Sequencing the future: stopping infectious disease in its tracks using high-throughput sequencing technologies in urban areas

Abstract

Infectious diseases are a significant cause of morbidity and mortality across the world, and there are unique features of life in densely populated cities that make them a particular challenge. They may lead to short or long term illness, with direct healthcare costs and indirect losses in workforce productivity. The use of high-throughput sequencing technologies will allow cities to respond to these challenges on a number of levels. These include human genome sequencing at birth to predict infectious disease susceptibility; genetic surveillance and epidemiology to track emerging infectious diseases; genetic monitoring to ensure the success of vaccine strategies and to prevent the spread of antimicrobial resistance; and genetic diagnosis of infectious disease. Cities of the future must bring together clinical, research and industrial stakeholders to join different pieces of technology and research into a coherent health strategy. Sequencing-based management of infectious diseases has the potential to improve the health and quality of life of residents of future cities, even as populations grow in size and density, and become increasingly globalised. High-throughput sequencing is essential for the continued positive development of cities such as Singapore over the next 50 to 100 years, and beyond.

Main text

Imagine a scene in a future Singapore: with your genome sequenced a few years ago you know that you have a common genetic variant that puts you at increased risk of severe influenza, and you undergo regular monitoring. While you have been vaccinated against strains circulating this year your doctor takes a sample of your saliva and less than an hour later you could have a list of all of the pathogens currently present in your body. The centralised system compares all of the pathogens in your body and those circulating in routine tests in your neighbourhood. Any new pathogen strains with pathogenic potential can be readily picked up. You, your family and your co-workers can be supplied with prophylactic antivirals. This system would be able to prevent individual sickness, formation of drug resistance, and outbreaks.

Every part of this scenario is currently possible in a lab, but at present we lack the understanding and infrastructure to pull these technologies and applications together. With population densities rising, if we are to maintain and improve on current standards of living we have no choice to but embrace a truly unified sequencing system for infectious disease epidemiology, diagnosis and management.

The inhabitants of large cities are at risk from infectious diseases for a number of reasons. Large, mobile populations, typically living at high population densities, are the perfect melting pot for infectious diseases. These range from the short-term and typically mild (colds, food poisoning) to the chronic (tuberculosis) and potentially deadly (epidemics of emerging diseases such as SARS or new strains of influenza). All of these infectious diseases have impacts, including financial costs. Acute respiratory and gastrointestinal infections affect millions of working adults every year and lead to significant reductions in productivity. Norovirus outbreaks are estimated to cost the UK's National Health Service (NHS) £100 million per annum, and food poisoning results $2 billion per annum in lost productivity to the US

economy. Cities at low latitudes have to contend with a burden of vector-driven infectious diseases such as malaria or dengue, with further costs. Antimicrobial and antiviral resistance is a growing problem at every level of the healthcare system, affecting every part of the population, and will lead to significant morbidity and mortality over the next century if not adequately managed.

It is well-known that urbanisation creates ideal circumstances for epidemic diseases to become endemic ones, and for infections to spread rapidly through a population. The story of HIV might be very different without the introduction of this virus to the cities of Kinshasa and Brazzaville in the Democratic Republic of Congo at the start of the 20th century[1]. More recently, Reuters reported the introduction of Middle East Respiratory Syndrome (MERS) to South Korea, the result of global air travel.

These challenges can all be met by better access to genetic data, which can be accessed through high-throughput sequencing (HTS) technology (also known as next-generation sequencing). HTS gives researchers access to rich data that has obvious applications in a health care setting: the complete genome sequence of the patient; the pathogens present in a sample, known and unknown (metagenomics); and for known pathogens, HTS can provide evidence of antimicrobial or antiviral resistance within those pathogens. Future developments in HTS, and the increasing penetrance of these technologies into healthcare will have a significant impact on the way we manage infectious disease.

In order for future cities to leverage these new technologies to maximum effect, I propose sequencing at every level: sequencing of the human genome at birth to predict which pathogens the population is particularly at risk from, and to allow personalised medicine; sequencing of pathogens during disease outbreaks; and sequencing-based surveillance in sentinel or reservoir groups, namely children, the elderly and the immunocompromised. Furthermore, sequence data combined with information technology will be able to predict which genes from new pathogens make good vaccine candidates. I will explore each of these applications in turn.

<u>Personalised medicine: genome sequencing to predict infection susceptibility</u>

Work over the last decade has shown that human genetic factors contribute to the susceptibility of many infectious diseases: dengue[2, 3]; influenza[4], tuberculosis[5] and HIV[6]. If an individual has their genome sequenced at birth, they could be counselled that they are at risk of certain infections (eg hepatitis C); vaccinated and screened regularly for infections that have long latency periods (TB); and targeted for vaccination during outbreaks (eg influenza and dengue).

For individuals, this approach would save money by preventing hospitalisations and lost workforce productivity. The cost of a human genome sequence is now below $1000; compare this to a day spent in intensive care (estimated at $2300-3100 per day by the UK NHS) and it is clear how host genomics could be combined with public health measures to reduce health care costs. On a population level, understanding which diseases the population is most at risk from may help with disease forecasting, deciding which vaccines and drugs to prioritise, and even in shaping the research focus of laboratories towards the biggest infectious disease threats for the population. For example, the severe influenza susceptibility polymorphism rs12252 in IFITM3[4] is present in 0.3% of healthy Europeans but is found in 5.7% of UK patients hospitalised with severe influenza. This polymorphism is much more common in

parts of Asia (25% of healthy Han Chinese and 44% of healthy Japanese) and was found in 69% of patients hospitalised with severe H1N1 flu in Beijing[7]. Influenza pandemics may therefore present a greater public health threat in parts of Asia where this mutation is common and cities in these regions could introduce universal free influenza vaccination programmes each year to reduce the burden of disease.

Sequencing for antibiotic resistant bacteria, drug resistant viral isolates, and transmission chains

Sequencing can be used in the monitoring, diagnosis and management of many viral and bacterial pathogens relevant to cities such as New York, Beijing, Singapore and London. Using TB as an example, the power of new sequencing technologies is being brought to bear on the cause of this increasingly urban disease, *Mycobacterium tuberculosis*.

The UK's Health Protection Agency recently reported that London has TB rates of 42 per 100,000 people (four times the average for the UK) and considerably higher than other European cities such as Paris (23/100,000) and Copenhagen (17/100,000), but comparable to Singapore. One area of London had 119 cases of TB per 100,000 people, higher than the average for India or Russia. Understanding how TB spreads and evolves multi-drug resistance are vital to managing this pathogen. High-throughput TB sequencing can: reduce the time from diagnosis to drug resistance testing from months to days, while also increasing the resolution and sensitivity of resistance testing[8]; distinguish between individuals who failed TB treatment and those who were successfully treated but became re-infected[9]; and rapidly validate transmission links between patients that were previously only inferred epidemiologically[10].

Beyond TB, sequencing can be used to track outbreaks of infection in healthcare settings, for example a norovirus outbreak within a children's hospital[11]; or demonstrate that a pathogen has been circulating in an area for longer than observed disease would suggest. For example, sequencing showed that a strain of dengue virus had been present in Jeddah and wider Saudi Arabia before the index case presented to hospital; furthermore, whole genome sequence data was able to establish the likely origin of this strain (Africa) and a potential transmission route (pilgrims)[12]. Worldwide infectious disease monitoring in highly technologically developed cities may even act as a proxy for surveillance in less developed regions[13].

Sequencing is just as important for detecting infectious diseases of unknown origin, and has been used to identify the aetiological agent of previously undescribed human pathogens including SARS[14] and MERS[15]. It can also be used to diagnose individual patients[16]. There are many diseases thought to have an infectious aetiology but for which a particular pathogen cannot be unambiguously implicated such as Kawasaki disease, a paediatric disease most common in Asian countries. Future cities can and must harness routine sequencing to deal with these threats to individual and public health.

Sequencing pathogens to ensure vaccine efficacy

Sequencing technology has an important role to play in managing vaccine-preventable diseases, for example in ensuring that vaccines provide protection against circulating pathogen diversity, or by

identifying whether live attenuated vaccines are recombining with wild-type strains to cause disease themselves.

Individuals co-infected with hepatitis B virus (HBV) and HIV provide the ideal circumstances for HBV to mutate and escape from current vaccine protection, as these co-infected individuals provide the perfect storm of poor immune responses, chronic infection requiring antiviral treatment for HBV, and a lifestyle that may include many of the risk factors for onwards HBV transmission, such as intravenous drug use[17]. Sequencing of HBV in sentinel groups such as the HIV/HBV co-infected can track the evolution and transmission of these escape mutants. However, HBV vaccine escape mutants also circulate naturally within the population, including in immunocompetent children[18]. A sequencing-based strategy to eradicate HBV would combine surveillance of all newly-vaccinated individuals and high-risk groups to ensure that the antigens in the HBV vaccine continue to provide protection as new escape mutants are detected.

Monitoring of live attenuated vaccine recombination with wild-type strains is being actively performed for the varicella-zoster vaccine with prevents the childhood and old-age manifestations of this common viral infection: chicken pox (varicella), and shingles (zoster). While no wild-type-vaccine recombination events have yet been found[19], it is a real possibility and may require a change a different vaccine strategy which will not allow recombination (eg a subunit vaccine or virus-like particle approach).

Pathogen sequencing for diagnosis and treatment management

Cities of the future, such as Singapore, can lead the way in integrating the approaches discussed above, because they can take the best new technologies developed by researchers and industry, and integrate them into health care systems. This will include developing standard operating procedures for sampling appropriate material (environmental, animal or human) on a rapid-turnaround sequencing platform. This method would need to capture all aspects of pathogen genetic material (DNA and RNA) and sequence them to high depth. The pipeline would then include automated bioinformatic analysis of the sample's genetic contents, with results scored against a database of known and potential pathogens to aid physician diagnosis. If this sequencing is sufficiently optimised and high coverage, it may even be possible to automate diagnoses and suggestion of drug treatments. For viral pathogens such as HIV and cytomegalovirus, there are published, well-characterised lists of mutations conveying resistance to particular drugs. The same is true of some bacterial pathogens. Patients would only be treated with drug to which their infection remains sensitive.

Regular molecular surveillance is an important part of infectious disease control for future cities. This could include taking nasopharyngeal swabs from every person attending primary healthcare as part of the standard work-up. School children and healthcare workers could be sampled regularly, as children are a significant reservoir of influenza-like illnesses. This would provide high-resolution epidemiological sampling, with the potential to identify emerging infectious diseases before they become a significant problem. For example, phylogenetic analysis has revealed that the 2009 swine flu outbreak in the UK was caused by a strain of influenza that was circulating in the UK for several months before the first clinical case was diagnosed[20]. Molecular epidemiology of norovirus suggests that asymptomatic or

chronic norovirus carriers, who may be elderly, malnourished or immune compromised, form a reservoir for this disease[21]. If molecular surveillance methods are used, scientists have the potential to detect novel, perhaps pandemic, disease outbreaks before they come to the notice of healthcare services. This would allow for better pandemic preparedness or quarantine for infection control.

The technology required to maintain these kinds of surveillance and response-based sequencing are gradually diffusing from large, dedicated sequencing centres such as BGI (Beijing Genomics Institute) China, the Broad Institute USA, A*STAR Singapore and the Wellcome Trust Sanger Institute UK to clinical settings in large urban areas. There are numerous hospitals pushing forward sequencing-based infectious disease management and diagnostics within the UK, mostly in or within 100 miles of London. These include Addenbrookes Hospital in Cambridge, and Great Ormond Street Hospital for Children and University College Hospital, both London. These hospitals greatly benefit from a close, synergistic working relationship between basic and translational research, bringing together academic, industrial, governmental and clinical partners. Similar models for treatment are emerging in other major urban areas around the world and represent the future of this kind of healthcare.

In our increasingly globalised cities, we see a melting pot of people – and pathogens – from all over the world. This diversifies the pathogens to which everyone is at risk and makes epidemiological forecasting increasingly challenging. Genetic surveillance and management of infectious diseases must become the new gold standard for preventing and controlling infectious disease in cities of the future.

Preventative medicine

Looking further into the future, large genetic datasets of diverse pathogens would allow computers to predict which proteins produced by novel pathogens make good vaccine candidates. Vaccines could be synthesised on a personalised basis for prevention or therapy, allowing the infected or exposed individual to mount a more directed immune response. Similar database and computational prediction could provide physicians with a list of known drugs the pathogen is likely to respond to.

Vaccines or drugs could be made available to the whole population if an infectious disease has epidemic potential, before such an epidemic starts. The vast majority of infectious diseases circulate for months or even years at a time before an outbreak occurs or they are detected by health care systems. By detecting these pathogens at the earliest opportunity, their ability to further adapt to humans can be curtailed.

Future cities are uniquely placed to bring all these technologies together to improve the health and well-being of their populations. They have the critical mass of knowledge and capacity from industry, research and medical care to unify the technologies and implement them in systems supported by the state to deliver solutions. While living in ever-denser cities presents a range of challenges for future governments and states, it is within our power to overcome them.

References

1. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ*, et al.* HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 2014, **346**(6205)**:** 56-61.

2. Whitehorn J, Chau TNB, Nguyet NM, Kien DTH, Quyen NTH, Trung DT*, et al.* Genetic Variants of MICB and PLCE1 and Associations with Non-Severe Dengue. *PloS one* 2013, **8**(3).

3. Khor CC, Tran NBC, Pang JX, Davila S, Long HT, Ong RTH*, et al.* Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. *Nature genetics* 2011, **43**(11)**:** 1139-U1134.

4. Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE*, et al.* IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* 2012, **484**(7395)**:** 519-523.

5. Curtis J, Luo Y, Zenner HL, Cuchet-Lourenco D, Wu CX, Lo K*, et al.* Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nature genetics* 2015, **47**(5)**:** 523-U128.

6. Apps R, Qi Y, Carlson JM, Chen HY, Gao XJ, Thomas R*, et al.* Influence of HLA-C Expression Level on HIV Control. *Science* 2013, **340**(6128)**:** 87-91.

7. Zhang YH, Zhao Y, Li N, Peng YC, Giannoulatou E, Jin RH*, et al.* Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated with severe influenza in Chinese individuals. *Nature communications* 2013, **4:** 1418.

8. Koser CU, Bryant JM, Becq J, Torok ME, Ellington MJ, Marti-Renom MA*, et al.* Whole-genome sequencing for rapid susceptibility testing of M. tuberculosis. *The New England journal of medicine* 2013, **369**(3)**:** 290-292.

9. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, van Helden P*, et al.* Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study. *The Lancet Respiratory medicine* 2013, **1**(10)**:** 786-792.

10. Torok ME, Reuter S, Bryant J, Koser CU, Stinchcombe SV, Nazareth B*, et al.* Rapid whole-genome sequencing for investigation of a suspected tuberculosis outbreak. *Journal of clinical microbiology* 2013, **51**(2)**:** 611-614.

11. Kundu S, Lockwood J, Depledge DP, Chaudhry Y, Aston A, Rao K*, et al.* Next-Generation Whole Genome Sequencing Identifies the Direction of Norovirus Transmission in Linked Patients. *Clin Infect Dis* 2013, **57**(3)**:** 407-414.

12. Azhar EI, Hashem AM, El-Kafrawy SA, Abol-Ela S, Abd-Alla AM, Sohrab SS*, et al.* Complete genome sequencing and phylogenetic analysis of dengue type 1 virus isolated from Jeddah, Saudi Arabia. *Virology journal* 2015, **12**(1)**:** 1.

13. Manor Y, Shulman LM, Kaliner E, Hindiyeh M, Ram D, Sofer D*, et al.* Intensified environmental surveillance supporting the response to wild poliovirus type 1 silent circulation in Israel, 2013. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2014, **19**(7)**:** 20708.

14. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S*, et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *The New England journal of medicine* 2003, **348**(20)**:** 1967-1976.

15. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England journal of medicine* 2012, **367**(19)**:** 1814-1820.

16. Brown JR, Morfopoulou S, Hubb J, Emmett WA, Ip W, Shah D*, et al.* Astrovirus VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients. *Clin Infect Dis* 2015, **60**(6)**:** 881-888.

17. Teo CG, Locarnini SA. Potential threat of drug-resistant and vaccine-escape HBV mutants to public health. *Antiviral therapy* 2010, **15**(3 Pt B)**:** 445-449.

18. Lin YM, Jow GM, Mu SC, Chen BF. Naturally occurring hepatitis B virus B-cell and T-cell epitope mutants in hepatitis B vaccinated children. *TheScientificWorldJournal* 2013, **2013:** 571875.

19. Norberg P, Depledge DP, Kundu S, Atkinson C, Brown J, Haque T*, et al.* Recombination of Globally Circulating Varicella Zoster Virus. *Journal of virology* 2015.

20. Baillie GJ, Galiano M, Agapow PM, Myers R, Chiam R, Gall A*, et al.* Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *Journal of virology* 2012, **86**(1)**:** 11-18.

21. Karst SM, Baric RS. What Is the Reservoir of Emergent Human Norovirus Strains? *Journal of virology* 2015, **89**(11)**:** 5756-5759.