Deep Sequencing in Pre- and Clinical Vaccine Research

P. Prachi   C. Donati   F. Masciopinto   R. Rappuoli   F. Bagnoli

Novartis Vaccines, Research Center, Siena, Italy

Abstract

Vaccine research has experienced a quantum leap after the beginning of the genomics era. High-throughput sequencing techniques, unlimited computing resources, as well as new bioinformatic algorithms are now changing the way we perform genomic studies. Whole genome sequencing will soon become the gold standard for phylogenetic and epidemiology studies and is already shedding new light on the dynamics of bacterial evolution. We believe that deep sequencing projects, together with structural studies on vaccine candidates, will allow targeting constant epitopes and avoid vaccine failure due to antigenic variability. Systems biology, which is expected to revolutionize vaccine research and clinical studies, greatly relies on high-throughput technologies such as RNA-seq. Furthermore, genomics is a key element to develop safer vaccines, and the accuracy of deep sequencing will allow monitoring vaccine coverage after their introduction on the market.

Copyright © 2013 S. Karger AG, Basel

Introduction

The first generation sequencing technology, based on the Sanger sequencing method (chain termination methodology), has been the molecular biology workhorse for more than 25 years. The first human genome sequence, accomplished using first generation technologies, took roughly 10 years and 3 billion dollars [1, 2]. In the last 7 years, the field of genomics has dramatically advanced, entering into the so-called era of next generation sequencing (NGS, also referred as 2nd generation). As compared to first generation technology, NGS is much cheaper and faster, and this has allowed to perform whole genome studies that could never be afforded before, which are also referred to as massively parallel sequencing or deep sequencing projects [3] (table 1). NGS technologies are based either on sequencing by synthesis on isolated groups of clonally amplified templates or sequencing by ligation reaction controlled by polymerase or ligase. Today many NGS platforms are commercially available. Although these platforms vary in their engineering configuration and sequencing chemistry, they share the common technological feature of massive parallel sequencing of either clonally amplified single molecule (Roche/454, Illumina/Solexa, Life/APG, Applied Biosystems/SoLid platforms, and Dover systems/The Polonator) [1] or of single molecules of DNA in real-time, where the DNA

Fabio Bagnoli
Novartis Vaccines
Via Fiorentina 1
IT–53100 Siena (Italy)
E-Mail fabio.bagnoli@novartis.com

Table 1. Summary of the major features of Sanger-based sequencing and deep sequencing technologies

|	Principle involved	Sample preparation	Read length bp	Cost/megabase Mb	Storage requirement	Microbiological applications	Strengths	Limitations
Sanger (1977)	Dideoxy sequencing or base-specific chain termination	Complex, PCR required and size-based separation of DNA fragments following capillary electrophoresis	800–1,000	USD 500	~1 Mb of data	De novo sequencing of complex genomes Verification of clones or engineered constructs	Relatively long and accurate reads (99.99%) Efficient algorithms for assembly and annotation	Low throughput (~10^6 bp/day) Long run time Labor intensive Cloning bias Time consuming Rare mutations in heterogeneous samples not detected
Deep sequencing (2005)	Sequencing by synthesis (SBS) or sequencing by ligation (SBL)	PCR required or not, depending on the technology	26–964 (depending on the technology)	USD 0.50–85 (depending on the technology)	~1–2 TB raw data (depending on the technology)	Sequencing of specific genomes and mixture of unknown genomes (metagenomics) Gene expression analysis from virtually any samples SNP-based epidemiology Non-coding RNA characterization [8]	Low cost per base Accurate (99.95%) Short-long read length Short run time (2 h–10 days depending on the technology) High-throughput (~10^9 bp/day to 100 Mb in an h)	Storage and management of terabytes of data New algorithms for sequence assembly and annotation required

synthesis is followed by its analysis without interruptions (Pacific BioSciences/PACBIO RS and Visigen biotechnologies/VisiGen).

Given the overwhelming amount of data generated by NGS projects, new algorithms have been recently generated for the alignment of the huge number of short reads released by NGS sequencers, the identification of operons, recombination events as well as for the construction of phylogenetic trees based on single nucleotide polymorphisms (SNPs) [4, 5].

The efficiency of the new sequencers, coupled with random PCR amplification, allows sequencing of virtually any nucleic acid present in a sample. This approach has been named metagenomics and can be used to identify microorganisms present in a sample without prior knowledge of their presence or their genome sequence [6, 7]. Since it can detect non-cultivable species, metagenomics is giving an unprecedented contribution to basic research, identifying new species in any ecological niche. Furthermore, as we will discuss in this review, deep sequencing is also becoming instrumental to vaccine research, phylogenetic and epidemiology studies as well as to predict and monitor vaccine efficacy and safety.

Genomics Is a Catalyzer for Vaccine Discovery

The first example of large scale use of genomic information for the identification of potential vaccine targets was the attempt to develop a vaccine against serogroup B *Neisseria meningitidis* [9] through the so-called reverse vaccinology approach [10]. The idea behind the method is to mine the pathogen's genome with bioinformatic algorithms to identify coding sequences predicted to encode for proteins exposed on the surface of the pathogen or to be secreted in the extracellular milieu. The rationale of this selection relies on the assumption that surface and secreted factors are exposed to the host's immune system and, therefore, potential vaccine targets. Since then, the same approach has been applied to most human bacterial pathogens determining an explosion in vaccine candidate identification [11]. In the attempt to measure the importance of the availability of bacterial genomes for vaccine discovery, we extrapolated all patent applications on bacterial vaccines containing genomic information filed since the 1990's when first microbial genome was sequenced [12]. The number of inventions enormously increased right after the first years of the genomics era, and since the year 2002, it has steadily decreased (fig. 1). We think that this phenomenon is due to the following
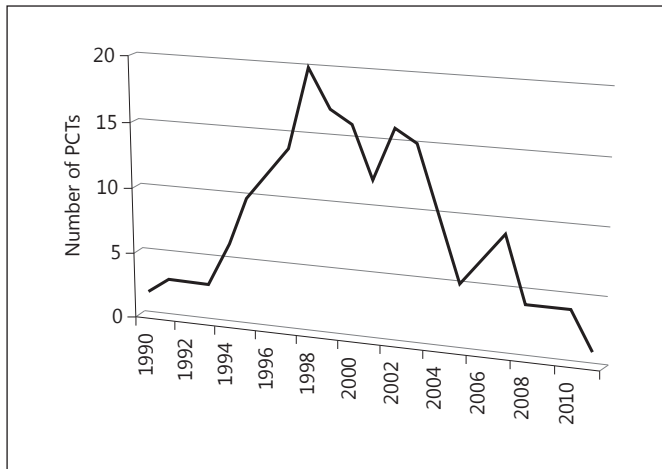
Fig. 1. Patent filing as an objective way of measuring the importance of genomics for vaccine discovery. The graph represents the frequency of most representative international patent applications (PCTs) published in the timeframe 1990–2011. 564 patent families were retrieved, and then selected for the most representative cases, which resulted in a total of 201 patent applications families. (Source: PatBase®)

reason. Initially, availability of genomes allowed the identification of many new vaccine targets as it was demonstrated with reverse vaccinology [11]. This is reflected by the sudden increase of patent filings after the publication of the first genomes in the 90's (fig. 1). Later, legal requirements for granting patent applications became more stringent and in silico data needed to be corroborated by empirical evidence. Therefore, a smaller number of patent applications, but containing more data, has been filed lately. We expect that the vaccine discovery rate will soon be boosted by the advent of deep sequencing studies. Indeed, deep sequencing together with other high-throughput technologies has the potential to provide important information on candidate vaccine targets in addition to the nucleotide sequence. For example, initially reverse vaccinology was performed on a single genome, therefore the predictive power of antigen coverage was low. After the introduction of deep sequencing technologies, a representative collection of epidemiologically relevant strains are sequenced and antigen selection is performed taking into consideration the level of conservation of the antigens. Another example is the use of RNA-seq to determine the expression level of antigens in different conditions, including infected tissues [13, 14]. Increased gene expression during infection, if paralleled by a role in the pathogenesis, is generally considered an important indi-

cation that a protein is a potential vaccine candidate. Indeed, vaccines targeting virulence factors may have 2 protective mechanisms: immune response against the target pathogen and inhibition of its virulence mechanisms. This vaccine discovery approach has been proposed earlier, measuring the transcription profile of antigens by DNA-microarrays [15, 16]. The technical advantages of RNA-seq over DNA-microarray analysis, including the more reliable and accurate quantitation of gene expression, cost-effectiveness and speed, is providing new fuel to this area of vaccine research. Furthermore, as discussed below, deep sequencing can be applied to the design of novel vaccines against variable pathogens and cancers.

Deep Sequencing for Designing Vaccines against Highly Variable Pathogens and Cancers

Most available vaccines are against pathogens whose antigens are relatively stable. Microbes that have rapid and extensive antigenic variability remain a major challenge for vaccine researchers [17]. The most striking example is the human immunodeficiency (HIV). Subunit vaccines derived from the HIV envelope were developed, tested in phase I and phase II clinical studies, and in the mid 1990s were ready to enter phase III efficacy studies. However, in vitro studies demonstrated that the antibodies induced by the vaccines only neutralized the virus strain used to make the vaccine and did not neutralize divergent viruses or primary viruses isolated from patients [17], due to the extremely high rate at which the virus is able to mutate its dominant antigens. A different approach, able to induce broadly neutralizing antibodies and CD8 T-cell response against conserved epitopes, will be probably needed to develop an efficacious vaccine. Similar issues are also hampering the progress towards broadly protective vaccines against several other viruses (e.g. rhinovirus and influenza). A combination of new technologies could help find the solution of this so far insurmountable problem. Deep sequencing can be used to find variable and constant regions of the pathogens genome from hundreds of isolates recovered from infected patients [18]. Once conserved epitopes have been identified, structural studies on the antigens can be performed to understand the degree of surface exposure of the epitopes and to design peptides optimized to generate neutralizing antibodies [19].

A similar, but reversed approach could be used to design anti-cancer vaccines. Indeed, during tumor genesis, cancer cells accumulate mutations generating antigenic

Prachi/Donati/Masciopinto/Rappuoli/
Bagnoli

variability. High-throughput sequencing has been recently used to identify mutations in murine melanome cells ('mutanome'), and the peptides containing the mutations were assessed for prophylactic and therapeutic vaccination in tumor transplant models [20]. The response triggered by the treatment was shown to be specific for the mutated antigens and to control tumor growth. This approach could be used to identify cancer-specific epitopes from biopsies and to generate personalized vaccines.

Genome-Based Phylogenetic Analysis and Vaccine Coverage

Comparative genomics has shown that genetic variability within bacterial species is much larger than expected, leading to the definition of the species pan-genome [21]. From a practical point of view, this unexpected finding led to the conclusion that a solid understanding of the population genetics of the bacterial species is fundamental for the formulation of vaccines with broad coverage [22]. At the present, phylogenetics and population genetics have been mostly based on molecular typing methods, the most successful of which is the multi locus sequence typing (MLST) [23]. However, it is now becoming evident that the level of resolution achievable with MLST is limited, and we should replace it with whole-genome based typing. Indeed, several recent studies have shown that strains belonging to the same clonal complex (CC), as defined on the basis of MLST, can have several substantial differences. One of the most striking examples is a recent publication on the analysis of *Staphylococcus aureus* CC30 [24]. High-throughput sequencing of a collection of historical and contemporary clones of MRSA *S. aureus* has shown that contemporary CC30 strains have a common ancestor with the phage type 80/81, which was responsible for an epidemic wave in Australia, Great Britain, Canada, and the United States in the 1950s. Phylogenetic analysis using genome-wide SNPs has shown that the contemporary CC30 strains harbor SNPs which correlate with substantial difference in the pathogenicity of the strains. These observations could not be predicted on the basis of MLST.

Genome wide screening of SNPs is also instrumental to track the evolution, demographic expansion and geographic dispersal of the species, as it was recently performed on European strains of the *S. aureus* ST225 lineage [25], to identify the dominant haplotypes and pathotypes within an single urban district of the mono-

morphic species *Salmonella enterica* serovar Tiphy [26, 27] and to investigate on historical pandemics of *Yersinia pestis* [28]. Genome studies are also demonstrating that MLST-based epidemiology can fail to identify macroscopic differences among strains. For example, *S. aureus* strains of sequence type ST239 (clonal complex CC8, as determined by eBURST) display evidence of a large recombination event involving a region of approximately 557 kb spanning the origin of replication that appears to have been donated by a CC30 strain [29].

Another important application of deep sequencing that is rapidly emerging is the post-marketing monitoring of vaccine coverage. The launch of the PCV7 pneumococcal vaccine in 2000 on the market provided an unprecedented opportunity to measure the vaccine-induced selective pressure. Since 2003, pneumococcal strains in which a capsular switch from serotype 4 (contained in the vaccine) to serotype 19A (not in the vaccine) were identified [30]. Genomic information was shown to be critical to understand the serological replacement event [31], showing that vaccine strains can switch to non-vaccine capsule types by homologous recombination of the entire capsular biosynthesis locus. Since genome-wide studies of large collection of strains of the same lineage has shown that these exchange events have occurred frequently in the past, also in the absence of a selective pressure against a specific serotype [30], it is conceivable that the efficacy of existing vaccines could last less than expected. On the basis of these observations, several authors have proposed the use of serotype-independent vaccines based on combination of conserved proteins that presumably would cover all circulating strains and avoid the capsular switch phenomena observed with PCV7 [32].

Characterization of Human Immune Response to Vaccines by RNA-seq

Recently, holistic approaches to identify and characterize immune responses to vaccines in humans have been proposed [17, 33–36]. The driving idea behind these approaches is to integrate different sets of biological data from as many hierarchical levels as possible to visualize 'emergent properties' that are not demonstrated by their individual parts and cannot be predicted from the parts alone. The sets of data that have been considered so far for these studies differ depending on the approach, but they have a certain degree of overlap. Vaccinomics focuses on unrevealing associations between the host genetic background and its responses to vaccination [34,

Table 2. Comparison of reverse vaccinology in the year 2000 and today, after the introduction of deep sequencing technologies

	RV in 2000 (Sanger sequencing-based)	RV today (deep sequencing-based)
	basic principle: identification of code sequences predicted to encode for surface exposed or secreted proteins	
Major outcomes	*Antigen selection* • Performed from 1 genome (low predictive power of antigen coverage)	*Antigen selection* • Performed on >10 genomes (high predictive power of antigen coverage) • High-throughput identification of antigens expressed in vivo by RNA-seq • Identification of conserved antigens and epitopes in hyper-variable pathogens • Identification of cancer-specific epitopes • Genome-based phylogenetic analysis for the identification of broadly conserved antigens
	Vaccine development • Fast identification of protein-based vaccines with no homology to human antigens and that can be developed using high standards of purity ensuring greater safety profile as compared to attenuated and killed vaccines	*Vaccine development* • Fast identification of protein-based vaccines with no homology to human antigens and that can be developed using high standards of purity ensuring greater safety profile as compared to attenuated and killed vaccines • Characterization of human immune response to vaccine antigens by RNA-seq • Use of deep sequencing to test vaccine lots for the presence of contaminant nucleic acids • Monitoring of vaccine coverage after their introduction in the market
Applicability	• Cultivable pathogens • All known microorganisms	• Cultivable and non-cultivable pathogens • All known and potentially unknown microorganisms

RV = Reverse vaccinology.

Note: while the basic principle of RV is still the same, the table summarizes the most relevant improvements brought by deep sequencing.

37]. On the other hand, systems biology studies have primarily analyzed the interrelationship existing between gene expression profiles and immune responses triggered by vaccination into the host [17, 38]. Recently, a systems biology approach has successfully been used to predict the immune responses induced by the live-attenuated yellow fever virus vaccine YF-17D. Gene expression profiles induced in the blood of vaccinees were used to identify genes that regulate virus innate sensing and type I interferon production. In addition, computational analyses identified a gene signature that predicted CD8 T-cell and neutralizing antibody responses to YF-17D [39, 40]. High-throughput technologies, such as deep sequencing, are vital to these kind of approaches in order to facilitate the discovery of rare polymorphisms and alternative splice variants as well as to measure how vaccines affect host gene expression (RNA-seq) [35, 38, 41].

Genomics to Develop Safer Vaccines

The availability of genomic sequences allows designing and developing safer vaccines. Indeed, by comparing the sequence of vaccine antigens with the human genome, it is possible to reveal homologies that could potential elicit autoimmunity. Another recent application of genomics to vaccine safety is the identification of contaminant nucleic acids in licensed human vaccines through deep sequencing. As compared to PCR-based strategies, which is used for the identification of specific target sequences, deep sequencing can reveal the presence of virtually every DNA or RNA molecule in the sample without prior knowledge of their sequence. Indeed, it has been shown that its application to live-attenuated vaccines can detect the presence of adventitious viruses, sequence changes in the attenuated virus sequence and minority variants [42–44]. For example, the rotavirus vaccine Ro-

Prachi/Donati/Masciopinto/Rappuoli/
Bagnoli

tarix was found to be unexpectedly contaminated with porcine circovirus-1 DNA by deep sequencing [42]. Furthermore, this technology can also be applied to the screening of cells and all the reagents used for the production of viral or subunit vaccines. Several viruses have been identified in mammalian as well as insect cells used in vaccine manufacturing [43]. Therefore, high-throughput sequencing technologies will soon become a key factor for vaccine lot release as well as characterization and improvement of reagents used in vaccine production.

Conclusions

The importance of genomics for vaccine research and development has been well established in the field. Deep sequencing is now bringing new fuel to virtually every area of vaccinology. As compared to the year 2000, when the first reverse vaccinology genome study was published [9], availability of NGS technology is significantly contributing to the improvement of the approach (table 2). Successful vaccines developed so far are the ones against relatively slow evolving pathogens [17]. To design vaccines that are able to cope with antigenic variability, we need deep sequencing to identify epitopes conserved virtually among any circulating strain. Using this technique, a medium-size laboratory can sequence hundreds of isolates of the same species and, on the basis of this information, identify vaccine candidates as well as perform much more accurate phylogenetic studies as compared to traditional methods.

The possibility to use these techniques to monitor the host response to vaccination and disease in a large collection of individuals and correlate it to the genetic background of the host will greatly facilitate vaccinomics and systems biology studies to predict and optimize vaccine outcomes (i.e. maximizing the appropriate immune responses and minimizing vaccine failure and adverse events) as well as to discover signature of protection in humans.

In conclusion, the level of analytical resolution and the quality of vaccines that we can now produce with the aid of deep sequencing is absolutely unprecedented. We believe that this is what the public opinion needs to hear to be reinsured on the safety of immunoprophylactic campaigns.

References

1 Metzker ML: Sequencing technologies – the next generation. Nat Rev Genet 2010;11:31–46.

2 Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, Niederweis M, Gundlach JH: Nanopore DNA sequencing with MspA. Proc Natl Acad Sci USA 2010;107:16060–16065.

3 Marguerat S, Wilhelm BT, Bahler J: Next-generation sequencing: applications beyond genomes. Biochem Soc Trans 2008;36:1091–1096.

4 Chan EY: Next-generation sequencing methods: impact of sequencing accuracy on SNP discovery. Methods Mol Biol 2009;578:95–111.

5 Hu B, Xie G, Lo CC, Starkenburg SR, Chain PS: Pathogen comparative genomics in the next-generation sequencing era: genome alignments, pangenomics and metagenomics. Brief Funct Genomics 2011;10:322–333.

6 Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al: The sequence of the human genome. Science 2001;291:1304–1351.

7 Relman DA: Microbial genomics and infectious diseases. N Engl J Med 2011;365:347–357.

8 Peng X, Gralinski L, Ferris MT, Frieman MB, Thomas MJ, Proll S, Korth MJ, Tisoncik JR, Heise M, Luo S, Schroth GP, Tumpey TM, Li C, Kawaoka Y, Baric RS, Katze MG: Integrative deep sequencing of the mouse lung transcriptome reveals differential expression of diverse classes of small RNAs in response to respiratory virus infection. MBio 2011;2:pii:e00198-11.

9 Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B, Galeotti CL, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Scarselli M, Storni E, Zuo P, Broeker M, Hundt E, Knapp B, Blair E, Mason T, Tettelin H, Hood DW, Jeffries AC, Saunders NJ, Granoff DM, Venter JC, Moxon ER, Grandi G, Rappuoli R: Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 2000;287:1816–1820.

10 Rappuoli R: Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 2001;19:2688–2691.

11 Bagnoli F, Baudner B, Mishra RP, Bartolini E, Fiaschi L, Mariotti P, Nardi-Dei V, Boucher P, Rappuoli R: Designing the next generation of vaccines for global public health. OMICS 2011;15:545–566.

12 Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995;269:496–512.

13 Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, Waldor MK: RNA-seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression. Cell Host Microbe 2011;10:165–174.

14 Azhikina T, Skvortsov T, Radaeva T, Mardanov A, Ravin N, Apt A, Sverdlov E: A new technique for obtaining whole pathogen transcriptomes from infected host tissues. Biotechniques 2010;48:139–144.

15 Grifantini R, Bartolini E, Muzzi A, Draghi M, Frigimelica E, Berger J, Ratti G, Petracca R, Galli G, Agnusdei M, Giuliani MM, Santini L, Brunelli B, Tettelin H, Rappuoli R, Randazzo F, Grandi G: Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. Nat Biotechnol 2002;20:914–921.

16 Stewart GR, Wernisch L, Stabler R, Mangan JA, Hinds J, Laing KG, Young DB, Butcher PD: Dissection of the heat-shock response in Mycobacterium tuberculosis using mutants and microarrays. Microbiology 2002;148:3129–3138.

17 Rappuoli R, Aderem A: A 2020 vision for vaccines against HIV, tuberculosis and malaria. Nature 2011;473:463–469.

18 Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, Sun L, Zhang T, Hu Y, Du J, Wang J, Jin Q: Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol 2011;49:3463–3469.

19 Dormitzer PR, Ulmer JB, Rappuoli R: Structure-based antigen design: a strategy for next generation vaccines. Trends Biotechnol 2008;26:659–667.

20 Castle JC, Kreiter S, Diekmann J, Lower M, van de Roemer N, de Graaf J, Selmi A, Diken M, Boegel S, Paret C, Koslowski M, Kuhn AN, Britten CM, Huber C, Tureci O, Sahin U: Exploiting the mutanome for tumor vaccination. Cancer Res 2012;72:1081–1091.

21 Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM: Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. Proc Natl Acad Sci USA 2005; 102:13950–13955.

22 Takala SL, Plowe CV: Genetic diversity and malaria vaccine design, testing and efficacy: preventing and overcoming 'vaccine resistant malaria'. Parasite Immunol 2009;31:560–573.

23 Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG: Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 1998;95: 3140–3145.

24 DeLeo FR, Kennedy AD, Chen L, Bubeck Wardenburg J, Kobayashi SD, Mathema B, Braughton KR, Whitney AR, Villaruz AE, Martens CA, Porcella SF, McGavin MJ, Otto M, Musser JM, Kreiswirth BN: Molecular differentiation of historic phage-type 80/81 and contemporary epidemic *Staphylococcus aureus*. Proc Natl Acad Sci USA 2011;108: 18091–18096.

25 Nübel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK, Zemlicková H, Leblois R, Wirth T, Jombart T, Balloux F, Witte W: A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. PLoS Pathog 2010;6:e1000855.

26 Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S, Chinh NT, Le TA, Acosta CJ, Farrar J, Dougan G, Achtman M: Evolutionary history of *Salmonella typhi*. Science 2006;314: 1301–1304.

27 Baker S, Holt K, van de Vosse E, Roumagnac P, Whitehead S, King E, Ewels P, Keniry A, Weill FX, Lightfoot D, van Dissel JT, Sanderson KE, Farrar J, Achtman M, Deloukas P, Dougan G: High-throughput genotyping of *Salmonella enterica* serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban district of Jakarta, Indonesia. J Clin Microbiol 2008;46:1741–1746.

28 Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li Y, Cui Y, Thomson NR, Jombart T, Leblois R, Lichtner P, Rahalison L, Petersen JM, Balloux F, Keim P, Wirth T, Ravel J, Yang R, Carniel E, Achtman M: *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. Nat Genet 2010;42:1140–1143.

29 Robinson DA, Enright MC: Evolution of *Staphylococcus aureus* by large chromosomal replacements. J Bacteriol 2004;186:1060–1064.

30 Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD: Rapid pneumococcal evolution in response to clinical interventions. Science 2011; 331:430–434.

31 Golubchik T, Brueggemann AB, Street T, Gertz RE, Jr., Spencer CC, Ho T, Giannoulatou E, Link-Gelles R, Harding RM, Beall B, Peto TE, Moore MR, Donnelly P, Crook DW, Bowden R: Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. Nat Genet 2012;44:352–355.

32 Harfouche C, Filippini S, Gianfaldoni C, Ruggiero P, Moschioni M, Maccari S, Pancotto L, Arcidiacono L, Galletti B, Censini S, Mori E, Giuliani M, Facciotti C, Cartocci E, Savino S, Doro F, Pallaoro M, Nocadello S, Mancuso G, Haston M, Goldblatt D, Barocchi MA, Pizza M, Rappuoli R, Masignani V: RrgB321, a fusion protein of the three variants of the pneumococcal pilus backbone RrgB, is protective in vivo and elicits opsonic antibodies. Infect Immun 2012;80:451–460.

33 Pulendran B, Li S, Nakaya HI: Systems vaccinology. Immunity 2010;33:516–529.

34 Poland GA, Kennedy RB, Ovsyannikova IG: Vaccinomics and personalized vaccinology: is science leading us toward a new path of directed vaccine development and discovery? PLoS Pathog 2011;7:e1002344.

35 Ovsyannikova IG, Poland GA: Vaccinomics: current findings, challenges and novel approaches for vaccine development. AAPS J 2011;13:438–444.

36 Buonaguro L, Pulendran B: Immunogenomics and systems biology of vaccines. Immunol Rev 2011;239:197–208.

37 Poland GA, Ovsyannikova IG, Jacobson RM: Personalized vaccines: the emerging field of vaccinomics. Expert Opin Biol Ther 2008;8: 1659–1667.

38 Nakaya HI, Li S, Pulendran B: Systems vaccinology: learning to compute the behavior of vaccine induced immunity. Wiley Interdiscip Rev Syst Biol Med 2012;4:193–205.

39 Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, Pirani A, Gernert K, Deng J, Marzolf B, Kennedy K, Wu H, Bennouna S, Oluoch H, Miller J, Vencio RZ, Mulligan M, Aderem A, Ahmed R, Pulendran B: Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol 2009;10:116–125.

40 Gaucher D, Therrien R, Kettaf N, Angermann BR, Boucher G, Filali-Mouhim A, Moser JM, Mehta RS, Drake DR, 3rd, Castro E, Akondy R, Rinfret A, Yassine-Diab B, Said EA, Chouikh Y, Cameron MJ, Clum R, Kelvin D, Somogyi R, Greller LD, Balderas RS, Wilkinson P, Pantaleo G, Tartaglia J, Haddad EK, Sekaly RP: Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. J Exp Med 2008;205:3119–3131.

41 Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63.

42 Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, Delwart EL: Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. J Virol 2010;84:6033–6040.

43 Onions D, Cote C, Love B, Toms B, Koduri S, Armstrong A, Chang A, Kolman J: Ensuring the safety of vaccine cell substrates by massively parallel sequencing of the transcriptome. Vaccine 2011;29:7117–7121.

44 Neverov A, Chumakov K: Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines. Proc Natl Acad Sci USA 2010;107:20063–20068.