

Commentary

Human genomics and preparedness for infectious threats

Nicole F Dowling*, Marta Gwinn* and Alison Mawle†

Addresses: *Office of Public Health Genomics, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA. †National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.

Correspondence: Nicole F Dowling. Email: ndowling@cdc.gov

Abstract

Public health preparedness requires effective surveillance of and rapid response to infectious disease outbreaks. Inclusion of research activities within the outbreak setting provides important opportunities to maximize limited resources, to enhance gains in scientific knowledge, and ultimately to increase levels of preparedness. With rapid advances in laboratory technologies, banking and analysis of human genomic specimens can be conducted as part of public health investigations, enabling valuable research well into the future.

Introduction

Despite major progress toward understanding infectious agents and controlling their spread, new and evolving infectious diseases - as well as old diseases in new contexts - continue to pose threats to humans worldwide. In 1992, the Institute of Medicine published an influential report calling attention to the emergence and re-emergence of human pathogens as a consequence of such factors as evolutionary changes in infectious agents and their human and non-human hosts; alterations in host behaviors and travel; and naturally occurring and man-made shifts in ecology, geography, and environment [1]. During the following decade, renewed concern about microbial threats to health spurred new investments in scientific research and public health infrastructure. In 2003 the Institute of Medicine published a report entitled *Microbial Threats to Health*, which highlighted the need for a global approach to preparedness [2]. That same year, the severe acute respiratory syndrome (SARS) epidemic acutely challenged the response capacity of scientists and public health officials across the globe [3,4]. Advances in high-throughput genome sequencing technology played a pivotal role in identifying the novel coronavirus associated with SARS and in facilitating the development of assays for diagnosis and control [5].

Technological advances

Public health investigations of infectious diseases have relied increasingly on molecular epidemiology since the

introduction of restriction fragment length polymorphism (RFLP) analysis in the 1980s. The first full genome sequence for a human bacterial pathogen, *Haemophilus influenzae*, was completed in 1995 [6]. Since then, the development of sequencing technologies has made genomic analysis of emerging pathogens easier, faster, and less expensive; instead of taking months or weeks, such investigations can often be accomplished in days. Recently, Musser and Shelburne reviewed a decade of progress in pathogenomic analysis of group A streptococcus infections, made possible by technical advances, including low-cost DNA sequencing, microarray technology, and high-throughput proteomics [7]. Application of these techniques has uncovered new virulence factors and provided insights into bacterial-host interactions, which are important for preventing invasive infections and developing effective vaccines.

More recently, within days of the initial identification of the first cases of 2009 pandemic influenza A (H1N1) in spring 2009, scientists had identified the origin of all eight influenza virus gene segments. Within two weeks, the Centers for Disease Control and Prevention (CDC) began to distribute RT-PCR diagnostic test kits to public health laboratories under a quickly granted emergency use authorization by the US Food and Drug Administration [8-10].

Developments in informatics have been crucial for the successful application of genomics to infectious disease research [11]. Making research data freely accessible in continuously updated, online databases further enhances their utility for public health investigation. Such resources recently allowed researchers to compare sequences of the pandemic 2009 H1N1 influenza virus with other influenza viruses, to quickly identify potentially important features [12]. As part of its Influenza Virus Resource [13], the National Center for Biotechnology Information (NCBI) has created a specific resource for H1N1 influenza genome

CDC, Centers for Disease Control and Prevention; GWAS, genome-wide association study; NCBI, National Center for Biotechnology Information; RFLP, restriction fragment length polymorphism; RT-PCR, reverse-transcriptase polymerase chain reaction; SARS, severe acute respiratory syndrome.

sequence data [14]; as of September 2009, the database contained more than 1,100 different nucleotide sequences and over 400 full-length viral genomes for 2009 H1N1 influenza viruses. NCBI's Entrez Genome database provides whole genome sequences for more than 1,000 organisms, including *Homo sapiens*, as well as bacteria, viruses, and parasites that cause human disease [15]. The CDC has quickly shared virus sequence data on public websites [16]. Other research organizations, such as the Viral Bioinformatics Resource Center and the Wellcome Trust's Sanger Institute, have also developed repositories of genomic information of public health importance [17,18].

Advances in genomics and the completion of the Human Genome and HapMap Projects have opened the door to research on the role of human genetic variation in the population distribution, transmission, and severity of infectious diseases. Published studies in 'human genome epidemiology' have been tracked since 2001 in a CDC-sponsored database, HuGE Navigator, which is freely available online for quick searching on human genes, diseases, and environmental factors, including pathogens [19,20]. Of the more than 40,000 studies in the database, most focus on chronic diseases; however, more than 2,000 so far are related to infectious diseases, including several genome-wide association studies (GWASs). In recent years, GWASs have become a powerful tool for systematically searching the human genome for novel associations with infectious diseases, including tuberculosis, malaria, and HIV [21-24].

Preparedness for research

Public health efforts to control and prevent infectious diseases are based on epidemiologic and laboratory surveillance systems that detect and monitor disease incidence, define pathogen characteristics, and track cases 'by person, place and time' [25]. Molecular epidemiology has long been a mainstay of public health surveillance; now, increasingly powerful molecular methods, including sequencing of whole pathogen genomes, help investigators to identify epidemiologically related cases, describe patterns of transmission, pinpoint sources of infection, and explain antimicrobial resistance [26-33]. Combining the methods of molecular pathogenomics with population genetics and epidemiology can provide new insights into the episodic behavior of epidemics of familiar pathogens, such as group A streptococcus [7]. Archived biological samples can also provide new insights into the emergence and evolution of infectious threats, from HIV [34] to influenza [35]. Building the capacity for human biological sample collection into existing surveillance networks has the potential to facilitate a more comprehensive, population-based evaluation of genomic and environmental determinants of health outcomes. For example, a meta-analysis of surveillance cohorts from Arizona, Colorado, California, and Illinois

demonstrated that human homozygosity for *CCR5delta32* (a non-functional variant of chemokine receptor CCR5) is consistently associated with symptomatic West Nile virus infection [36]. When disease surveillance is conducted independently in multiple jurisdictions, the capacity to share biological samples and epidemiologic and clinical data provides important infrastructure for research.

Laboratory infrastructure for pathogen genotyping and sequencing is readily adaptable to the analysis of human genomes; indeed, human genome studies can now be accomplished in the time that pathogen studies required only a decade ago. Additional planning and investment are needed to support the collection and storage of specimens that allow for comprehensive evaluation of key attributes of pathogen and host. In the United States, public health agencies routinely collect, store, and analyze data from individuals to identify and control public health threats; regulations regarding privacy and human subjects' research protections do not always apply to these activities [37]. Preparedness for research in the public health setting should explicitly address protection of human subjects, for example, through development of research protocols that have been pre-approved by review boards to pave the way for systematic research and data collection.

In an infectious disease outbreak, the immediate priority is to limit and contain the threat to human health and wellbeing. Research during and following the outbreak can also be important for developing effective treatments and preventive measures, and for guiding public health policies. Recently, research conducted within ongoing investigations of H1N1 infection has demonstrated that pregnant women may be at elevated risk for complications from infection [38]. These findings have led to recommendations to prioritize pregnant women for vaccination and, if infected, for antiviral therapy. Several published candidate gene studies of susceptibility to viral and bacterial diseases have demonstrated the feasibility and utility of integrating host genomics into epidemiologic studies and surveillance [36,39-42]. The greatest public health impact of such research may be through accelerating the development of preventive vaccines; however, novel diagnostic tests developed for clinical use may also prove useful for epidemic monitoring and triage during outbreaks. For example, scientists recently demonstrated that human gene expression profiles are an effective tool for determining the etiology of respiratory infections, providing a striking example of rapid translation from basic research to potential clinical and public health application [43].

In a recent editorial titled 'Epidemic science in real time', Fineberg and Wilson underscored the critical importance of 'conducting the right science and communicating expert judgment' to 'enable policies to be adjusted appropriately as an epidemic scenario unfolds' [44]. They emphasized

that in times of diminishing public health resources, scientists from diverse disciplines - epidemiology, laboratory, social sciences - must work together to respond to immediate threats and follow through with research to understand key attributes of the affected populations and the disease process. The results of such research are needed to inform policy, to develop treatments and interventions, and to update and adjust recommendations as the state of knowledge changes.

Conclusions

Integrating genomics research into the context of public health surveillance and response can help maximize the use of limited resources, enhance the exchange and growth of scientific knowledge, and increase preparedness for infectious threats. Such research should be based on sound protocols that protect human subjects. Specimens should be processed and banked, enabling future research on genetic variation of both pathogen and host, as well as gene expression profiles, proteomics, and other measures. Epidemiologic data about environment and behaviors should be collected and stored to support additional analysis of gene-environment interactions. Such efforts will require a shift in culture and broadening of traditional public health definitions of preparedness and response, research, and collaboration.

Competing interests

The authors declare that they have no competing interests

Authors' contributions

NFD, MG, and AM were involved in drafting the manuscript and have given final approval of the version to be published.

Acknowledgements

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

- Institute of Medicine, edited by Lederberg J, Shope R, Oaks S: *Emerging Infections: Microbial Threats to Health in the United States*. Washington, DC: National Academy Press; 1992.
- Institute of Medicine, edited by Smolinski M, Hamburg M, Lederberg J: *Microbial Threats to Health: Emergence, Detection, and Response*. Washington, DC: National Academy Press; 2003.
- Finlay BB, See RH, Brunham RC: **Rapid response research to emerging infectious diseases: lessons from SARS**. *Nat Rev Microbiol* 2004, **2**:602-607.
- Snowden FM: **Emerging and reemerging diseases: a historical perspective**. *Immunol Rev* 2008, **225**:9-26.
- Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattri J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girm N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, et al.: **The genome sequence of the SARS-associated coronavirus**. *Science* 2003, **300**:1399-1404.
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM, McKenney K, Sutton G, 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, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, et al.: **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd**. *Science* 1995, **269**:496-512.
- Musser JM, Shelburne SA 3rd: **A decade of molecular pathogenomic analysis of group A *Streptococcus***. *J Clin Invest* 2009, **119**:2455-2463.
- Centers for Disease Control and prevention (CDC): **Swine influenza A (H1N1) infection in two children - Southern California, March-April 2009**. *MMWR Morb Mortal Wkly Rep* 2009, **58**:400-402.
- Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivaller P, Smagala J, de Graaf M, Burke DF, Fouchier RA, Pappas C, Alpuche-Aranda CM, López-Gatell H, Olivera H, López I, Myers CA, Faix D, Blair PJ, Yu C, et al.: **Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans**. *Science* 2009, **325**:197-201.
- Maurer-Stroh S, Ma J, Lee RT, Sirota FL, Eisenhaber F: **Mapping the sequence mutations of the 2009 H1N1 influenza A virus neuraminidase relative to drug and antibody binding sites**. *Biol Direct* 2009, **4**:18.
- Greene JM, Collins F, Lefkowitz EJ, Roos D, Scheuermann RH, Sobral B, Stevens R, White O, Di Francesco V: **National Institute of Allergy and Infectious Diseases bioinformatics resource centers: new assets for pathogen informatics**. *Infect Immun* 2007, **75**:3212-3219.
- Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, Garten RJ, Gubareva LV, Xu X, Bridges CB, Uyeki TM: **Emergence of a novel swine-origin influenza A (H1N1) virus in humans**. *N Engl J Med* 2009, **360**:2605-2615.
- Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D: **The influenza virus resource at the National Center for Biotechnology Information**. *J Virol* 2008, **82**:596-601.
- Influenza Virus Resource: GenBank sequences from pandemic (H1N1) 2009 viruses** [<http://www.ncbi.nlm.nih.gov/genomes/FLU/SwineFlu.html>] (accessed 10 December 2009).
- NCBI Entrez Genome** [<http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome>] (accessed 10 December 2009).
- Centers for Disease Control and Prevention** [<http://www.cdc.gov/>](accessed 10 December 2009).
- Viral Bioinformatics Resource Center** [<http://athena.bioc.uvic.ca/>](accessed 10 December 2009).
- Wellcome Trust Sanger Institute: Sequencing Genomics Projects** [<http://www.sanger.ac.uk/Projects>] (accessed 10 December 2009).
- HuGE Navigator** [<http://www.hugenavigator.net/>] (accessed 10 December 2009).
- Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ: **A navigator for human genome epidemiology**. *Nat Genet* 2008, **40**:124-125.
- Davila S, Hibberd ML: **Genome-wide association studies are coming for human infectious diseases**. *Genome Med* 2009, **1**:19.
- Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, Kivinen K, Bojang KA, Conway DJ, Pinder M, Sirugo G, Sisay-Joof F, Usen S, Auburn S, Bumpstead SJ, Campino S, Coffey A, Dunham A, Fry AE, Green A, Gwilliam R, Hunt SE, Inouye M, Jeffreys AE, Mendy A, Palotie A, Potter S, Ragoussis J, Rogers J, Rowlands K, et al.: **Genome-wide and fine-resolution association analysis of malaria in West Africa**. *Nat Genet* 2009, in press.
- Limou S, Le Clerc S, Coulonges C, Carpentier W, Dina C, Delaneau O, Labib T, Taing L, Sladek R, Deveau C, Ratsimandresy R, Montes M, Spadoni JL, Lelièvre JD, Lévy Y, Therwath A, Schächter F, Matsuda F, Gut I, Froguel P,

- Delfraissy JF, Hercberg S, Zagury JF; ANRS Genomic Group: **Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02).** *J Infect Dis* 2009, **199**: 419-426.
24. Vannberg FO, Chapman SJ, Khor CC, Tosh K, Floyd S, Jackson-Sillah D, Crampin A, Sichali L, Bah B, Gustafson P, Aaby P, McAdam KP, Bah-Sow O, Lienhardt C, Sirugo G, Fine P, Hill AV: **CD209 genetic polymorphism and tuberculosis disease.** *PLoS ONE* 2008, **3**:e1388.
 25. Thacker SB, Berkelman RL: **Public health surveillance in the United States.** *Epidemiol Rev* 1988, **10**:164-190.
 26. Charlton B, Crossley B, Hietala S: **Conventional and future diagnostics for avian influenza.** *Comp Immunol Microbiol Infect Dis* 2009, **32**:341-350.
 27. Cummings CA, Relman DA: **Genomics and microbiology. Microbial forensics – 'cross-examining pathogens'.** *Science* 2002, **296**:1976-1979.
 28. Karpathy SE, Dasch GA, Ereemeeva ME: **Molecular typing of isolates of *Rickettsia rickettsii* by use of DNA sequencing of variable intergenic regions.** *J Clin Microbiol* 2007, **45**: 2545-2553.
 29. Liu JY, Lin HH, Liu YC, Lee SS, Chen YL, Hung CC, Ko WC, Huang CK, Lai CH, Chen YS, Chung HC, Liang SH, Lin JN: **Extremely high prevalence and genetic diversity of hepatitis C virus infection among HIV-infected injection drug users in Taiwan.** *Clin Infect Dis* 2008, **46**:1761-1768.
 30. Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM, Solomon D, Keim P, Fraser CM: **Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*.** *Science* 2002, **296**:2028-2033.
 31. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB: **Tuberculosis drug resistance mutation database.** *PLoS Med* 2009, **6**:e2.
 32. Smith DM, May SJ, Tweeten S, Drumright L, Pacold ME, Kosakovsky Pond SL, Pesano RL, Lie YS, Richman DD, Frost SD, Woelk CH, Little SJ: **A public health model for the molecular surveillance of HIV transmission in San Diego, California.** *AIDS* 2009, **23**:225-232.
 33. Rota PA, Featherstone DA, Bellini WJ: **Molecular epidemiology of measles virus.** *Curr Top Microbiol Immunol* 2009, **330**: 129-150.
 34. Galvani AP, Novembre J: **The evolutionary history of the CCR5-Delta32 HIV-resistance mutation.** *Microbes Infect* 2005, **7**:302-309.
 35. Tumpey TM, Belser JA: **Resurrected pandemic influenza viruses.** *Annu Rev Microbiol* 2009, **63**:79-98.
 36. Lim JK, Louie CY, Glaser C, Jean C, Johnson B, Johnson H, McDermott DH, Murphy PM: **Genetic deficiency of chemokine receptor CCR5 is a strong risk factor for symptomatic West Nile virus infection: a meta-analysis of 4 cohorts in the US epidemic.** *J Infect Dis* 2008, **197**:262-265.
 37. Lee LM, Gostin LO: **Ethical collection, storage, and use of public health data: a proposal for a national privacy protection.** *JAMA* 2009, **302**:82-84.
 38. Jamieson DJ, Honein MA, Rasmussen SA, Williams JL, Swerdlow DL, Biggerstaff MS, Lindstrom S, Louie JK, Christ CM, Bohm SR, Fonseca VP, Ritger KA, Kuhles DJ, Eggers P, Bruce H, Davidson HA, Lutterloh E, Harris ML, Burke C, Cocoroc N, Finelli L, MacFarlane KF, Shu B, Olsen SJ; Novel Influenza A (H1N1) Pregnancy Working Group: **H1N1 2009 influenza virus infection during pregnancy in the USA.** *Lancet* 2009, **374**:451-458.
 39. Crawford DC, Zimmer SM, Morin CA, Messonnier NE, Lynfield R, Yi Q, Shephard C, Wong M, Rieder MJ, Livingston RJ, Nickerson DA, Whitney CG, Lingappa J: **Integrating host genomics with surveillance for invasive bacterial diseases.** *Emerg Infect Dis* 2008, **14**:1138-1140.
 40. Lingappa J, Kuffner T, Tappero J, Whitworth W, Mize A, Kaiser R, McNicholl J: **HLA-DQ6 and ingestion of contaminated water: possible gene-environment interaction in an outbreak of Leptospirosis.** *Genes Immun* 2004, **5**:197-202.
 41. Thorven M, Grahn A, Hedlund KO, Johansson H, Wahlfriid C, Larson G, Svensson L: **A homozygous nonsense mutation (428G→A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections.** *J Virol* 2005, **79**:15351-15355.
 42. Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, Baric RS: **Mechanisms of GII.4 norovirus persistence in human populations.** *PLoS Med* 2008, **5**:e31.
 43. Zaas AK, Chen M, Varkey J, Veldman T, Hero AO 3rd, Lucas J, Huang Y, Turner R, Gilbert A, Lambkin-Williams R, Øien NC, Nicholson B, Kingsmore S, Carin L, Woods CW, Ginsburg GS: **Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans.** *Cell Host Microbe* 2009, **6**:207-217.
 44. Fineberg HV, Wilson ME: **Epidemic science in real time.** *Science* 2009, **324**:987.

Published: 29 December 2009

doi:10.1186/gm119

© 2009 BioMed Central Ltd