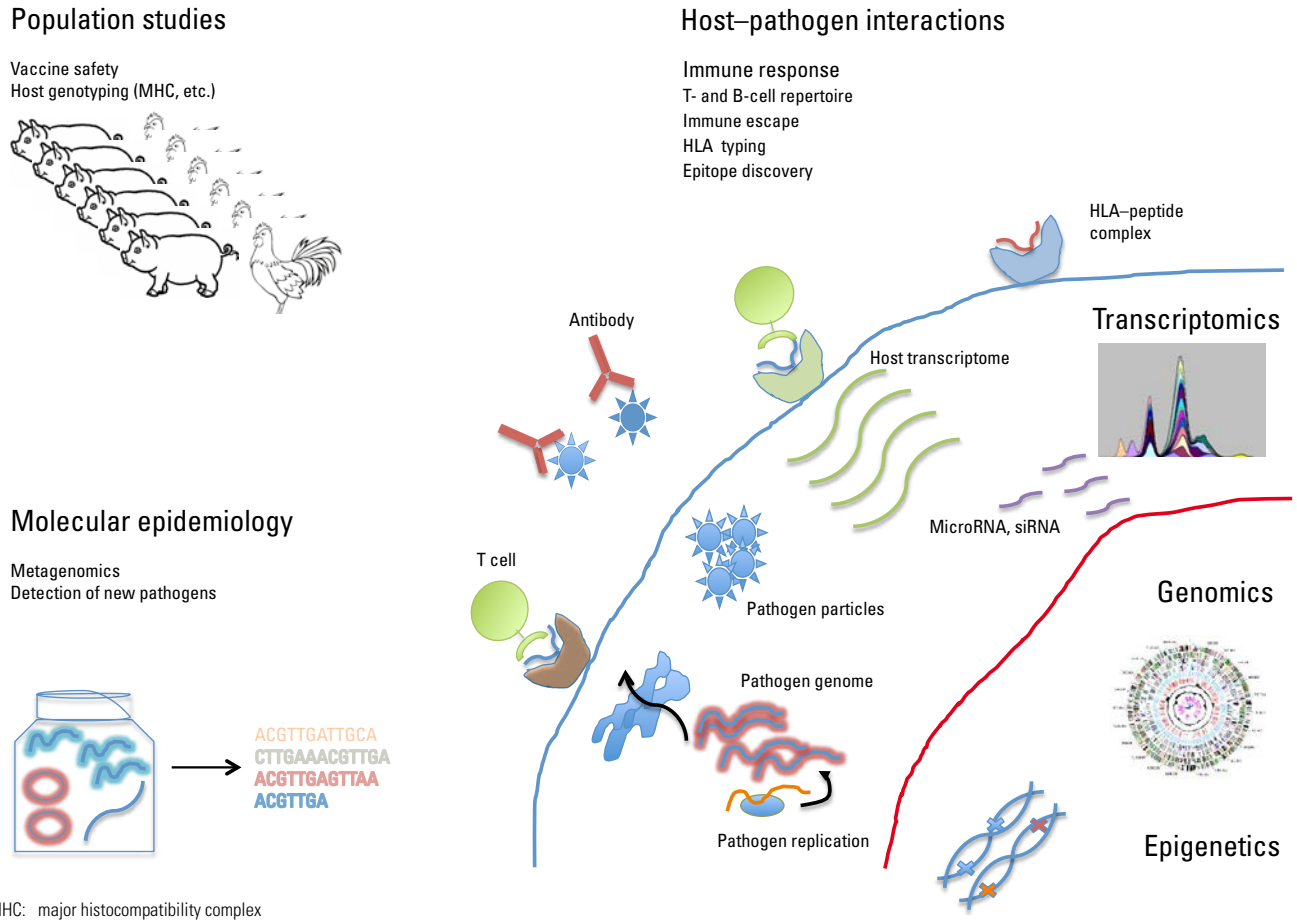
There are three main benefits of high-throughput sequencing technologies: deep sequencing of mixed populations of DNA or RNA genomes is now possible; a large volume of data is generated; and their versatility allows the rapid, direct measurement of whole genomes and transcriptomes under a variety of sample conditions. Today, the major applications of NGS are genome sequencing, transcriptome analysis (i.e. RNA-Seq), DNA–protein binding analysis (i.e. chromatin immunoprecipitation [CHIP]-Seq) and analysis of histone modification (NGS-methylation; for a comprehensive review, see [4]). Recent developments mean that these analyses can now be performed in single cells, thus providing further insight into the heterogeneous complexity of biological systems (18).

For vaccine research, the most obvious application of NGS is the analysis of pathogen genomes (19). In the '-omics' era, a new systems biology approach is influencing vaccine research; hence, high-throughput sequencing now forms an integral component of research into host–pathogen interactions (Fig. 2; adapted from [20]). Several studies have used NGS to analyse the genomes and transcriptomes of both host and pathogen (21, 22). This technology has also been used to investigate diversity in host immune responses, e.g. of T and B cells (23, 24, 25).

Next-generation sequencing and genomics

Deep sequencing the genomes of infectious agents

A key advantage of NGS is its ability to detect low-frequency genomic variants (26), which is important for both genetic and infectious diseases research (27). NGS analyses have also revealed complex genomic alterations: somatic gene rearrangements in host tissues or cells are far more common than expected; and far more variation between individuals is caused by copy number variations than by the vast number of recognised SNPs (28). Research into the virology of human diseases has pioneered the deep sequencing approach by combining the small genome size and fast evolution of viruses with rapid developments in NGS technology. This has led to important developments in data analysis, algorithms and pipelines, in addition to significantly improving our understanding of viral pathogens.



MHC: major histocompatibility complex
HLA: human leukocyte antigen
siRNA: small interfering RNA

Fig. 2
Overview of the current applications of next-generation sequencing in molecular, cellular and population biology that are revolutionising biomedical research, including vaccine research

Applications to human pathogens

The application of deep sequencing of human pathogens includes the comprehensive characterisation of within-host HCV evolution during the early acute phase of infection (29) and investigations into the evolution of viral escape dynamics (30). In these examples, deep sequencing of HCV genomes identified the transmitted (or founder) virus that successfully established the primary HCV infection. Bioinformatics prediction tools were used to identify the viral epitopes presented on MHC class I molecules. There was evidence of strong cytotoxic T-cell responses to target these epitopes during the first few weeks of infection. However, these epitopes were subsequently found to be mutated at a very high frequency during the chronic phase of infection, thus contributing to immune escape. The transmitted virus has also been targeted in HIV research, with the goal of identifying early phenotypes of HIV infections for preventative vaccination (31). However, a major hurdle to overcome in the design of T-cell-based vaccines against rapidly mutating viruses is the rapid evolution of escape variants from the MHC-I antigen used in the vaccine.

Applications in veterinary medicine

Next-generation sequencing technologies are now being used in veterinary medicine to study genomes, transcriptomes and host–pathogen interactions, with a particular focus on disease control and management (see [3] for a review). More recently, with the reduction in cost and technological improvements, deep sequencing approaches have been used on a larger scale in molecular epidemiology for the analysis of large sample sets and in the forensic tracing of viral transmission pathways (16). For instance, NGS-mediated deep sequencing of FMDV has revealed important factors driving disease transmission and genetic bottlenecks both

within and between hosts (32). Another example is the use of NGS to trace the evolution of a low-pathogenicity strain of avian influenza virus (33). In this study, epidemiological information, whole-genome sequence data and deep NGS were combined to characterise the evolution of highly pathogenic strains of avian influenza in a large outbreak in northern Italy. The most virulent strains had evolved from low-pathogenicity variants.

Deep sequencing has also been used to compare the molecular characteristics of circulating viral genomes and follow their evolution from strains used in live attenuated vaccines. For example, in 21 swine herds from Ontario, Canada, with clinical problems due to respiratory disease, deep sequencing confirmed the identity and origin of the circulating influenza viral strains. This study revealed a large, heterogeneous population of circulating H3N2 influenza viruses: evidence for the reassortment of gene segments derived from strains used in live attenuated vaccines led to the identification of three distinct genetic groups (34). NGS has also been used to identify the genetic relationship between several vaccines against circulating strains of classic swine fever (35). The recent use of NGS to identify viral pathogens in clinical and environmental samples led to the discovery of many new, uncharacterised viruses from a number of viral families. For instance, the use of a single unbiased protocol for rapidly sequencing multiple RNA virus genomes from bovine virus stock isolates revealed the simultaneous presence of pestivirus and coronavirus (36). This approach is likely to provide new insight into viral co-infection, evolution and genomic stability, and may aid vaccine design.

Finally, NGS has also been used in vaccine production to test the biosafety of vaccines. For example, NGS analysis of fluids and embryos from specific-pathogen-free eggs has been used to screen for encapsidated viral genomes and viral transcripts, respectively (37).

Towards complex genomic rearrangements

Before the advent of NGS in viral genomics, the canonical method of investigating viral genomes was analysis of the most common variant within the population and identification of its most common mutations, including the fixation of single nucleotide changes (38). An important finding of high-throughput NGS was the discovery of complex genomic rearrangements within relatively small genomes, such as those of viruses. For example, the NGS technologies have led to the discovery of novel structural mutations (such as large deletions and insertions) in the human genome (39), as well as complex recombinant variants in viral genomes. For instance, NGS recently detected chimeric genomes in single-stranded DNA viruses (40). A metagenomics approach revealed that horizontal gene transfer can lead to the formation of novel DNA virus groups with traits similar to those of single-stranded RNA viruses. Another example is the study of arenaviruses, which comprise one of the largest families of human haemorrhagic fever viruses, and are also known to infect both mammals and snakes. In a recent animal study, NGS analysis identified multiple infections and a high level of viral recombination within captive arenavirus-infected snakes, with most snakes being infected with up to four distinct S and 11 distinct L segment genotypes (41). Similarly, sequencing the large genomes of many pathogens, including DNA viruses, bacteria and parasites, has led to the identification of new pathogens via 'metagenomics' (42). In a remarkable example, whole-genome analysis of a comprehensive sample set revealed that a catastrophic event – termed 'chromothripsis' – had occurred in at least 2–3% of human cancers, leading to tens to hundreds of somatic genomic rearrangements, with many genomic segments derived from distinct chromosomes reassembled in a random order to form a derivative chromosome (43).

The genomic stability of viral vaccine strains is increasingly becoming a hot topic in veterinary vaccine research and design. NGS technologies are being increasingly used as powerful tools to simultaneously identify and compare multiple virus strains carrying structural variations in their genomes. Understanding how the genomes of circulating virus strains evolve and differ from those of the original viral vaccine strains may have important consequences for downstream diagnostics, vaccine potency and the outcome of eradication programmes. One example was the discovery of interspecies recombination between co-circulating Australian-origin and European-origin herpesviruses by whole-genome NGS, which indicated that emergent highly virulent strains are recombinant variants of circulating vaccine strains (44).

The unprecedented power to detect complex genomic rearrangements in both small viral genomes and large bacterial genomes is particularly important for vaccine design. In addition to single nucleotide mutations that may impair a specific stage of immune recognition (e.g. T-cell recognition), more complex recombination events such as horizontal gene transfer can result in the loss of entire genomic segments, resulting in almost complete failure of a potential vaccine. The first breakthrough in the application of NGS to reverse vaccinology was whole-genome sequencing of pathogens and their hosts within a time frame of hours to days at a moderate cost.

Next-generation sequencing and transcriptomics

Owing to its versatility and efficiency, RNA-Seq is rapidly becoming the gold standard technology for gathering comprehensive transcriptional information (21). RNA-Seq can detect 25% more transcripts compared with microarrays (45) and can be used in both animals and human cell samples (21) and for obtaining the transcriptome of large-genome pathogens. For example, in human *Cytomegalovirus*, four previously unrecognised protein-coding regions were identified. In addition, large RNA-splicing events involving 229 potential donor and 132 acceptor sites (affecting 58 protein-coding genes) were identified during virion production (46). RNA-Seq can also be used to investigate whether differences in immune responses to a pathogen or candidate vaccine result from alterations in the expression of coding genes or whether these responses are regulated by non-coding portions of the genome. NGS has been used to characterise temporal changes in both host and pathogen gene expression during infection (see [19] for a review). Other NGS studies have focused on non-coding RNAs, for example, microRNAs (miRNAs). MiRNAs constitute a large family of small non-coding RNAs that post-transcriptionally regulate mRNAs and thus influence gene expression programmes and thereby fundamental cellular processes. There is growing evidence that these molecules are relevant to human disease (47).

Next-generation sequencing and epigenetic modifications

'Epigenetics' is the study of heritable chemical changes to DNA and histone molecules, notably DNA methylation and histone deacetylation. These changes are reported to alter gene expression profiles or cellular phenotypes in many human and animal models (48). Epigenetic mechanisms can target both host and viral genomes, and hence may have important effects on the response to viral vaccines. High-throughput DNA sequencing technologies can now be adapted to allow genome-scale mapping of methylation sites. Publication of the first whole-genome 'methylome' of human cells in 2009 revealed that 4–6% of cytosine sites were methylated (49). A major current challenge is to delineate the extent of epigenetic modifications and their importance in infectious diseases and cancer.

Chromatin immunoprecipitation (ChIP) technologies were first developed to study DNA–protein binding sites (50). ChIP-Seq has been combined with NGS to study how transcription factors and other chromatin-associated proteins, such as polymerases, interact with DNA to regulate gene expression. These combined technologies have already revealed that a substantial number of transcription factors are involved in B- and T-cell immune responses (51). In the foreseeable future, it is likely that these technologies will help to achieve a comprehensive understanding of the DNA-binding profiles and epigenetic modification patterns associated with immune responses to both pathogens and vaccines.

Limitations of current high-throughput technologies

For NGS, genomic material must first be obtained from a target pathogen sample. In heterogeneous community samples (i.e. metagenomes), the choice of DNA extraction method can significantly affect the likelihood of detecting an individual member (species or subspecies) of the community (52, 53). Fortunately, this is likely to be less problematic for deep sequencing of single species populations, which generally consist of closely related individuals that have similar properties with respect to cell lysis and recovery of genomic material. However, particularly for viruses with low titres or low sample amounts, the nucleic acid yield may be insufficient and polymerase chain reaction (PCR) amplification may be required prior to sequencing. For whole-virus sequencing, it may be necessary to amplify overlapping fragments along the entire length of the genome prior to random shearing and the ligation of sequencing adapters (54, 55). For RNA viruses such as HCV and HIV, reverse transcription PCR may be necessary (54, 56). Alternatively, reverse transcription of viral RNA may be performed as a separate procedure before regular PCR amplification (57). If whole-genome sequencing is not required, then smaller target regions can be amplified either directly or using multiple, 'nested' rounds of PCR. Hybrid primers with integrated sequencing adapters are often used for the final round (i.e. 'amplicon sequencing') (58).

Sequence-independent RNA virus analysis

For RNA viruses, there has been a significant reduction in the technical bias caused by PCR amplification, which is known to alter the relative frequency of quasispecies components and generate PCR-related chimera. It may be better to use sequence-independent methods for sequencing viral genomes because they do not rely on viral genome sequence knowledge. Several papers have now reported sequence-independent methods of detecting RNA viral genomes, for example recent reports of the complete

Ebola virus, HIV and dengue virus genome sequences (59, 60). These studies used viral RNA obtained from random amplification strategies, such as random hexamers, instead of using the classic amplicon-based approach to target virus-specific sequences. There was thus a significant reduction in selection bias for viral genomes and also in the choice of conserved genomic regions used for primer design.

Bioinformatics

Key challenges of NGS technology are bioinformatics and statistical analyses of the resultant large error-prone datasets to ensure high-quality analysis and data interpretation. Notably, the application of NGS to study genome diversity and structure can be significantly hampered by current methods of sample preparation. For example, the process used for RNA viruses often involves reverse transcription and PCR amplification, both of which can introduce point mutations and recombination events that change the output sequence. This is particularly important when studying RNA virus populations, which are highly diverse. It is therefore essential to distinguish low-frequency variants from technical errors (38, 61).

It is necessary to use quantitative methods to achieve a complete analysis of NGS data. There are still a number of limitations to NGS applications, such as the relatively short read length and very high error rate, which can compromise sequencing quality and thus the range of potential applications. For instance, reconstruction of sequence haplotypes of kilobases in length from short-read NGS data is challenging (62). Rapid advances in new technologies (e.g. Pacific Bioscience SMRT sequencing) are enabling longer read lengths, which makes it possible to sequence genome haplotypes up to 30 kb in length (30 times the length of standard Sanger sequencing). Analysis of entire RNA virus genomes within a single read is now possible, thus enabling the co-evolving multiple mutations to be investigated within the same variant. This could provide an unprecedented ability to identify the mutations responsible for multidrug resistance. However, for large DNA viruses and bacteria, the current read length is still short compared with the genome length (Mb). Furthermore, analysis of these genomes requires sophisticated bioinformatics approaches, which are becoming an increasingly important part of modern vaccinology.

Future directions

Approaches to vaccine development are now taking full advantage of the explosion of high-throughput techniques (63) by utilising genomics, transcriptomics and proteomics (collectively termed 'omics') (18, 20, 51), as well as computational and statistical analyses of high-throughput data. High-throughput sequencing is thus rapidly becoming a core technology in 'vaccinomics' (13, 64), a systems biology approach to vaccine research. The ultimate goal of vaccinomics is to delineate the cellular and molecular pathways involved in the protective immune responses induced by pathogens, and to recapitulate (and potentially enhance) these responses via vaccination utilising genetic signatures with predicted immunogenicity and safety and, hence, efficacy. Since 2008, when the first whole human genome sequence was completed (18), thousands of human genomes have been sequenced. More will become available with the further development of technologies which allow the sequencing of a whole genome within a single day and at an affordable cost. It is therefore likely that the low cost of these technologies will translate to the possibility of large-scale animal genome sequencing, hence raising animal vaccine research to the level of human vaccine research.

Next-generation sequencing can improve current vaccine design

Several difficulties are associated with current vaccines that target rapidly mutating viruses, such as influenza: antigenic change is fast and vaccines are often outpaced within a year. The current protocol has several limitations; for example, the decision about which strains to target is based on a limited set of strains and does not take within-host diversification into account (65). In fact, it remains unclear whether prolonged infections in immunocompromised individuals contribute to viral evolution and, hence, the generation of new strains (66). Deep NGS may also reveal the evolutionary dynamics within a single host and significantly contribute to our understanding of the epidemiological evolution of pathogens. In addition, inactivated vaccines provide only partial immunity and lack cross-protection, especially against emerging variants that have dramatically altered antigenic domains. The technical limitations of current NGS approaches could prevent their successful application in vaccine design by reducing the accuracy of measurements and the reliability of results. Nevertheless, ongoing developments in this field are slowly but convincingly establishing a high-quality standard and more accurate experimental design. For example, it is now possible to reliably obtain the genomic, transcriptomic and proteomic profiles of a single cell (17, 67).

Vaccines that induce a broader immune response

Most current vaccines aim to generate a neutralising antibody response against the pathogen. This response,

however, is only one of several closely interconnected, protective arms of the host immune system. The major reasons for using such a focused approach are that the neutralising antibody response is relatively easy to induce and downstream immunoprotective events may also involve T-cell immunity. Characteristics of the vaccine, however, may hinder broad stimulation of the immune system because pathogen-derived antigens may be highly specific and non-immunogenic, or synthetic antigens may not represent life-cycle stages of the pathogen. T-cell responses require specific antigen processing and presentation within both infected cells and antigen-presenting cells (via cross-presentation). These mechanisms are tightly regulated by costimulatory signalling pathways driven by early innate responses.

High-throughput genomics can help to generate antigens that can be used together to stimulate a broader immune response than can be induced by a live attenuated vaccine. For instance, a cocktail of multiple antigens could be engineered to directly target T cells. An example of this approach is the HIV mosaic vaccine, which was recently trialled for safety in humans. This vaccine was based on bioinformatics-based antigen prediction from HIV sequencing data (68). In the future, genomic and bioinformatics tools will be increasingly applied to identify T- and B-cell epitopes that can be used in vaccine development (69, 70).

Genomics and cancer vaccines

Sequencing large genomes is becoming more feasible, which is rapidly increasing research into anti-cancer vaccines. One example is the recent discovery that a facial tumour of the Tasmanian devil is caused by a transmissible disease. This discovery resulted from accurate annotation of the Tasmanian devil genome using NGS and a sophisticated comparative genomics analysis. This novel approach has opened new avenues into the development of vaccines to prevent this disease (71).

Le séquençage à haut débit et la conception des vaccins

F. Luciani

Résumé

Les technologies de séquençage de nouvelle génération (SNG) ont donné une impulsion nouvelle à la recherche sur le génome. Le niveau accru de profondeur et de résolution de séquençage qui en résulte se traduit par une précision génomique inégalée, ce qui à son tour donne lieu à une meilleure perception de la complexité des génomes animaux et humains et de ceux des agents pathogènes. Cela a également ouvert de nouvelles perspectives à la recherche sur les vaccins. D'une part, la complexité accrue du génome nous invite à étudier et à mieux comprendre la biologie des agents pathogènes et les interactions entre ceux-ci et leurs hôtes. D'autre part, les données de plus en plus nombreuses sur le génome permettent d'obtenir des informations cruciales pour mettre au point et concevoir de meilleurs vaccins contre certains agents pathogènes précédemment difficiles à traiter, par exemple les bactéries ou les virus à ARN soumis à des mutations rapides et présentant des interactions complexes avec le système immunitaire de l'hôte. Cette synthèse décrit l'impact des nombreuses applications des technologies de séquençage de nouvelle génération sur la recherche sur les vaccins, en particulier les conséquences dans le domaine de la génomique virale et les travaux récents de virologie humaine et vétérinaire.

Mots-clés

Évolution des virus – Génome viral – Mise au point de vaccins – Quasi-espèce – Réponse immune – Séquençage de nouvelle génération.

Secuenciación de alto rendimiento y concepción de vacunas

F. Luciani

Resumen
Las técnicas de secuenciación de próxima generación han transformado la investigación sobre el genoma. El incremento de la profundidad y la resolución de la secuenciación que estas técnicas han hecho posible ha servido para conocer el genoma con un nivel de detalle sin precedentes y para tomar cada vez más conciencia de la complejidad que reviste el genoma de animales, seres humanos y patógenos. Ello, a su vez, ha traído consigo nuevos métodos de investigación en materia de vacunas. Por un lado, la creciente complejidad que observamos en el genoma pone a prueba nuestra capacidad para estudiar y entender la biología de los patógenos y las interacciones entre patógeno y anfitrión. Por otro lado, del creciente volumen de datos genómicos podemos extraer información básica para concebir y obtener vacunas más eficaces contra patógenos que hasta ahora eran muy difíciles de combatir, como virus ARN o bacterias que mutan con gran rapidez e interaccionan de forma compleja con el sistema inmunitario del anfitrión. El autor explica cómo está influyendo en la investigación sobre vacunas el uso generalizado de técnicas de secuenciación de próxima generación para estudiar e investigar el genoma, centrándose especialmente en las repercusiones que de ahí se siguen en el ámbito de la genómica vírica y refiriéndose a una serie de recientes estudios realizados en animales y personas.

Palabras clave
Concepción de vacunas – Cuasiespecie – Evolución vírica – Genoma vírico – Respuesta inmunitaria – Secuenciación de próxima generación.

References

1. Rappuoli R., Miller H.I. & Falkow S. (2002). – Medicine. The intangible value of vaccination. *Science*, **297** (5583), 937–939. doi:10.1126/science.1075173.

2. Meeusen E.N., Walker J., Peters A., Pastoret P-.P., & Jungersen G. (2007). – Current status of veterinary vaccines. *Clin. Microbiol. Rev*., **20** (3), 489–510. doi:10.1128/CMR.00005-07.

3. Van Borm S., Belák S., Freimanis G., Fusaro A., Granberg F., Höper D., King D.P., Monne I., Orton R. & Rosseel T. (2015). – Next-generation sequencing in veterinary medicine: how can the massive amount of information arising from high-throughput technologies improve diagnosis, control, and management of infectious diseases? *Meth. Molec. Biol*., **1247**, 415–436. doi:10.1007/978-1-4939-2004-4_30.

4. Metzker M.L. (2010). – Sequencing technologies: the next generation. *Nat. Rev. Genet*., **11**, 31–46. doi:10.1038/nrg2626.

5. Schadt E.E., Turner S. & Kasarskis A. (2010). – A window into third-generation sequencing. *Hum. Molec. Genet*., **19** (R2), R227–R240. doi:10.1093/hmg/ddq416.

6. Zimmer S.M. & Burke D.S. (2009). – Historical perspective: emergence of influenza A (H1N1) viruses. *N. Engl. J. Med*., **361**, 279–285. doi:10.1056/NEJMra0904322.

7. Beerenwinkel N. & Zagordi O. (2011). – Ultra-deep sequencing for the analysis of viral populations. *Curr. Opin. Virol*., **1** (5), 413–418. doi:10.1016/j.coviro.2011.07.008.

8. Marz M., Beerenwinkel N., Drosten C., Fricke M., Frishman D., Hofacker I.L., Hoffmann D., Middendorf M., Rattei T., Stadler P.F. & Töpfer A. (2014). – Challenges in RNA virus bioinformatics. *Bioinformatics*, **30** (13), 1793–1799. doi:10.1093/bioinformatics/btu105.

9. McElroy K., Thomas T. & Luciani F. (2014). – Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions. *Microb. Inform. Exp.*, **4**, 1. doi:10.1186/2042-5783-4-1.

10. Tan K.K., Tan Y.C., Chang L.Y., Lee K.W., Nore S.S., Yee W.Y., Mat Isa M.N., Jafar F.L., Hoh C.C. & AbuBakar S. (2015). – Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genomics*, **16**, 93. doi:10.1186/s12864-015-1294-x.

11. Rappuoli R. (2004). – From Pasteur to genomics: progress and challenges in infectious diseases. *Nat. Med.*, **10**, 1177–1185. doi:10.1038/nm1129.

12. Rappuoli R. & Del Giudice G. (1999). – Identification of vaccine targets. *In* Vaccines: from concept to clinic. CRC Press, Boca Raton, Florida, 1–17.

13. Poland G.A. (2007). – Pharmacology, vaccinomics, and the second golden age of vaccinology. *Clin. Pharmacol. Ther.*, **82** (6), 623–626. doi:10.1038/sj.clpt.6100379.

14. Seib K.L., Zhao X. & Rappuoli R. (2012). – Developing vaccines in the era of genomics: a decade of reverse vaccinology. *Clin. Microbiol. Infect.*, **18** (Suppl. 5), 109–116. doi:10.1111/j.1469-0691.2012.03939.x.

15. Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M. *et al.* (1995). – Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269** (5223), 496–512. doi:10.1126/science.7542800.

16. Kao R.R., Haydon D.T., Lycett S.J. & Murcia P.R. (2014). – Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol.*, **22** (5), 282–291. doi:10.1016/j.tim.2014.02.011.

17. Sette A. & Rappuoli R. (2010). – Reverse vaccinology: developing vaccines in the era of genomics. *Immunity*, **33** (4), 530–541. doi:10.1016/j.immuni.2010.09.017.

18. Wang Y. & Navin N.E. (2015). – Advances and applications of single-cell sequencing technologies. *Molec. Cells*, **58** (4), 598–609. doi:10.1016/j.molcel.2015.05.005.

19. Ley T.J., Mardis E.R., Ding L., Fulton B., McLellan M.D., Chen K., Dooling D., Dunford-Shore B.H., McGrath S., Hickenbotham M., Cook L., Abbott R., Larson D.E., Koboldt D.C., Pohl C., Smith S., Hawkins A., Abbott S., Locke D., Hillier L.W., Miner T., Fulton L., Magrini V., Wylie T., Glasscock J., Conyers J., Sander N., Shi X., Osborne J.R., Minx P., Gordon D., Chinwalla A., Zhao Y., Ries R.E., Payton J.E., Westervelt P., Tomasson M.H., Watson M., Baty J., Ivanovich J., Heath S., Shannon W.D., Nagarajan R., Walter M.J., Link D.C., Graubert T.A., DiPersio J.F. & Wilson R.K. (2008). – DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, **456**, 66–72. doi:10.1038/nature07485.

20. Luciani F., Bull R.A. & Lloyd A.R. (2012). – Next generation deep sequencing and vaccine design: today and tomorrow. *Trends Biotechnol.*, **30** (9), 443–452. doi:10.1016/j.tibtech.2012.05.005.

21. Ozsolak F. & Milos P.M. (2010). – RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.*, **12** (2), 87–98. doi:10.1038/nrg2934.

22. Margeridon-Thermet S., Shulman N.S., Ahmed A., Shahriar R., Liu T., Wang C., Holmes S.P., Babrzadeh F., Gharizadeh B., Hanczaruk B., Simen B.B., Egholm M. & Shafer R.W. (2009). – Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. *J. Infect. Dis.*, **199** (9), 1275–1285. doi:10.1086/597808.

23. Burwitz B.J., Ende Z., Sudolcan B., Reynolds M.R., Greene J.M., Bimber B.N., Almeida J.R., Kurniawan M., Venturi V., Gostick E., Wiseman R.W., Douek D.C., Price D.A. & O'Connor D.H. (2011). – Simian immunodeficiency virus SIVmac239 nef vaccination elicits different Tat_{28-35}SL8-specific CD8+ T-cell clonotypes compared to a DNA prime/adenovirus type 5 boost regimen in rhesus macaques. *J. Virol.*, **85** (7), 3683–3689. doi:10.1128/JVI.02112-10.

24. Wu X., Zhou T., Zhu J., Zhang B., Georgiev I., Wang C., Chen X., Longo N.S., Louder M., McKee K., O'Dell S., Perfetto S., Schmidt S.D., Shi W., Wu L., Yang Y., Yang Z.Y., Yang Z., Zhang Z., Bonsignori M., Crump J.A., Kapiga S.H., Sam N.E., Haynes B.F., Simek M., Burton D.R., Koff W.C., Doria-Rose N.A., Connors M., NISC Comparative Sequencing Program, Mullikin J.C., Nabel G.J., Roederer M., Shapiro L., Kwong P.D. & Mascola J.R. (2011). – Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*, **333** (6049), 1593–1602. doi:10.1126/science.1207532.

25. Birzele F., Fauti T., Stahl H., Lenter M.C., Simon E., Knebel D., Weith A., Hildebrandt T. & Mennerich D. (2011). – Next-generation insights into regulatory T cells: expression profiling and FoxP3 occupancy in human. *Nucleic Acids Res.*, **39** (18), 7946–7960. doi:10.1093/nar/gkr444.

26. Meyerson M., Gabriel S. & Getz G. (2010). – Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.*, **11**, 685–696. doi:10.1038/nrg2841.

27. Altshuler D.M., Gibbs R.A., Peltonen L., Dermitzakis E., Schaffner S.F. *et al.* (2010). – Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–58. doi:10.1038/nature09298.

28. Gonzaga-Jauregui C., Lupski J.R. & Gibbs R.A. (2012). – Human genome sequencing in health and disease. *Annu. Rev. Med.*, **63**, 35–61. doi:10.1146/annurev-med-051010-162644.

29. Bull R.A., Luciani F., McElroy K., Gaudieri S., Pham S.T., Chopra A., Cameron B., Maher L., Dore G.J., White P.A. & Lloyd A.R. (2011). – Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog.*, **7**, e1002243. doi:10.1371/journal.ppat.1002243.

30. Bull R.A., Leung P., Gaudieri S., Deshpande P., Cameron B., Walker M., Chopra A., Lloyd A.R. & Luciani F. (2015). – Transmitted/founder viruses rapidly escape from CD8+ T cell responses in acute hepatitis C virus infection. *J. Virol.*, **89** (10), 5478–5490. doi:10.1128/JVI.03717-14.

31. Joseph S.B., Swanstrom R., Kashuba A.D. & Cohen M.S. (2015). – Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nat. Rev. Microbiol.*, **13**, 414–425. doi:10.1038/nrmicro3471.

32. Morelli M.J., Wright C.F., Knowles N.J., Juleff N., Paton D.J., King D.P. & Haydon D.T. (2013). – Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. *Vet. Res.*, **44**, 12. doi:10.1186/1297-9716-44-12.

33. Monne I., Fusaro A., Nelson M.I., Bonfanti L., Mulatti P., Hughes J., Murcia P.R., Schivo A., Valastro V., Moreno A., Holmes E.C. & Cattoli G. (2014). – Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. *J. Virol.*, **88** (8), 4375–4388. doi:10.1128/JVI.03181-13.

34. Grgić H., Costa M., Friendship R.M., Carman S., Nagy E., Wideman G., Weese S. & Poljak Z. (2014). – Molecular characterization of H3N2 influenza A viruses isolated from Ontario swine in 2011 and 2012. *Virol. J.*, **11**, 194. doi:10.1186/s12985-014-0194-z.

35. Zhou W., Gao S., Podgorska K., Stadejek T., Qiu H.J., Yin H., Drew T. & Liu L. (2014). – Rovac is the possible ancestor of the Russian lapinized vaccines LK-VNIVViM and CS strains but not the Chinese strain (C-strain) vaccine against classical swine fever. *Vaccine*, **32** (49), 6639–6642. doi:10.1016/j.vaccine.2014.09.058.

36. Neill J.D., Bayles D.O. & Ridpath J.F. (2014). – Simultaneous rapid sequencing of multiple RNA virus genomes. *J. Virol. Meth.*, **201**, 68–72. doi:10.1016/j.jviromet.2014.02.016.

37. Gagnieur L., Cheval J., Cochet M., Breard E., Gratigny M., Hébert C., Muth E., Viarouge C., Dumarest M., Coulpier M. & Eloit M. (2014). – Analysis by high throughput sequencing of specific pathogen free eggs. *Biologicals*, **42** (2), 218–219. doi:10.1016/j.biologicals.2014.05.004.

38. Beerenwinkel N., Gunthard H.F., Roth V. & Metzner K.J. (2012). – Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front. Microbiol.*, **3**, 329. doi:10.3389/fmicb.2012.00329.

39. Chaisson M.J., Huddleston J., Dennis M.Y., Sudmant P.H., Malig M., Hormozdiari F., Antonacci F., Surti U., Sandstrom R., Boitano M., Landolin J.M., Stamatoyannopoulos J.A., Hunkapiller M.W., Korlach J. & Eichler E.E. (2015). – Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, **517**, 608–611. doi:10.1038/nature13907.

40. Krupovic M., Zhi N., Li J., Hu G., Koonin E.V., Wong S., Shevchenko S., Zhao K. & Young N.S. (2015). – Multiple layers of chimerism in a single-stranded DNA virus discovered by deep sequencing. *Genome Biol. Evol.*, **7** (4), 993–1001. doi:10.1093/gbe/evv034.

41. Stenglein M.D., Jacobson E.R., Chang L.W., Sanders C., Hawkins M.G., Guzman D.S., Drazenovich T., Dunker F., Kamaka E.K., Fisher D., Reavill D.R., Meola L.F., Levens G. & DeRisi J.L. (2015). – Widespread recombination, reassortment, and transmission of unbalanced compound viral genotypes in natural arenavirus infections. *PLoS Pathog.*, **11** (5), e1004900. doi:10.1371/journal.ppat.1004900.

42. Gilbert J.A. & Dupont C.L. (2011). – Microbial metagenomics: beyond the genome. *Ann. Rev. Mar. Sci.*, **3**, 347–371. doi:10.1146/annurev-marine-120709-142811.

43. Stephens P.J., Greenman C.D., Fu B., Yang F., Bignell G.R., Mudie L.J., Pleasance E.D., Lau K.W., Beare D., Stebbings L.A., McLaren S., Lin M.L., McBride D.J., Varela I., Nik-Zainal S., Leroy C., Jia M., Menzies A., Butler A.P., Teague J.W., Quail M.A., Burton J., Swerdlow H., Carter N.P., Morsberger L.A., Iacobuzio-Donahue C., Follows G.A., Green A.R., Flanagan A.M., Stratton M.R., Futreal P.A. & Campbell P.J. (2011). – Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, **144** (1), 27–40. doi:10.1016/j.cell.2010.11.055.

44. Lee S.W., Markham P.F., Coppo M.J., Legione A.R., Markham J.F., Noormohammadi A.H., Browning G.F., Ficorilli N., Hartley C.A. & Devlin J.M. (2012). – Attenuated vaccines can recombine to form virulent field viruses. *Science*, **337** (6091), 188. doi:10.1126/science.1217134.

45. Sultan M., Schulz M.H., Richard H., Magen A., Klingenhoff A., Scherf M., Seifert M., Borodina T., Soldatov A., Parkhomchuk D., Schmidt D., O'Keeffe S., Haas S., Vingron M., Lehrach H. & Yaspo M.L. (2008). – A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, **321** (5891), 956–960. doi:10.1126/science.1160342.

46. Gatherer D., Seirafian S., Cunningham C., Holton M., Dargan D.J., Baluchova K., Hector R.D., Galbraith J., Herzyk P., Wilkinson G.W. & Davison A.J. (2011). – High-resolution human cytomegalovirus transcriptome. *Proc. Natl Acad. Sci. USA*, **108** (49), 19755–19760. doi:10.1073/pnas.1115861108.

47. Esteller M. (2011). – Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874. doi:10.1038/nrg3074.

48. Bernstein B.E., Meissner A. & Lander E.S. (2007). – The mammalian epigenome. *Cell*, **128** (4), 669–681. doi:10.1016/j.cell.2007.01.033.

49. Lister R., Pelizzola M., Dowen R.H., Hawkins R.D., Hon G., Tonti-Filippini J., Nery J.R., Lee L., Ye Z., Ngo Q.M., Edsall L., Antosiewicz-Bourget J., Stewart R., Ruotti V., Millar A.H., Thomson J.A., Ren B. & Ecker J.R. (2009). – Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322. doi:10.1038/nature08514.

50. Robertson G., Hirst M., Bainbridge M., Bilenky M., Zhao Y., Zeng T., Euskirchen G., Bernier B., Varhol R., Delaney A., Thiessen N., Griffith O.L., He A., Marra M., Snyder M. & Jones S. (2007). – Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods*, **4**, 651–657. doi:10.1038/nmeth1068.

51. Northrup D.L. & Zhao K. (2011). – Application of ChIP-Seq and related techniques to the study of immune function. *Immunity*, **34** (6), 830–842. doi:10.1016/j.immuni.2011.06.002.

52. Willner D., Daly J., Whiley D., Grimwood K., Wainwright C.E. & Hugenholtz P. (2012). – Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One*, **7** (4), e34605. doi:10.1371/journal.pone.0034605.

53. Yuan S., Cohen D.B., Ravel J., Abdo Z. & Forney L.J. (2012). – Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One*, **7** (3), e33865. doi:10.1371/journal.pone.0033865.

54. Bull R.A., Eden J.S., Luciani F., McElroy K., Rawlinson W.D. & White P.A. (2012). – Contribution of intra- and interhost dynamics to norovirus evolution. *J. Virol.*, **86** (6), 3219–3229. doi:10.1128/JVI.06712-11.

55. Henn M.R., Boutwell C.L., Charlebois P., Lennon N.J., Power K.A., Macalalad A.R., Berlin A.M., Malboeuf C.M., Ryan E.M., Gnerre S., Zody M.C., Erlich R.L., Green L.M., Berical A., Wang Y., Casali M., Streeck H., Bloom A.K., Dudek T., Tully D., Newman R., Axten K.L., Gladden A.D., Battis L., Kemper M., Zeng Q., Shea T.P., Gujja S., Zedlack C., Gasser O., Brander C., Hess C., Günthard H.F., Brumme Z.L., Brumme C.J., Bazner S., Rychert J., Tinsley J.P., Mayer K.H., Rosenberg E., Pereyra F., Levin J.Z., Young S.K., Jessen H., Altfeld M., Birren B.W., Walker B.D. & Allen T.M. (2012). – Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog.*, **8** (3), e1002529. doi:10.1371/journal.ppat.1002529.

56. Tapparel C., Cordey S., Junier T., Farinelli L., Van Belle S., Soccal P.M., Aubert J.D., Zdobnov E. & Kaiser L. (2011). – Rhinovirus genome variation during chronic upper and lower respiratory tract infections. *PLoS One*, **6** (6), e21163. doi:10.1371/journal.pone.0021163.

57. Parameswaran P., Charlebois P., Tellez Y., Nunez A., Ryan E.M., Malboeuf C.M., Levin J.Z., Lennon N.J., Balmaseda A., Harris E. & Henn M.R. (2012). – Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity. *J. Virol.*, **86** (16), 8546–8558. doi:10.1128/JVI.00736-12.

58. Escobar-Gutiérrez A., Vazquez-Pichardo M., Cruz-Rivera M., Rivera-Osorio P., Carpio-Pedroza J.C., Ruíz-Pacheco J.A., Ruiz-Tovar K. & Vaughan G. (2012). – Identification of hepatitis C virus transmission using a next-generation sequencing approach. *J. Clin. Microbiol.*, **50** (4), 1461–1463. doi:10.1128/JCM.00005-12.

59. Gire S.K., Goba A., Andersen K.G., Sealfon R.S., Park D.J., Kanneh L., Jalloh S., Momoh M., Fullah M., Dudas G., Wohl S., Moses L.M., Yozwiak N.L., Winnicki S., Matranga C.B., Malboeuf C.M., Qu J., Gladden A.D., Schaffner S.F., Yang X., Jiang P.P., Nekoui M., Colubri A., Coomber M.R., Fonnie M., Moigboi A., Gbakie M., Kamara F.K., Tucker V., Konuwa E., Saffa S., Sellu J., Jalloh A.A., Kovoma A., Koninga J., Mustapha I., Kargbo K., Foday M., Yillah M., Kanneh F., Robert W., Massally J.L., Chapman S.B., Bochicchio J., Murphy C., Nusbaum C., Young S., Birren B.W., Grant D.S., Scheiffelin J.S., Lander E.S., Happi C., Gevao S.M., Gnirke A., Rambaut A., Garry R.F., Khan S.H. & Sabeti P.C. (2014). – Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*, **345** (6202), 1369–1372. doi:10.1126/science.1259657.

60. Malboeuf C.M., Yang X., Charlebois P., Qu J., Berlin A.M., Casali M., Pesko K.N., Boutwell C.L., DeVincenzo J.P., Ebel G.D., Allen T.M., Zody M.C., Henn M.R. & Levin J.Z. (2013). – Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res.*, **41** (1), e13. doi:10.1093/nar/gks794.

61. Mardis E.R. (2008). – The impact of next-generation sequencing technology on genetics. *Trends Genet.*, **24** (3), 133–141. doi:10.1016/j.tig.2007.12.007.

62. Beerenwinkel N. (2011). – Ultra-deep sequencing for the analysis of viral populations. *Curr. Opin. Virol.*, **1** (5), 413–418. doi:10.1016/j.coviro.2011.07.008.

63. Rappuoli R., Black S. & Lambert P.H. (2011). – Vaccine discovery and translation of new vaccine technology. *Lancet*, **378** (9788), 360–368. doi:10.1016/S0140-6736(11)60440-6.

64. Poland G.A., Kennedy R.B. & Ovsyannikova I.G. (2011). – Vaccinomics and personalized vaccinology: is science leading us toward a new path of directed vaccine development and discovery? *PLoS Pathog.*, **7** (12), e1002344. doi:10.1371/journal.ppat.1002344.

65. Krammer F. & Palese P. (2015). – Advances in the development of influenza virus vaccines. *Nat. Rev. Drug Discov.*, **14**, 167–182. doi:10.1038/nrd4529.

66. Reperant L.A., Kuiken T., Grenfell B.T. & Osterhaus A.D. (2014). – The immune response and within-host emergence of pandemic influenza virus. *Lancet*, **384** (9959), 2077–2081. doi:10.1016/S0140-6736(13)62425-3.

67. Bendall S.C., Simonds E.F., Qiu P., Amir E.D., Krutzik P.O., Finck R., Bruggner R.V., Melamed R., Trejo A., Ornatsky O.I., Balderas R.S., Plevritis S.K., Sachs K., Pe'er D., Tanner S.D. & Nolan G.P. (2011). – Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*, **332** (6030), 687–696. doi:10.1126/science.1198704.

68. Barouch D.H., Stephenson K.E., Borducchi E.N., Smith K., Stanley K., McNally A.G., Liu J., Abbink P., Maxfield L.F., Seaman M.S., Dugast A.S., Alter G., Ferguson M., Li W., Earl P.L., Moss B., Giorgi E.E., Szinger J.J., Eller L.A., Billings E.A., Rao M., Tovanabutra S., Sanders-Buell E., Weijtens M., Pau M.G., Schuitemaker H., Robb M.L., Kim J.H., Korber B.T. & Michael N.L. (2013). – Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. *Cell*, **155** (3), 531–539. doi:10.1016/j.cell.2013.09.061.

69. Desai D.V. & Kulkarni-Kale U. (2014). – T-cell epitope prediction methods: an overview. *Meth. Molec. Biol.*, **1184**, 333–364. doi:10.1007/978-1-4939-1115-8_19.

70. He L. & Zhu J. (2015). – Computational tools for epitope vaccine design and evaluation. *Curr. Opin. Virol.*, **11**, 103–112. doi:10.1016/j.coviro.2015.03.013.

71. Deakin J.E. & Belov K. (2012). – A comparative genomics approach to understanding transmissible cancer in Tasmanian devils. *Annu. Rev. Genomics Hum. Genet.*, **13**, 207–222. doi:10.1146/annurev-genom-090711-163852.