

The Application of Metagenomic Approaches in the Management of Infectious Diseases

Jinhua Wu¹, Kaiye Cai¹ and Qiang Feng^{1,2*}

¹Beijing Genomics Institute- Shenzhen Main Building, Beishan Industrial Zone, Yantian District, Shenzhen, China

²Shenzhen Engineering Laboratory of Detection and Intervention of Human Intestinal Microbiome, Shenzhen, China

*Corresponding author: Qiang Feng, Beijing Genomics Institute-Shenzhen Main Building, Beishan Industrial Zone, Yantian District, Shenzhen, 518083, China; Tel: +86-755-25273620; Fax: +86-755-25273620; E-mail: fengqiang@genomics.cn

Received date: July 04, 2015, Accepted date: September 01, 2015, Published date: September 04, 2015

Copyright: © 2015 Wu J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Abstract

Over the past decades, although billions of dollars has been invested and much progress has been made in the prevention and surveillance of infectious diseases, the mortality caused by these diseases is still hovering at high level. Infectious diseases are constantly threatening lives of millions of people around the world. For the diagnosis of these diseases, traditional culture-dependent methods are playing vital roles in clinical practice, but they still cannot meet the requirement of rapid detection of both known and unknown pathogens in a high-throughput mode. The emergence of next-generation sequencing (NGS) technology has greatly promoted the development of genome research, and the NGS-based metagenomics is gaining more and more attention as a potential technique for the management of infectious diseases. This review gives a brief summarization of the application of metagenomics in infectious diseases in recent years.

Keywords: Metagenomics; Infectious diseases; Pathogen detection

Introduction

Infectious diseases are defined by World Health Organization as diseases that are caused by viruses, bacteria, parasites or fungi. Common infectious diseases include HIV/AIDS, hepatitis B and C, diarrheal diseases, respiratory infections, tuberculosis (TB), pneumonia, condylomas etc. These diseases can be spread directly or indirectly from one person to another, and in some cases the number of infected people may increase enormously in a very short time. Despite the fast-pacing development of modern medicine, the incidence and mortality for infectious diseases remains high in the past few decades. Each year, infectious diseases have taken millions of lives around the world. According to the Global Burden of Disease Study 2013, the number of deaths caused by HIV, tuberculosis, and malaria was nearly 3.6 million (from 3.3 million to 4.0 million) in 2013 alone [1].

For the management of infectious diseases, one important step is to find out the potential pathogens (if possible). Traditional methods for identifying pathogens include microscopy, culture, serology, as well as antigen detection [2,3]. In recent years, the progress of next-generation sequencing technologies has greatly prompted the development of metagenomics (genomic methods). Metagenomics refers to the genomic analysis of microbial communities from their natural environment. While the concept of “metagenomics” first appeared early in 1998 [4], it was only in the past few years that metagenomics has gained attention as an approach for pathogen detection. In 2008, by the method of metagenomics sequencing, Nakamura et al. directly detected *Campylobacter jejuni* as the bacterial pathogen in a diarrhea patient [5]. Since then, metagenomic approaches are drawing much more attention from clinical researchers. With the advantage of being culture-independent, high-

throughput and fast-in-speed, metagenomics is now playing an increasingly role in the diagnosis of infectious diseases.

This review firstly gives a brief introduction to the metagenomic techniques that are frequently employed. We then summarize the recent progresses of applying metagenomics in management of infectious diseases, with an emphasis on viral infectious diseases such as HIV/AIDS, hepatitis B and C, the yearly panic-causing influenza; bacterial diseases such as tuberculosis (TB), bacterial diarrheal diseases, as well as some other infectious diseases caused by fungi or parasites.

Metagenomic Approaches for Detecting Pathogens of Infectious Diseases

While new sequencing technologies are constantly emerging, two commonly used metagenomic approaches are deep amplicon sequencing and whole-genome shotgun metagenomics [6-8]. Generally, both of these two metagenomic-based sequencing approaches involves laboratory pipeline and bioinformatics pipeline [8] (Figure 1). In the laboratory pipeline, firstly, nucleic acid is extracted from samples, and then sequencing libraries are prepared. In the case of whole-genome shotgun metagenomics, all nucleic acids (including viral RNA/DNA) were extracted and amplified; the amplification step can be skipped only if the PCR-free library construction is possible. Whereas in the case of deep amplicon sequencing, targeted variable regions of the conserved genes (16S rRNA genes from bacteria, 18S rRNA genes from eukaryotes etc.) were amplified from the extracted nucleic acids. When the sequencing libraries were constructed, the bioinformatics pipeline is outlined as follows: First, DNA sequencing is performed on high-throughput sequencing platforms (Illumina, Hiseq, CG, Roche 454 and Ion proton). The obtained raw reads go through quality check and filter in order to remove adapters, low quality reads, and the host's reads. Next, the filtered reads are aligned against known reference databases

through various alignment tools (BLAST, BWA, Bowtie2, SOAP, MG-RAST etc.), so that known bacterial and viral reads are identified with annotations. The remaining unmapped reads can be *de novo* assembled and the contigs can be classified by using different algorithms. In case when reads are too short, the assembly work would be done first. Gene prediction and taxonomic/functional analysis will be conducted on the assembled sequences, thus new microorganism is possibly to be detected. Finally, laboratory methods such as PCR, isolation and culture are performed to confirm the identified

pathogens. From above, we can see that the advantages of metagenomics are culture-independent, high-throughput, rapid, feasible, specific, and more importantly, it can be used to detect unknown microorganisms. However, the disadvantage of metagenomics method is that the sequencing cost remains high and the process of informatics analysis is complex. Moreover, metagenomics has a relatively low sensitivity and it needs further confirmation for its identification results.

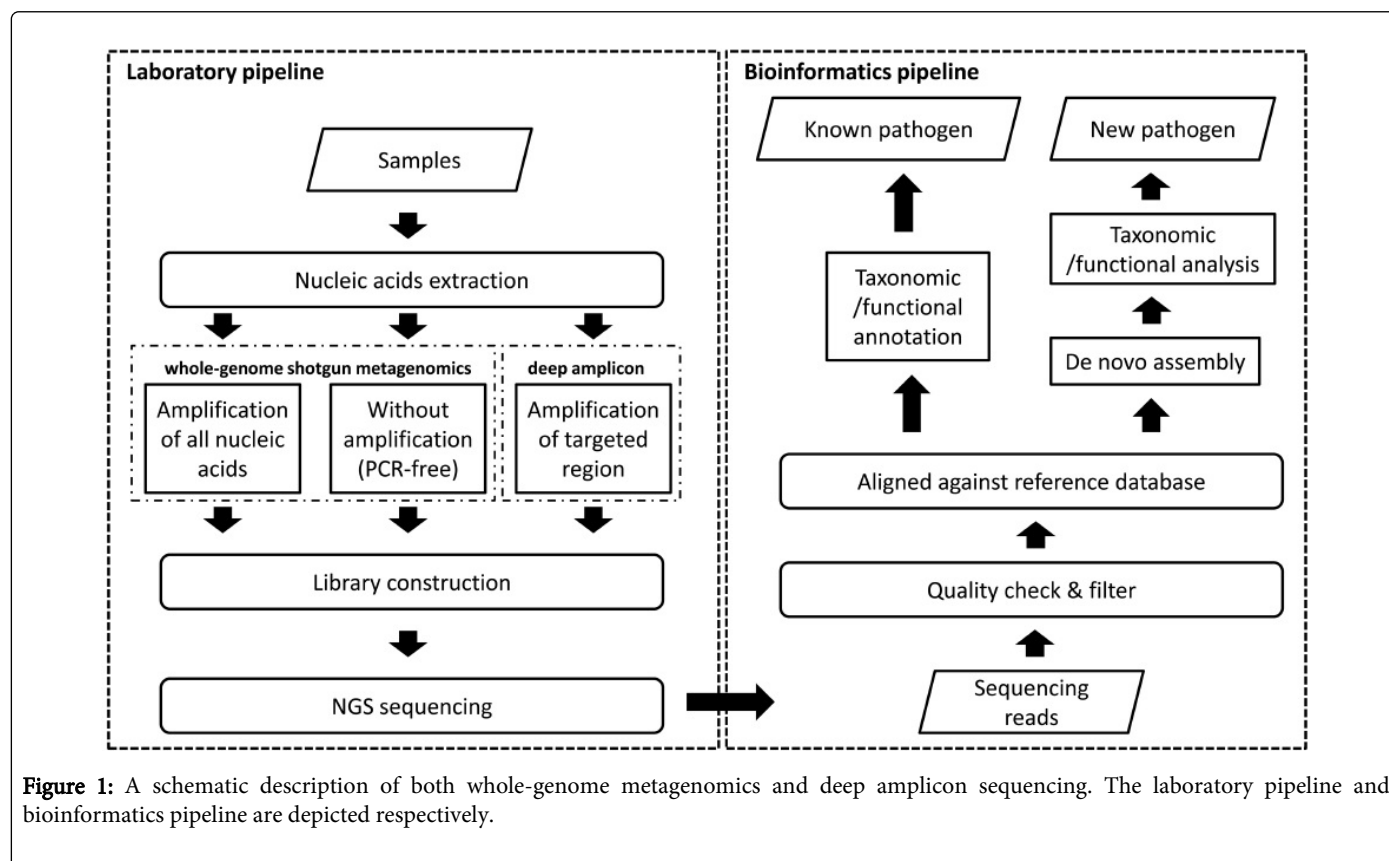


Figure 1: A schematic description of both whole-genome metagenomics and deep amplicon sequencing. The laboratory pipeline and bioinformatics pipeline are depicted respectively.

The Application of Metagenomics in the Management of Infectious Diseases

Viral infectious disease

AIDS: AIDS is a deadly infectious disease, causing 1.3 million deaths worldwide in 2013 alone [1]. AIDS is caused by human immunodeficiency virus (HIV). Using next generation sequencing method, Li et al. compared the plasma bacterial and viral elements in HIV/AIDS patients with that of healthy controls. It was shown that while both the HIV/AIDS and normal plasma DNA virome shared some common eukaryotic viruses, the former is more abundant in bacteriophages whereas the latter mainly contained viruses from Anelloviridae. Additionally, it was also found that the bacterial elements in HIV/AIDS plasma resembled that of human gut microbes [9]. By using 16S rRNA gene-based pyrosequencing and quantitative PCR, Liu et al. compared the composition of semen microbiome in samples from 22 HIV-uninfected men and 27 HIV-infected men, respectively. It was found that the semen microbiome of HIV-infected men decreased in diversity and richness as compared with that of HIV-uninfected men. Since semen serves as an important vector for

HIV transmission, this suggests that semen microbiome may play a role in sexual transmission process of HIV [10]. Altogether, the above findings indicate that metagenomic methodology may help elucidate the transmission mechanism of HIV and improve future management of AIDS [9,10].

Hepatitis B: It is a very common infectious disease in the world, with the total number of infected people reaches approximately 240 million [11]. Hepatitis B is caused by hepatitis B virus, which is characterized by a high rate of mutation and numerous virus variants. Sometimes, current detecting methods such as direct PCR sequencing, clonal sequencing, and point mutation assays could fail because of the high variability of HBV genome. By using ultra-deep pyrosequencing (UDPS), Margeridon-Thermet et al. were able to detect low-prevalence HBV variants at a high level of sensitivity [12]. In this study, plasma samples were collected from HBV patients treated with nucleoside and nucleotide reverse-transcriptase inhibitors (NRTIs) and from NRTI-naïve patients. Then DNA was extracted, and direct PCR sequencing and UDPS were performed. Sequence analysis showed that drug-resistance mutations were detected by UDPS in NRTI-treated patients, whereas direct PCR failed to detect those mutations. This suggests that UDPS could be employed to detect more

HBV variants and thus enhance management of the disease in the future.

Hepatitis C: It is another common viral hepatitis that is caused by hepatitis C virus, affecting nearly 180 million people around the world [13]. HCV is a single-stranded, positive-sense RNA virus. Like many other RNA viruses, HCV also shows a high degree of genetic variability, which brings much difficulty to identification and development of vaccines [13,14]. By employing the method of Illumina deep sequencing, Ninomiya et al. successfully differentiated variants of hepatitis C virus. In this study, total RNA was extracted from serum samples from two hepatitis C patients and one healthy control. Then deep sequencing was conducted and the yielded sequences were further analyzed. Their result showed that both major and minor HCV variants were detectable, and that amino substitutions could be determined, which suggest that deep sequencing technology could be applied to study the HCV quasispecies in a serum sample [14].

Viral diarrheal diseases: Diarrheal diseases, the worldwide incidence of which is estimated to be about 1.7 billion in 2010, remain to be a leading cause of mortality for children under 5 years old [15]. Diarrheal diseases could be caused by viruses, bacteria and protozoa, but it is estimated that the etiologic agent of nearly 40% cases cannot be determined. To characterize the viral diversity in human diarrhea, Finkbeiner et al. analyzed the viral communities in the fecal samples of 12 diarrhea children by metagenomic approaches [16]. A strategy of what they called "micro-mass sequencing" was employed, the pipeline of which can be briefly described as follows: First, minimal fecal samples were collected, and then nucleic acids were extracted and PCR amplified from each sample. Sample libraries were constructed and then sequenced, generating 2, 013 qualified and unique sequences in total. The sequences are then analyzed by the method of tBLASTx (E-value $\leq 10^{-5}$) alignments. Through this method, they not only identified sequences from known enteric viruses, but also detected a number of novel viruses [16]. Although the detected new viruses may not necessarily be the causative agent for the disease, the findings suggest that metagenomics can serve as a powerful tool for identification of novel viruses.

Influenza: Influenza is an acute viral infection that spreads easily and occurs globally. Each year, influenza attacks nearly 5%–10% of adults and 20%–30% of children around the world. Influenza is caused by influenza viruses, which is characteristic of highly antigenic shift and drift and can be classified into three types-influenza virus A, B and C [17]. In this review, only the famous panic-causing H1N1 influenza A is concerned. The H1N1 influenza A broke out in 2009 and had caused more than 50 million infections and 18,449 deaths by August, 2010 [18]. While the first patient was sent to hospital at the end of March, conventional point-of-care diagnostics failed in detecting pathogenic viruses. It was about two weeks later that the novel swine-origin influenza A (H1N1) virus(S-OIV) was identified by real-time RT-PCR [19]. However, by using of PCR primers, RT-PCR may have low sensitivity in discriminating the 2009 pandemic H1N1 influenza A virus from seasonal H3N1 or H1N1 virus. To identify and characterize the 2009 H1N1 virus in a more sensitive and specific way, Greninger et al. employed two metagenomics-based methods-one is pan-viral microarray (Virochip) and the other is deep sequencing. Result from Virochip showed a novel swine influenza virus that differentiated from seasonal H3N1 or H1N1 virus was detected, whereas deep sequencing detected reads of the novel 2009 H1N1 virus in all 17 clinical samples [18]. Other representative examples of applying metagenomics in

influenza also include: Na kamura et al. used an unbiased high-throughput sequencing approach to detect influenza virus (Flu) in nasopharyngeal samples [20].

Condylomas: Condylomas is a sexually-transmitted disease that is caused human papillomavirus (HPV). Johansson et al. were able to detect both known and novel putative HPV types in "HPV-negative" condylomas using metagenomic sequencing [21]. In this study, forty swab samples were collected from the seemingly "HPV-negative" condylomas, then the samples were pooled randomly into ten pools and were subjected to metagenomic sequencing. Results showed that 12 different HPV types or putative types were detected in those samples. This suggests that metagenomics may play important roles in detecting novel HPVs.

Bacterial infectious diseases

Tuberculosis (TB): TB is a widespread infectious disease that mainly affects the lungs and occasionally other sites of our body. TB ranks second in terms of global mortality. It was reported that in 2012, the worldwide tuberculosis incidence was 8, 6 million, and the deaths number was 1.3 million [22]. TB is caused by *Mycobacterium tuberculosis* and other closely associated species in the *M. tuberculosis* complex. Traditional methods for diagnosing TB mainly include sputum smear microscopy and culture-based method. Since microscopy cannot identify pathogen in species- or lineage-level and the culture method has disadvantages such as complex and time-consuming, Doughty et al. employed shotgun-metagenomics to detect and characterize the strains in the *M. tuberculosis* complex [23]. Initially, eight smear- and culture-positive sputum samples were chosen, then DNA was extracted and libraries were constructed respectively. The constructed libraries were then sequenced. Results from large sequence polymorphisms and single nucleotide polymorphisms (SNPs) analysis showed that seven of eight metagenome-derived genomes could be assigned to a species and lineage within the *M. tuberculosis* complex. In a previous study, Chan et al. also applied metagenomic method to analyze DNA sample which was extracted from the lung tissue of an 18th mummy [24]. It was found that the person, who was diagnosed with tuberculosis, was actually infected by two *M. tuberculosis* genotypes. Altogether, the above findings suggest that metagenomics can be applied both to freshly collected samples and to historical samples.

Bacterial diarrheal diseases: Nakamura et al. reported one case of employing metagenomics in the diagnosis of bacterial diarrhea. Firstly, traditional culture methods and specific reverse transcriptase-PCR were used to analyze two different samples from the same person, but failed to detect any potential pathogens. Then DNA was extracted from these two samples, referred to as ill DNA sample and recovery DNA sample, and high-throughput DNA sequencing were performed on both DNA samples. It was shown that sequences from *Campylobacter jejuni* were found uniquely in illness DNA sample. The presence of *C. jejuni* was further confirmed by culture and biochemical analysis. These findings suggest that bacterial pathogens can be directly detected by high-throughput DNA sequencing (metagenomic sequencing) [5]. Another example that needs to be mentioned is the application of metagenomics in the research of Shiga-toxicogenic *Escherichia coli* (STEC) O104:H4, which hit Germany as well as many other countries in Europe in 2011. Through metagenomics method, Loman et al. successfully recovered a draft genome sequence of the German STEC strain, and sequences from other potential pathogens as well. This indicates that during an

outbreak of diarrheal disease, metagenomics may serve as a powerful tool in detecting related bacterial pathogens [25].

Infectious diseases caused by fungi or parasites

Chlamydia trachomatis: *Chlamydia trachomatis* is an obligate pathogen which causes infection to the eyes or genital tracts. By using ultra-deep Illumina sequencing, Andersson et al. were able to obtain sequences of a *Chlamydia trachomatis* genotype and other bacterial genomes directly from a diagnostic specimen. In this study, total nucleic acid was firstly purified from a *C. trachomatis*-positive vaginal swab. Next, the extracted nucleic acid was whole-genome amplified and then paired-end sequencing was performed. The consequent sequence analysis showed that a great number of SNPs from *C. trachomatis* were identified and that sequences from *Prevotella melaninogenica*, *Gardnerella vaginalis*, *Clostridiales genomosp* were also present abundantly in the sample [26]. Altogether, this study shows that metagenomic approaches has the potential of producing explicit sequences data of the microbial community in a diagnostic sample.

Malassezia yeasts: *Malassezia* yeasts are parasitic fungi that may cause opportunistic skin infections in human. Conventional methods for identification of *Malassezia* species mainly include morphological observation and biochemical reactions, the results of which are prone to be affected by many factors. To circumvent these limitations and to explore the applicability and plausibility of pyrosequencing method for analyzing the *Malassezia* species, Kim et al. conducted pyrosequencing on samples collected from different body sites [27]. Results from sequences analysis showed that pyrosequencing may serve as a rapid and effective method for diagnosis of *Malassezia* yeasts-associated infections.

Pneumonia: It is a leading cause of death in children, causes 15% of all deaths of children under 5 years old [28]. It was estimated that nearly 935,000 children under age five died of this disease in 2013. Pneumonia can be caused by various infectious agents, such as viruses, bacteria and fungi. In one case, when all diagnostic tests failed to detect any commonly known causative pathogens of pneumonia in a police officer who was observed of acute respiratory distress syndrome (ARDS), next generation sequencing (NGS) was performed to analyze bronchoalveolar lavage (BAL) samples from the patient [29]. The whole workflow could be briefly described as follows (www.virus-genomics.org/supplementaries/EID1406.pdf): First, unbiased sequencing of total RNA and DNA were performed. Second, 16S rRNA gene amplicon sequencing was performed and the produced sequences data were analyzed. Finally, *Chlamydophila psittaci* infection was confirmed by quantitative real-time PCR (qPCR). The above procedures were completed in less than 50 hours. Sequence analysis result identified *Chlamydophila psittaci* to be the potential disease-causing agent, which is so rare that is normally excluded in conventional diagnostic panels. This strongly suggests that metagenomics has great potential in rapid diagnostics of diseases outbreaks.

Conclusions and Future Perspectives

In general, both whole-genome shotgun metagenomic and deep amplicon sequencing could be applied to detect any all potential pathogens (including viruses, bacteria, fungi or parasites) in a single assay, but they may be adopted separately or in combination for different purposes in clinical practice. Compared to shotgun

metagenomics, deep amplicon sequencing yields shorter reads, and thus is cost-less, time-saving, but lower in accuracy. On the other hand, much information could be inferred from the obtained whole-genome sequences of each pathogen, such as drug resistance, virulence, origin and evolution, transmission et al. [8]. Thus, both metagenomic approaches have been widely applied in diagnosis and management of infectious diseases. However, the bioinformatic pipeline of metagenomics remains a big challenge: First, the cost of sequencing is high; second, massive sequencing data are produced and multiple alignment/classification algorithms are required to analyze these data; third, identification of pathogens by computational analysis still takes several days and does not establish a causation of the disease. In recent years, efforts have been made to deal with these problems: The third-generation sequencing technology, single-molecule real-time (SMRT) was developed to provide a comparatively long read length with median accuracy [30,31]. A cloud-compatible bioinformatics pipeline called "sequence-based ultrarapid pathogen identification" (SURPI) was developed [32]. Compared to blast alignment, SNAP and RAPSearch were employed in SURPI. The new computational pipeline SURPI was so effective that it took only a few hours to detect viruses and bacteria. Other improved analysis tools like SOAP, Bowtie2, BWA, etc., which are specifically targeted for alignment of NGS short reads, are constantly emerging. With the development of sequencing technology and the improvement of alignment algorithms, metagenomic approaches will continue to increase in speed, decrease in cost and become friendly in using. Thus, it is expected that metagenomics will play a more significant role in the management of infectious diseases in the future.

Acknowledgement

This work was supported by grant DRC-SZ [2015]162.

References

1. Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, et al. (2014) Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 384: 1005-1070.
2. Houpijian P, Raoult D (2002) Traditional and molecular techniques for the study of emerging bacterial diseases: one laboratory's perspective. *Emerg Infect Dis* 8: 122-131.
3. Laupland KB, Valiquette L (2013) The changing culture of the microbiology laboratory. *Can J Infect Dis Med Microbiol* 24: 125-128.
4. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: R245-249.
5. Nakamura S, Maeda N, Miron IM, Yoh M, Izutsu K, et al. (2008) Metagenomic diagnosis of bacterial infections. *Emerg Infect Dis* 14: 1784-1786.
6. Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. *Clin Chem* 55: 856-866.
7. Shah N, Tang H, Doak TG, Ye Y (2011) Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput*.
8. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P (2013) Metagenomics for pathogen detection in public health. *Genome Med* 5: 81.
9. Li SK, Leung RK, Guo HX, Wei JF, Wang JH, et al. (2012) Detection and identification of plasma bacterial and viral elements in HIV/AIDS patients in comparison to healthy adults. *Clin Microbiol Infect* 18: 1126-1133.

10. Liu CM, Osborne BJ, Hungate BA, Shahabi K, Huibner S, et al. (2014) The semen microbiome and its relationship with local immunology and viral load in HIV infection. *PLoS Pathog* 10: e1004262.
11. WHO. Hepatitis B Fact sheet N204. July 2014. Retrieved 4 November 2014.
12. Margeridon-Thermet S, Shulman NS, Ahmed A, Shahriar R, Liu T, et al. (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: 1275-1285.
13. Rossi LM, Escobar-Gutierrez A, Rahal P (2015) Advanced molecular surveillance of hepatitis C virus. *Viruses* 7: 1153-1188.
14. Ninomiya M, Ueno Y, Funayama R, Nagashima T, Nishida Y, et al. (2012) Use of illumina deep sequencing technology to differentiate hepatitis C virus variants. *J Clin Microbiol* 50: 857-866.
15. Fischer Walker CL, Perin J, Aryee MJ, Boschi-Pinto C, Black RE (2012). Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. *BMC Public Health* 12: 220.
16. Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, et al. (2008) Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog* 4: e1000011.
17. WHO. Influenza (Seasonal), Fact sheet N 211. March 2014. 2014; 4.
18. Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, et al. (2010) A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* 5: e13381.
19. Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team, Dawood FS, Jain S, Finelli L, Shaw MW, et al. (2009) Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med* 360: 2605-2615.
20. Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, et al. (2009) Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 4: e4219.
21. Johansson H, Bzhalava D, Ekström J, Hultin E, Dillner J, et al. (2013) Metagenomic sequencing of "HPV-negative" condylomas detects novel putative HPV types. *Virology* 440: 1-7.
22. WHO. Global tuberculosis report 2013.
23. Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ (2014) Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ* 2: e585.
24. Chan JZ, Sergeant MJ, Lee OY, Minnikin DE, Besra GS, et al. (2013) Metagenomic analysis of tuberculosis in a mummy. *N Engl J Med* 369: 289-290.
25. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, et al. (2013) A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxicogenic *Escherichia coli* O104:H4. *JAMA* 309: 1502-1510.
26. Andersson P, Klein M, Lilliebridge RA, Giffard PM (2013) Sequences of multiple bacterial genomes and a *Chlamydia trachomatis* genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen. *Clin Microbiol Infect* 19: E405-408.
27. Kim JY, Hahn HJ, Choe YB, Lee YW, Ahn KJ, et al. (2013) Molecular biological identification of malassezia yeasts using pyrosequencing. *Ann Dermatol* 25: 73-79.
28. WHO. Pneumonia, Fact sheet 331. Updated November 2014. In. October 2014.
29. Fischer N, Rohde H, Indenbirken D, Günther T, Reumann K, et al. (2014) Rapid metagenomic diagnostics for suspected outbreak of severe pneumonia. *Emerg Infect Dis* 20: 1072-1075.
30. Eid J, Fehr A, Gray J, Luong K, Lyle J, et al. (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133-138.
31. Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, et al. (2011) The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364: 33-42.
32. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D et al. (2014) A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res* 24: 1180-1192.