Review Article

High throughput genomic and proteomic technologies in the fight against infectious diseases

Alessandro Tanca[1,2], Massimo Deligios[1,2], Maria Filippa Addis[2], Sergio Uzzau[1,2]

[1] *Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy*
[2] *Porto Conte Ricerche, Tramariglio, Alghero, Italy*

Abstract

New technologies have shown significant promise in the fight against infectious diseases, with the discovery of novel molecular targets for *in vitro* diagnostics and the improved design of vaccines. In developing countries, especially in areas of neglected diseases and resources-poor settings, a number of technological innovations are further needed, such as the integration of old and new biomarkers in suitable analysis platforms, the simplification of existing analysis systems, and the improvement of sample preservation and management. However, in these areas, identification of new biomarkers for infectious diseases is still a core issue in the diagnostic quest. Similarly, new technologies will allow scientists to design vaccines with improved immunogenicity, efficacy and safety in the local area, according to the circulating pathogenic strains and the genetic background of the population to be immunized.

In this work we review the current omics-based technologies and their potential for accelerating the development of next generation vaccines and the identification of biomarkers suitable for point-of-care (POC) diagnostic applications.

Key words: genomics; proteomics; proteogenomics; reverse vaccinology; biomarkers

J Infect Dev Ctries 2013; 7(3):182-190.

(Received 25 September 2012 – Accepted 16 February 2013)

Key features and solutions of genomics and proteomics

In 1995, with the whole sequencing of the first bacterial genome using the automated Sanger sequencing technology (today called first-generation sequencing), microbial genomics was officially born [1]. So far, 3,363 bacterial genome projects have been completed (http://www.genomesonline.org). In the first years, a bacterial sequencing project needed a huge investment, up to millions of dollars, and years of labor time. Since 2005, however, Roche/454, Life Technologies SOLiD, and Illumina have been responsible for an unprecedented technological evolution with the first high-throughput sequencing technologies (so-called next-generation sequencing, NGS). Any laboratory with access to these technologies has gained the possibility to invest in a bacterial sequencing project with a cost of a few hundred dollars. Most recently, new single-molecule sequencing technologies have been reported that enable eliminating template amplification steps, avoiding errors in the DNA polymerization chemistry and reaching multikilobase reads (third-generation sequencing technologies) [2]. However, the progress of technology in automated sequencing does not allow completely finished genomes to be obtained and further work is generally required for closing genome sequence gaps with combinatorial PCR and Sanger sequencing [3]. Most importantly, annotation of genes is required to convert genome sequences into a better understanding of the biology of microorganisms. Annotation provides details on the bacterial proteins and on the pathways involved in metabolism, pathogenicity, horizontal transfer, and other highly relevant features related to vaccine development and innovative diagnostic and therapeutic strategies. To date, several automated pipelines are available to reduce the labor time considerably, but manual curation is still required [4]. Software systems often introduce poor annotation and propagate errors from other genomes. Additionally, manual curation is necessary to correctly identify pseudogenes (genes that have lost their function because of an insertion or a frameshift mutation compared with an orthologue in a correlated species). Finally, unfinished high-quality drafts provide reference databases enabling the deep proteomic analysis of bacteria.

Following the astonishing success of genomics at the end of the last century, "classical" protein biochemistry evolved into a high-throughput, systematic, and holistic science called proteomics [5]. The term "proteome" was originally defined in 1995 as "the total protein complement able to be encoded by a given genome" [6], and subsequently specified as "the entire protein complement expressed by a genome, or by a cell or tissue type", to highlight the dynamic nature of protein expression trends depending on several factors, such as cell type, cell cycle state, and environmental influences [7]. Furthermore, proteomics has not been restricted to knowing the whole list of proteins expressed by a cell, but it also comprises the analysis of splicing variants and co- and post-translational modifications, as well as the identification of protein complexes and protein-protein interactions [8].

However, the greater part of the most biologically important proteins are present in a few copies per cell, and have to be identified and quantified in the presence of a large excess of many other proteins. Therefore, handling the huge complexity and high dynamic range of a proteome represents a key technical challenge in proteomics. Several strategies have been successfully employed (and need to be developed) to address this issue, including protein and peptide fractionation, selective enrichment based on specific interaction with antibodies or other molecules, as well as abundant protein depletion or dynamic range "normalization" by means of various systems [9].

Globally speaking, proteomic research can be divided into two categories: "unbiased" (or discovery-oriented) and "targeted" (or system-oriented) proteomics. In a typical discovery-oriented experiment, a complex biological sample is analyzed by separating and identifying as many proteins as possible, thus considering the entire proteome (comprising known and unknown proteins) independently from specific hypotheses based on previous knowledge. Conversely, in a systems-oriented study, a subset of proteins is selected by the investigator, and then analyzed and precisely quantified, specifically focusing on protein panels known to be related by sequence, biological function, or diagnostic potential [10].

Current technologies enabling qualitative and quantitative analysis of microbial proteomes

Two-dimensional electrophoresis (2-DE) was developed two decades before the term "proteomics" was coined [11,12]. This technique entails separation of complex protein mixtures on the basis of pI using isoelectric focusing (first dimension) and further of molecular mass using SDS-PAGE (second dimension). After gel staining for protein visualization, image analysis is performed to match and compare protein patterns within gel replicates and therefore detect quantitative changes based on spot intensity. Usually, protein spots are then picked from the 2-D gel, digested with trypsin, and eventually analyzed by MALDI mass spectrometry (MS), which enables protein identification through peptide mass fingerprinting. Although the forerunner of proteomic techniques, 2-DE is still extensively used because of its ability to separate, display, and store thousands of proteins in one gel; however, 2-DE has some important shortcomings which should be mentioned. In fact, the gel-based approach is complex, labor-intensive, and usually poorly reproducible; furthermore, proteins with extreme MW, pI and hydrophobicity values cannot be properly analyzed by 2-DE [13].

A great step forward was achieved with the development of difference gel electrophoresis (DIGE). With this method, protein samples are labeled with different cyanine dyes, mixed, and co-separated in the same gel. The co-migrated protein spots of the different samples are detected by scanning at different wavelengths, and their abundance ratios are determined with dedicated software [14]. Moreover, gel-to-gel variability is minimized by using the same internal standard among different gels, thus allowing a significant increase in experimental reproducibility and throughput [15].

Around the turn of the millennium, proteomics researchers began to look for alternatives to the gel-based workflow. The most successful approach, named "shotgun" or Multidimensional protein identification technology (MudPIT),), employs tryptic digestion of the entire protein mixture and analysis of the peptides with the combination of nanoscale liquid chromatography and electrospray tandem mass spectrometry [16,17]. In the gel-free workflow, several combinations of peptide separation methods can be used prior to LC-MS/MS analysis, including strong cation exchange and reversed phase high-performance liquid chromatography, liquid phase isoelectric focusing, and capillary electrophoresis [18]. As a "hybrid" alternative, a combination of SDS-PAGE, in-gel digestion and LC-MS/MS (named GeLC-MS/MS) can be also employed [19]. Furthermore, methods enabling quantitative analysis of proteomes by mass spectrometry have been developed. Among them, two main approaches can be distinguished, namely stable isotope labeling (including ICAT [20], iTRAQ [21], SILAC [22], AQUA [23]) and label-free quantitation (including spectral counting and peak intensity/area quantitation) [24].

Among targeted proteomics techniques, array-based and MS-based approaches should be mentioned. Protein arrays are solid-phase ligand binding assay systems, in which proteins are immobilized on surfaces; these assays are multiplexed and often miniaturized. They are comprised of "functional arrays", able to detect protein-protein, protein-DNA, protein-small molecule interactions, and "capture arrays", and are used to detect and quantify analytes in complex mixtures such as plasma/serum or tissue extracts, and "lysate (reverse phase) arrays", in which the complex samples are printed on the surface and targets detected with antibodies [25]. More recently, selected reaction monitoring (or multiple reaction monitoring) mass spectrometry (MRM-MS) has been introduced, in which the mass spectrometer is programmed to analyze and absolutely quantify a preselected group of proteins. Precise and specific quantitation of individual proteins is achieved by incorporating stable isotope labeled standards and without the requirement for antibodies [26].

Improving gene annotation by proteogenomics in a quest for microbial biomarkers

The ability to sequence DNA rapidly, inexpensively, and in a high-throughput fashion provides a unique opportunity to characterize whole genomes of a large number of microbial species, particularly pathogens. To date, an impressive catalogue of genomes has been established, with an additional thousand genomes currently under investigation. It is worth noting that, to realize the full biological value of the sequenced genome, an accurate identification of the protein-coding genes in each genome is required. Accordingly, in the last years part of the scientific and technical challenge has shifted from genome sequencing to genome annotation. Although manual annotation of protein-coding genes is generally considered more reliable, such efforts may not be feasible because of time constraints; therefore, genome annotations of most sequenced genomes are almost exclusively based on predictions [27,28]. However, gene annotation is still far from trivial, whatever the genome under consideration. For instance, in the case of the *Mycoplasma genitalium* genome, inconsistencies in gene predictions among three different groups resulted in 8% error [29]; in the case of *Mycobacterium tuberculosis*, two sequencing centers disagreed by 12% in predicting open-reading frames [30]. Specifically, predicting short genes which are not yet annotated, identifying genes with abnormal codon usage, determining the precise start codons, assessing splicing and alternative splicing in eukaryotes, recognizing programmed frameshifts, and correcting overcalled open reading frames (ORFs) due to the use of multiple gene prediction algorithms are some of the current challenges in genome annotation [31]. In-frame stop codons can be read as sense depending on their context: for instance, TGA and TAG codons may specify selenocysteine and pyrrolysine insertions, respectively. The problem of perpetuating a previous mistake over the next genomes under annotation is still a very important concern. In addition, the description of the gene product itself may be erroneous; therefore, the use of the "hypothetical protein" nomenclature is the current norm to avoid overconfidence in these types of annotations [32].

The term "proteogenomics" refers to the correlation of the proteomic data with the genomic data, with the goal of enhancing the understanding and the annotation of the genome [33]. In a proteomic experiment, tandem mass spectra are generally searched against a protein database composed of in silico predicted peptides; correct matching between theoretical and measured values provides confident protein identification. The idea of searching MS/MS spectra against nucleic acid sequences was first demonstrated by Yates and colleagues [34]. In this approach, the nucleic acid sequence representing the "genomic" database is translated in all six reading frames, and then queried with MS/MS spectra to identify protein-coding genes. Interestingly, this strategy is able to minimize the inherent biases derived from gene prediction methods, thus allowing novel peptide sequences to be identified that were not present in the original protein databases [35,36]. Moreover, using a proteogenomic approach, it is possible to assign correct start sites, as well as to validate the expression of predicted genes (or pseudogenes). As such, proteomics represents a potentially essential tool for integrating protein-level information into the genome annotation process and improving genome annotation quality [31,37].

Although still in its infancy, proteogenomics has been applied to date for improving genome annotation of various pathogenic microbes, including *Plasmodium falciparum* [38], *Toxoplasma gondii* [39], *Leishmania donovani* [40], *Candida glabrata* [41], *Mycoplasma pneumoniae* [33], *Mycobacterium tuberculosis* [30], *Shigella flexneri* [42], and *Yersinia* spp. [43,44]. On the whole, the expression of a high number of sequenced microbial genes has been experimentally validated, several novel protein-coding genes have been discovered for each pathogen, various existing gene models have been corrected (for instance concerning start sites or errors in genome sequencing), and new signal peptides have been identified [36]. These refinements are starting to pave the way to a better understanding of microbial biology and to the identification of novel targets for diagnosis and therapy of infectious diseases. Proteogenomic results may deserve particular attention by those dealing with vaccine development. In fact, accuracy in gene annotation, validation of gene product expression, and identification of novel signal peptide sequences are considered discriminating factors within reverse vaccinology projects, as described below.

Contribution of "omic" technologies to the development of point-of-care (POC) diagnostic tools

While there are many diseases for which a suitable biomarker has not yet been identified, in diseases for which biomarkers are available, these often identify only a subset of patients, or can be measured only in biological samples that require complex procedures for collection and/or analysis. Other biomarkers are dependent on the host response to infection, and are therefore variable and lack sensitivity. Once developed, a diagnostic device has a manyfold impact on reducing the disease burden: 1) it can enable timely detection of the disease, and since patients can be treated earlier, additional costs either in terms of health-care or income losses are not incurred; 2) it can facilitate selection of the most appropriate treatment, increasing the chances of therapy success (*e.g.*, in diseases with significant drug resistance issues); 3) it can assist the monitoring of treatment effectiveness, such as in chronic illnesses and in lengthy therapeutic procedures; 4) it can enable disease monitoring in areas or in patient categories where the incidence and prevalence of a disease need to be kept under constant control, therefore enabling its efficient surveillance.

However, the uncertainty of the time frame spanning from research to discovery to application needs to be taken into consideration, since a thorough validation process is required to integrate the new biomarkers into diagnostic devices for the clinical setting. This step is usually the most time-consuming part of the process, and the one with the highest candidate biomarker "death rate". On the other hand, mass spectrometry-based proteomics can assist also in this translational step; with the advent of targeted approaches (such as the above-described MRM technique [26]), the fast and high-throughput validation of protein and peptide biomarkers on a large scale becomes more feasible. The mass spectrometry-based approach is fast, reliable, quantitative, and highly specific, although its implementation in the diagnostic setting is limited by the elevated investment that is initially required in instrumentation and method set-up. However, once implemented, the MRM approach becomes a cost-effective and efficient method for enabling large-scale validation of a biomarker before investing into its implementation within a diagnostic device.

The proteomic approach, leading to the identification of diagnostic protein/peptides, is of particular interest, since in the infectious disease field the available POC solutions usually consist of rapid diagnostic tests (RDTs) based mostly on antibody-based recognition methods, including agglutination, enzyme immunoassays, and lateral-flow immunochromatography. Usually, these can take the form of dipsticks (with the appearance of symbols or lines), latex agglutination systems (appearance of coagulation), or in-solution systems (color change), provided that complicated instrumentation for performing, reading or interpreting the test is not needed and that the reaction occurs in a short time frame. Other rapid procedures can be based on direct microscopical examination of the sample with simple preparation protocols [45,46,47]. Resource-poor settings have gained considerable benefit from immunochromatography technologies, which form the basis of the dipstick and lateral flow systems. Devices enabling the diagnosis of malaria, HIV, and syphilis have been effectively developed and implemented. Among these, an example of significant success is the case of the lateral flow test for the detection of antibodies to HIV-1 and HIV-2, and the more recent generation of tests for the detection of the p24 capsid protein. In less than 20 minutes, and without requiring complex instrumentation for analysis and interpretation, these tests are able to provide qualitative information on the HIV status of the patient, with sensitivity comparable to that of laboratory-based assays.

Conversely, POC nucleic acid-based testing is lagging significantly behind. This situation is probably dependent on the extreme difficulty in integrating most molecular amplification technologies in POC diagnosis systems, due to staff training, cost, energy requirements, and regulatory compliance [48,49,50]. However, the rapid progress in genomics and proteomics and the significant breakthroughs in innovative technologies are creating significant opportunities for the expansion of this field. Promising technologies that are entering the market or are at the late stages of development include: magnetic beads for enrichment of nucleic acids or antigens; isothermal nucleic acid amplification techniques; nanobiosensors; hand-held microfluidic devices enabling multiplex detection of infectious diseases; and robust and cost-effective optical instruments for the detection of fluorescent signals, enabling an increase in sensitivity of POC tests [51,52,53,54,55,56,57,58]. The list of tests waived by the Clinical Laboratory Improvement Amendments (CLIA) can be found at http://www.cms.hhs.gov/CLIA/downloads/waivetbl.pdf.

Upon rigorous evaluation and validation, correct use, and appropriate regulation, the integration of novel biomarkers identified by means of genomics and proteomics into dedicated POC tests can provide a significant contribution to the fight against infectious diseases also in the resource-poor settings of developing countries.

Reverse vaccinology and the promise for next-generation vaccines

For more than 100 years, since the initial and successful attempts by Louis Pasteur, candidate vaccines have been developed in large part by the classical approach of isolating a bacterial or viral pathogen, inactivating or attenuating its infectious capabilities then administering the vaccine to the individual to be protected. Even when this general approach has failed or has been only partially successful, it has served to guide improvements in vaccine design. The trial-and-error approach has also been performed to obtain vaccine products designed as pooled or single purified antigens chosen among the most relevant according to their role in microbial pathogenicity and to the immunodominance of their specific epitopes. Unfortunately, this classical approach has not been successful in a number of the most dangerous infectious diseases. Yet, deepening/extending our knowledge on the structure and the function of many antigens allowed vaccinologists to identify a number of key features for the ideal antigen. First, the antigen should be located on the surface of the microorganism and it should preferably play a direct role in pathogenesis. It should necessarily be expressed during certain phase(s) of the infection and, specifically, when presenting cells (APCs) may efficiently process it. The antigen should carry CD4 and/or CD8 epitopes to activate humoral and/or cellular specific immune response that might contrast survival and replication according to the biology of the infection. Furthermore, its presence and structure should be highly stable within strains of the same species.

In the late 1990s, reverse vaccinology was proposed by scientist Rino Rappuoli to address all these constraints to attain a reliable and efficient approach that is expected to allow a relentless progression toward novel and efficient vaccines [59]. Reverse vaccinology is based on technologies that allow an outstanding wide availability of information about any possible antigen elaborated by the pathogen of interest. The first powerful tool came from the ability to access the genomes of microorganisms, since 1995, when Craig Venter published the genome of the first free living organism [1]. Noteworthy, the wealth of sequenced genomes have rapidly grown in the past five years thanks to wide access to massive parallel sequencing and, consequently, a decrease in the economic effort required to obtain a fully sequenced genome. A second powerful tool came more recently from the ability to obtain the proteome of microorganisms, thanks to the new technologies available as described in the previous paragraphs.

In brief, reverse vaccinology initiates the search for candidate antigens by sequencing the whole genome of the specific pathogen. However, this is only the first step required to gather a database of candidate ORF that must be thoroughly analyzed by an array of sophisticated biocomputing tools. According to defined signatures occurring within its sequence, each potential gene/protein is evaluated to identify its specific subcellular compartment, category of functions, epitope contents, and other key features that lead to the definition of a limited group of antigens. Pools of peptides, predicted to bind specific common HLA types, can be screened. Alternatively, protein antigens might be chosen according to their peptide contents. Peptides predicted to bind multiple alleles within an HLA supertype might be of special interest, since they provide a wide coverage of all populations without ethnic bias.

As mentioned above, in addition to the genome sequence, the reverse vaccinology approach is also supported by the investigation of the microbial proteome antigenic repertoire. This might be performed using libraries of expressed antigens and screening for the immunogenicity of the proteins expressed during infection [60], or by a proteomic systematic search on bacteria grown in a variety of conditions and stimuli, or even directly recovered from infected tissues. Other approaches to identify antigens with a strong potential as vaccine candidates include the analysis of the bacterial cell surface proteome. This latter technology aims to identify and quantify those antigens that are present on the bacterial surface [61]. Proteins that are exposed to the extracellular milieu are first partially digested by treatment with proteases (*i.e.*, trypsin), and the resulting peptides are then subjected to mass spectrometry analysis.

The discovery of protective antigens has progressed significantly since the application of reverse vaccinology-based technologies. The first reverse vaccinology process to develop protective immunity against serogroup B *Neisseria meningitidis* (MenB) served as a proof-of-concept study, and has led to the first vaccine product that went through clinical trials (4CMenB – commercial name Bexero, Novartis, Italy). Whole sequence and bioinformatic analysis of the MenB genome initially enabled the selection of nearly 600 antigens as vaccine candidates. Most of them were successfully expressed in *E. coli* and 28 recombinant proteins that were found to induce bactericidal antibodies in immunized mice. A pool of four of these proteins, each with a specific pathogenetic role, was used to formulate the 4 Component MenB vaccine [62].

Genome and proteome-based vaccines have also been designed to induce protection against pathogenic species when a high level of diversity occurs among circulating epidemic strains. A combination of pan-genome comparative analysis of circulating strains (n = 8) has been applied to the Gram-positive pathogen *Streptococcus agalactiae*. Starting from a core genome content of 1,806 genes, researchers predicted 396 surface-exposed proteins. This study, in addition to extending the knowledge of adhesion structures elaborated by *S. agalactiae*, identified four proteins that induce a protective immune response in a mouse model of maternal-neonatal pup immunization [63]. Interestingly, three of these proteins were found to assemble into pili. Based on these pioneering findings, a search for molecular signatures for typical pilus regions was also applied by Mora *et al*. [64] in a comparative reverse vaccinology study aimed to identify *S. pyogenes* vaccine candidates. Genome sequence analysis of five different circulating strains suggested that a multicomponent vaccine combining 12 backbone variants of pilus protein genes might provide protection against over 90% of circulating strains of *S. pyogenes*.

The quality of these approaches has been more recently confirmed in similar studies aimed to develop new vaccines for *S. pneumoniae* and extra-intestinal pathogenic *E. coli* [65,66].

Conclusion and perspectives

The power of high-throughput “omic” technologies can provide a significant contribution to the discovery of biomarker candidates for the development of efficient vaccines and cost-effective diagnostic systems. The populations of developing countries are exposed to a wide range of diseases that are almost nonexistent in developed countries, such as malaria and African sleeping sickness. Others, such as HIV/AIDS and tuberculosis, although also limitedly affecting developed countries, pose a tremendously higher burden on these populations. In the fight against these diseases, reverse vaccinology is providing the possibility of identifying pools of peptides that are predicted to bind multiple alleles within an HLA supertype, enabling a wide coverage in all populations without ethnic bias.

Currently, most of the investments are dedicated to development of vaccines and new therapies, while the research on more efficient and effective diagnosis tools is somehow lagging behind. Indeed, despite their incidence and the life toll taken, there are still significant unmet needs in terms of molecules and methods enabling a better diagnosis. Highly impacting infectious diseases, such as Chagas disease, cholera, dengue, lymphatic filariasis, schistosomiasis, acute lower respiratory tract infections, and diarrheal diseases, prompt for new and improved diagnosis systems, with the potential of saving thousands to millions of lives [67]. To provide the most influential benefits, diagnostic tests must meet the requirement to be reliably used in the most rural regions. It might be perceived that a POC test should somehow be less effective or efficient than a laboratory-based test in fighting infectious diseases. As a matter of fact, in some cases the overall treatment rate can be higher when using POC devices. This has been defined as the rapid test “paradox” [68]. In fact, the shorter turnaround time to diagnosis (that is, time from sample collection to results) enables test and treatment in the same encounter, enabling a faster treatment and reducing patient loss to follow-up. Furthermore, a POC test includes only three steps: sample collection, on-site testing, and treatment, while a laboratory test requires at least 6 steps: sample collection, sample shipment, notification of results, shipment of results to physician, second patient encounter, and eventually treatment. The higher "need for speed" typical of infectious diseases as opposed to other pathologies amplifies the benefits of POC testing.

Further perspectives have been recently opened by a novel science named "microbiomics", lying at the interface between microbiology and omic science. Microbiomics can be defined as the application of genomic, transcriptomic, and/or proteomic approaches to the investigation of microbial communities and their genetic potential and expression (*i.e.*, microbiome). Most of the microbes colonizing human, animal, or environmental sites are in fact not amenable to cultivation, and are therefore less studied, if not completely unknown. In this respect, the impressive depth and sensitivity of analysis which can be reached by exploiting omic techniques may considerably help in knowing the composition and the functional properties of microbial communities, whose changes are known to be involved in, or even causing, the development of several diseases. In this context, it is worth highlighting the recent launch of the Human Microbiome Project, which is aimed at screening the main human body sites through a high-throughput metagenomic approach [69,70]. Shotgun sequencing of complex microbial populations can provide, on the one hand, precise sequence information about core genes, that is, genes conserved in all species and evidently essential to live in that environment; on the other hand, the taxonomic composition of a given microbiome can be thoroughly investigated, primarily by sequencing the 16S ribosomal RNA gene. To date, several approaches are available for sequencing a microbiome with next-generation platforms for its analysis via bioinformatic pipelines [71]. Metagenomic and metaproteomic projects are expected to elucidate the abundance and variety of the human microbiome, thus providing an exceptional framework for future research developments concerning, for instance, the impact of the gut microbiome composition in malnutrition, as well as the mechanisms underlying the shift from microflora imbalance to intestinal, vaginal, or urinary tract infections, or even the relationship between the structure of the microbial communities living in the human body and the development of immunological disorders.

References

1. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton S, FitzHugh W, Fields C, Gocayne JD, Scott J, Shirley R, Liu L-I, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Hedblom E, Cotton MD, Teresa R. Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Craig Venter JC (1995) An approach for genome analysis based on sequencing and assembly Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496-512.
2. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol 10: 599-606.
3. Nagarajan N, Cook C, Di Bonaventura M, Ge H, Richards A, Bishop-Lilly KA, DeSalle R, Read TD, Pop M (2010) Finishing genomes with limited resources: lessons from an ensemble of microbial genomes. BMC Genomics 11: 242.
4. Richardson EJ and Watson M (2012) The automatic annotation of bacterial genomes. Brief Bioinform.
5. Lottspeich F (2009) Introduction to proteomics. Methods Mol Biol 564: 3-10.
6. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I (1995) Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis 16: 1090-1094.
7. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev 13: 19-50.
8. Ahrens CH, Brunner E, Qeli E, Basler K, Aebersold R (2010) Generating and navigating proteome maps using mass spectrometry. Nat Rev Mol Cell Biol 11: 789-801.
9. Westermeier R, Naven T, Höpker HR (2008) Proteomics in Practice: A Guide to Successful Experimental Design: Westheim: Wiley-VCH 482p.
10. MacBeath G (2002) Protein microarrays and proteomics. Nat Genet 32 Suppl: 526-532.
11. Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26: 231-243.
12. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250: 4007-4021.
13. Görg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. Proteomics 4: 3665-3685.
14. Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18: 2071-2077.
15. Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I (2003) A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics 3: 36-44.

16. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd (1999) Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol 17: 676-682.
17. Washburn MP, Wolters D, Yates JR, 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19: 242-247.
18. Lambert JP, Ethier M, Smith JC, Figeys D (2005) Proteomics: from gel based to gel free. Anal Chem 77: 3771-3787.
19. Schirle M, Heurtier MA, Kuster B (2003) Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics 2: 1297-1305.
20. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17: 994-999.
21. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3: 1154-1169.
22. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1: 376-386.
23. Stemmann O, Zou H, Gerber SA, Gygi SP, Kirschner MW (2001) Dual inhibition of sister chromatid separation at metaphase. Cell 107: 715-726.
24. Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA (2011) Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics 11: 535-553.
25. LaBaer J and Ramachandran N (2005) Protein microarrays as tools for functional proteomics. Curr Opin Chem Biol 9: 14-19.
26. Gallien S, Duriez E, Domon B (2011) Selected reaction monitoring applied to proteomics. J Mass Spectrom 46: 298-312.
27. Frishman D (2007) Protein annotation at genomic scale: the current status. Chem Rev 107: 3448-3466.
28. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC (2010) The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 38: D346-354.
29. Brenner SE (1999) Errors in genome annotation. Trends Genet 15: 132-133.
30. de Souza GA, Malen H, Softeland T, Saelensminde G, Prasad S, Jonassen I, Wiker HG (2008) High accuracy mass spectrometry analysis as a tool to verify and improve gene annotation using *Mycobacterium tuberculosis* as an example. BMC Genomics 9: 316.
31. Ansong C, Purvine SO, Adkins JN, Lipton MS, Smith RD (2008) Proteogenomics: needs and roles to be filled by proteomics in genome annotation. Brief Funct Genomic Proteomic 7: 50-62.
32. Armengaud J (2009) A perfect genome annotation is within reach with the proteomics and genomics alliance. Current Opinion in Microbiology 12: 292-300.
33. Jaffe JD, Berg HC, Church GM (2004) Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics 4: 59-77.
34. Yates JR, 3rd, Eng JK, McCormack AL (1995) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal Chem 67: 3202-3210.
35. Mann M and Pandey A (2001) Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. Trends Biochem Sci 26: 54-61.
36. Renuse S, Chaerkady R, Pandey A (2011) Proteogenomics. Proteomics 11: 620-630.
37. Gupta N, Tanner S, Jaitly N, Adkins JN, Lipton M, Edwards R, Romine M, Osterman A, Bafna V, Smith RD, Pevzner PA (2007) Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. Genome Res 17: 1362-1377.
38. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW, Eling WM, Hall N, Waters AP, Stunnenberg HG, Mann M (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature 419: 537-542.
39. Xia D, Sanderson SJ, Jones AR, Prieto JH, Yates JR, Bromley E, Tomley FM, Lal K, Sinden RE, Brunk BP, Roos DS, Wastling JM (2008) The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. Genome Biol 9: R116.
40. Pawar H, Sahasrabuddhe NA, Renuse S, Keerthikumar S, Sharma J, Kumar GS, Venugopal A, Sekhar NR, Kelkar DS, Nemade H, Khobragade SN, Muthusamy B, Kandasamy K, Harsha HC, Chaerkady R, Patole MS, Pandey A (2012) A proteogenomic approach to map the proteome of an unsequenced pathogen - *Leishmania donovani*. Proteomics 12: 832-844.
41. Prasad TS, Harsha HC, Keerthikumar S, Sekhar NR, Selvan LD, Kumar P, Pinto SM, Muthusamy B, Subbannayya Y, Renuse S, Chaerkady R, Mathur PP, Ravikumar R, Pandey A (2012) Proteogenomic analysis of *Candida glabrata* using high resolution mass spectrometry. J Proteome Res 11: 247-260.
42. Zhao L, Liu L, Leng W, Wei C, Jin Q (2011) A proteogenomic analysis of *Shigella flexneri* using 2D LC-MALDI TOF/TOF. BMC Genomics 12: 528.
43. Payne SH, Huang ST, Pieper R (2010) A proteogenomic update to *Yersinia*: enhancing genome annotation. BMC Genomics 11: 460.
44. Schrimpe-Rutledge AC, Jones MB, Chauhan S, Purvine SO, Sanford JA, Monroe ME, Brewer HM, Payne SH, Ansong C, Frank BC, Smith RD, Peterson SN, Motin VL, Adkins JN (2012) Comparative omics-driven genome annotation refinement: application across *Yersiniae*. PLoS ONE 7: e33903.
45. Nichols JH (2007) Point of care testing. Clin Lab Med 27: 893-908, viii.
46. Trevino EA and Weissfeld AS (2007) The case for point-of-care testing in infectious-disease diagnosis. Clinical Microbiology Newsletter 29: 177-179.
47. Overturf GD (2008) CLIA waived testing in infectious diseases. Pediatr Infect Dis J 27: 1009-1012.
48. Bissonnette L and Bergeron MG (2010) Diagnosing infections--current and anticipated technologies for point-of-care diagnostics and home-based testing. Clin Microbiol Infect 16: 1044-1053.

49. Weile J and Knabbe C (2009) Current applications and future trends of molecular diagnostics in clinical bacteriology. Anal Bioanal Chem 394: 731-742.
50. Nougairede A, Ninove L, Zandotti C, De Lamballerie X, Gazin C, Drancourt M, La Scola B, Raoult D, Charrel RN (2009) Point of care strategy for rapid diagnosis of novel A/H1N1 influenza virus. PLoS One 5:e9215.
51. Peeling RW and Mabey D (2010) Point-of-care tests for diagnosing infections in the developing world. Clin Microbiol Infect 16: 1062-1069.
52. Nargessi D and Ou CY (2010) MagaZorb: a simple tool for rapid isolation of viral nucleic acids. J Infect Dis 201 Suppl 1: S37-41.
53. Piepenburg O, Williams CH, Stemple DL, Armes NA (2006) DNA detection using recombination proteins. PLoS Biol 4: e204.
54. Zelada-Guillen GA, Riu J, Duzgun A, Rius FX (2009) Immediate detection of living bacteria at ultralow concentrations using a carbon nanotube based potentiometric aptasensor. Angew Chem Int Ed Engl 48: 7334-7337.
55. Rodriguez WR, Christodoulides N, Floriano PN, Graham S, Mohanty S, Dixon M, Hsiang M, Peter T, Zavahir S, Thior I, Romanovicz D, Bernard B, Goodey AP, Walker BD, McDevitt JT (2005) A microchip CD4 counting method for HIV monitoring in resource-poor settings. PLoS Med 2: e182.
56. Sia SK, Linder V, Parviz BA, Siegel A, Whitesides GM (2004) An integrated approach to a portable and low-cost immunoassay for resource-poor settings. Angew Chem Int Ed Engl 43: 498-502.
57. Sorger PK (2008) Microfluidics closes in on point-of-care assays. Nat Biotechnol 26: 1345-1346.
58. Obeid PJ and Christopoulos TK (2003) Continuous-flow DNA and RNA amplification chip combined with laser-induced fluorescence detection. Analytica Chimica Acta 494: 1-9.
59. Rappuoli R (2000) Reverse vaccinology. Curr Opin Microbiol 3: 445-450.
60. Giefing C, Meinke AL, Hanner M, Henics T, Bui MD, Gelbmann D, Lundberg U, Senn BM, Schunn M, Habel A, Henriques-Normark B, Ortqvist A, Kalin M, von Gabain A, Nagy E (2008) Discovery of a novel class of highly conserved vaccine antigens using genomic scale antigenic fingerprinting of pneumococcus with human antibodies. J Exp Med 205: 117-131.
61. Rodriguez-Ortega MJ, Norais N, Bensi G, Liberatori S, Capo S, Mora M, Scarselli M, Doro F, Ferrari G, Garaguso I, Maggi T, Neumann A, Covre A, Telford JL, Grandi G (2006) Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. Nat Biotechnol 24: 191-197.
62. Giuliani MM, Adu-Bobie J, Comanducci M, Aricò B, Savino S, Santini L, Brunelli B, Bambini S, Biolchi A, Capecchi B, Cartocci E, Ciucchi L, Di Marcello F, Ferlicca F, Galli B, Luzzi E, Masignani V, Serruto D, Veggi D, Contorni M, Morandi M, Bartalesi A, Cinotti V, Mannucci D, Titta F, Ovidi E, Welsch JA, Granoff D, Rappuoli R, Pizza M (2006) A universal vaccine for serogroup B meningococcus. Proc Natl Acad Sci USA 103: 10834-10839.
63. Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Nardi Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G (2005) Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science 309: 148-150.
64. Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AG, Maggi T, Taddei AR, Grandi G, Telford JL (2005) Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc Natl Acad Sci U S A 102: 15641-15646.
65. Gianfaldoni C, Censini S, Hilleringmann M, Moschioni M, Facciotti C, Pansegrau W, Masignani V, Covacci A, Rappuoli R, Barocchi MA, Ruggiero P (2007) *Streptococcus pneumoniae* pilus subunits protect mice against lethal challenge. Infect Immun 75: 1059-1062.
66. Moriel DG, Bertoldi I, Spagnuolo A, Marchi S, Rosini R, Nesta B, Pastorello I, Corea VA, Torricelli G, Cartocci E, Savino S, Scarselli M, Dobrindt U, Hacker J, Tettelin H, Tallon LJ, Sullivan S, Wieler LH, Ewers C, Pickard D, Dougan G, Fontana MR, Rappuoli R, Pizza M, Serino L (2010) Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. Proc Natl Acad Sci U S A 107: 9072-9077.
67. Mehta P, Cook D (2010) The Diagnostic Innovation Map: Medical Diagnostics for the Unmet Needs of the Developing World - Report. BIO Ventures for Global Health. Available: www.bvgh.org/LinkClick.aspx?fileticket=-a1C6u2LE4w%3d&tabid=91
68. Gift TL, Pate MS, Hook EW, 3rd, Kassler WJ (1999) The rapid test paradox: when fewer cases detected lead to more cases treated: a decision analysis of tests for *Chlamydia trachomatis*. Sex Transm Dis 26: 232-240.
69. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, Nelson KE, White O, Methé BA, Huttenhower C (2012) The human microbiome project: a community resource for the healthy human microbiome. PLoS Biol 10: e1001377.
70. Human Microbiome Project Consortium (2012) A framework for human microbiome research. Nature 486: 215-221.
71. Davenport CF and Tummler B (2013) Advances in computational analysis of metagenome sequences. Environ Microbiol 15:1-5.

Corresponding author

Prof. Sergio Uzzau, MD, PhD
Dipartimento di Scienze Biomediche
Università di Sassari
V.le S. Pietro 43/B
07100 Sassari, Italy
Telephone: +39079228303
Email: uzzau@uniss.it

Conflict of interests: No conflict of interests is declared.