The use of whole genome sequencing in the investigation of a nosocomial influenza virus outbreak

Catherine Houlihan[1*], Dan Frampton[1], R. Bridget Ferns[1,2], Jade Raffle[1], Paul Grant[3], Myriam Reidy[4], Leila Hail[4], Kirsty Thomson[5], Frank Mattes[3], Zisis Kozlakidis[1,6], Deenan Pillay[1], Andrew Hayward[6,7], Eleni Nastouli[3,8]

* Corresponding Author

Catherine Houlihan.

Email Catherine.houlihan@lshtm.ac.uk

1. Division of Infection and Immunity, University College London, London, United Kingdom

2. NIHR UCLH/UCL Biomedical Research Centre, London, United Kingdom

3. Department of Clinical Virology, University College London Hospitals NHS Foundation Trust, London, United Kingdom

4. Infection Control, University College London Hospitals NHS Foundation Trust, London, United Kingdom

5. Department of Blood Diseases, University College London Hospitals NHS Foundation Trust, London, United Kingdom

6. Department of Infectious Disease Informatics, Farr Institute of Health Informatics Research, London, United Kingdom

7. Institute of Epidemiology and Health Care, University College London, London, United Kingdom

8. Department of Population, Policy and Practice, UCL GOS Institute of Child Health, London, United Kingdom

Key words

Influenza; nosocomial; sequencing; transmission

Table of contents summary

Influenza virus whole genome sequencing (WGS) was applied to an outbreak of influenza A in a haematology/oncology ward; two separate introductions were identified. The role of WGS for nosocomiail outbreak investigation is discussed.

Abstract

Traditional epidemiological investigation of nosocomial transmission of influenza involves the identification of patients who have the same influenza virus type and who have overlapped in time and place. This method may miss-identify transmission where it has not occurred or miss transmission when it has. We applied influenza virus whole genome sequencing (WGS) to an outbreak of influenza A in a haematology/oncology ward and identified two separate introductions; one which resulted in 5 additional infections and 79 bed-days lost. Results from WGS are becoming rapidly available and

2

may supplement traditional infection control procedures in the investigation and management of nosocomial outbreaks.

Background

Nosocomial transmission of influenza A virus is of significant concern since infection in individuals who are immunocompromised, immunosuppressed, at extremes of age or pregnant have an increased risk of severe illness, morbidity and death[1,2]. The risk of nosocomial acquisition of influenza virus is high since patients commonly share open bays before viral respiratory infection is diagnosed, and significant health care costs are associated with hospital ward closures due to influenza virus outbreaks. Influenza can also be asymptomatic and prolonged shedding has been demonstrated in those who are immunocompromised[3]. Low rates of influenza vaccination in Health care staff (HCS) in England (49.5% in January 2016) combined with reduced vaccine efficacy raises the possibility that HCS are involved in nosocomial transmission[2].

Commonly, nosocomial transmission of influenza virus within the health care setting has been identified through traditional molecular diagnostic methods (real-time PCR; RT-PCR) for the detection of viral species, combined with data collected on patient and staff movement within the hospital. Patients who have overlapped in time and space and who have evidence of infection with the same viral type, and subtype if available, are assumed to have transmitted to each other. Direct Sanger sequencing methodologies have more recently been used to identify possible clusters of nosocomial influenza infection[4], but next generation sequencing (NGS) of the whole virus genome (WGS), which may offer increased discriminatory capacity, has been infrequently reported. We describe an outbreak of influenza A virus, during early 2016, on a

3

haematology/oncology ward in an NHS hospital in London, in which timely use of WGS would have identified the presence or absence of nosocomial transmission, and allowed a more targeted infection control response.

Clinical cases

Two patients with haematological/oncological malignancies (A and B, Figure 1), who had each been admitted for 2 and 5 weeks on the North side of Ward 1, a haematology/oncology ward in a London NHS Hospital developed coryzal symptoms with fever in the same 24 hour period. Combined nose and throat swabs were taken from each patient at symptom onset (day 0), as per the hospital policy, and were tested using multiplex RT-PCR for a standard panel of 6 respiratory viruses, including influenza A and B. Results were available on day 1, and influenza A virus was detected in samples from both patients. Nosocomial transmission was not assumed to have occurred between these patients since they were in separate positive pressure rooms. The mechanism was therefore uncertain. On day 1, a further patient (C, Figure 1) on the South side of the ward, developed symptoms and tested positive for influenza A. Since this south-side patient had no clear link to patients A and B on the north side, it was assumed that an outbreak was not occurring and the ward was not closed. Over the next 48 hours, a further 2 patients tested positive for influenza A virus (patients D on day 2, and E on day 3) and the ward was closed on day 3. Testing of one further symptomatic patient and screening all asymptomatic patients on day 4 identified two further cases (F was symptomatic and G was asymptomatic). A visiting relative of patient B later revealed that they had coryzal symptoms which preceeded, and continued throughtout, the outbreak. This relative tested positive for influenza A virus one week after the first patients were tested.

4

A total of 7 patients developed influenza A infection on this ward. One of these patients (E) required a prolonged period in intensive care. Thirteen patients who were contacts of the patients illustrated were screened, tested negative and commenced on prophylaxis in keeping with Public Health England guidance and the trust policy(5). The ward was closed for a total of 6 days, and resulted in 79 bed-days lost. Five health care workers developed influenza-like illness and were not permitted to attend work as per the trust's occupational health policy. They were not tested for influenza virus.

Whole genome sequencing and assembly of influenza A virus genomes

Samples from the 7 patients on the same ward, along with 9 samples which were submitted to the same laboratory on the same days from the same hospital, and the symptomatic relative who had tested positive for influenza A virus using standard in-house RT-PCR, were sequenced using the Mi-Seq™ (Illumina Inc.) platform.

Amplification of influenza A virus RNA was carried out using the One-Step SuperScript III RT-PCR kit (Invitrogen™) and a modified eight-segment PCR method(6). Fifty μl reactions containing: 10.0μl RNA, a final concentration of 1xSuperScript III one step RT-PCR buffer, 0.1mM of each primer and 1.0μl SuperScript III RT/Platinum Taq high-fidelity enzyme were prepared. RT-PCR thermal cycling conditions were as described(6). Two PCR amplicons were generated; equal volumes of both were combined and total DNA concentration determined using Qubit HS DNA assay (Invitrogen™)

5

Library preparations were generated using the Nextera®XT DNA Sample Preparation Kit (Illumina Inc.) according to manufacturer's instructions using 1ng of input DNA. Tagged PCR samples were purified with 30µl AMPure XP beads (Agencourt, Cat. No. A63881). Sample library normalization and MiSeq™ sample loading were carried out according to the Nextera®XT protocols. Pooled normalized samples, including a Phi-X control at a final concentration of 0.1pM, were loaded onto a Mi-Seq™ reagent Kit V3 600 cycle (Illumina Inc.) and sequenced on a MiSeq™ (Illumina Inc.).

Consensus genomes were generated from short reads using ICONIC's in-house *de novo* assembly pipeline, applying a read depth cut-off of >=20 reads to the final sequences. Phylogenetic analysis was undertaken firstly by separately aligning each set of segments using MAFFT(7) and concatenating the coding regions within Aliview(8). Maximum-likelihood phylogenetic trees were inferred for each alignment using RAxML(9). Phylogenies were inferred under a general-time reversible substitution model with rate heterogeneity among sites modelled under a 4-category discrete approximation of a gamma distribution. Branch support was assessed through non-parametric bootstrapping of 1000 pseudoreplicates. Direct linkage of whole virus genomes was considered to have statistical support if the observed number of mutations between them occurred within the 95% confidence interval of the expected number (given the mutation rate, time interval between sample collection, genome and pairwise alignment lengths) and they clustered with respective bootstrap values >=95% (see supplementay methods).

6

Results

Seven samples from 7 haematology/oncology patients, 1 sample from a visiting relative of patient B, and 9 samples from other patients in the same hospital tested positive for Influenza A virus on the multiplex RT-PCR. Whole genome sequencing was successful for 15 of these 17 with a median read depth across all samples of 800 (Q1-Q3 600-5000). Attempts to sequence the sample from the relative of patient A were unsuccessful. The sample from patient D failed to generate sequence across all eigth segments of the genome with sufficient depth, resulting in reliable sequence for NS1 alone (segment 8). Despite this limitation, when reliable sequence from this patient's sample was included in a maximum-likelihood phylogenetic tree analysis, patients B, C, D, E, F and G clustered tightly with high bootstrap support (>95%), with almost identical sequences (Figure 2). The same result was obtained when patient D was excluded from the phylogenetic analysis (appendix Figure A1).

There was no evidence of transmission between patient A and B, who both had influenza A detected by RT-PCR on day 1. These are likely to have been two separate nosocomial acquisitions. Incidentally, a further probable (unrecognised) nosocomial transmission was identified on a separate ward (patients H and I, Figure 2).

Discussion

Using WGS, we have illustrated the dynamics of an influenza virus outbreak on a haematology/oncology ward, clarifying that two patients A and B, considered to be 'index cases' using traditional epidemiological methods, were not linked infections and one of these individuals was not involved in the outbreak. The use of traditional epidemiological infection control measures erroneously linked these two unrelated infections and missed spread of the infection from the North to the South side of the

7

ward, leading to a delay in infection control action. Had WGS data been available in real time, the cross-ward transmission would have been identified on day 1 or 2, and the ward closed. Identification of the separate introduction of influenza A (patient A) would additionally have led to a separate investigation and instigation of infection control procedures specific to this introduction.

The integration of WGS data with epidemiological analysis in nosocomial infections could identify common pathways of transmission on wards, for example; communal areas, bays, shared patient equipment, healthcare staff, or visitors interacting with each other and other patients. Early identification of pathways of transmission could prevent further nosocomial cases of influenza. Interventions would include earlier initiation of; ward closure, prophylaxis, emphasis on staff influenza vaccination and treatment if unwell, enhanced equipment or room cleaning, and the limiting of visitor-visitor or visitor-patient contact. Prophylaxis has been shown to reduce the risk of symptomatic influenza infection in immunocompetent and immunocompromised adults by up to 80%(11). Staff vaccination rates on this ward were extremely low at the time of the outbreak; 25% of staff working on the North end had received the seasonal influenza vaccine, and 55% on the south (Figure 1). Analysis of effectiveness of the WHO-recommended seasonal influenza vaccination against the circulating Influenza A(H1N1)pdm09 at the time has been reported as 65% overall(12). Demonstration of the nosocomial outbreak described on this ward led anaecdotally to an increased vaccine uptake in staff.

We were able to achieve full length genome sequencing on all but two of the isolates in this case, allowing more accurate linkage of infections compared to limited sequencing of the haemaglutanin gene or neuraminidase gene. One of these two failed attempts at

8

sequencing may be attributable to sampling the relative of patient B close to the end of their illness, where quantity and quality of viral RNA is less robust.

NGS has been used to identify common sources of nosocomial outbreaks of norovirus and MRSA, but less frequently for identification of transmission of respiratory viruses including influenza virus (13). Possible transmission between patients can be identified from relatively small amounts of viral RNA in specimens. However, bioinformatics analysis, equipment and consumables costs have limited the implementation of this technology into routine clinical care. These limitations, as well as reduction in turnaround times, are improving with time and recognition of their clinical utility. Whole genome sequencing is now available using small, lightweight and easily transportable devices such as the MinION™ (Oxford Nanopore). It is expected that NGS results for recognised pathogens with established pipelines (i.e. influenza sequencing) will be available 24 hours after sample receipt in the near future.

Two large studies from an epidemiological surveillance unit in Germany and a retrospective analysis of nosocomial transmissions in care-facilities in Canada have confirmed nosocomial influenza virus transmissions where it was thought to have occurred, using WGS(14,15). Although it could be argued that confirmation of a suspected transmission has high cost with limited benefit, we have shown that WGS can demonstrate separate introductions and allow tailored infection control approaches. Further, the use of NGS during outbreaks of highly infectious diseases with significant consequences and the sharing of data with public health teams has been instrumental in understanding transmission, highlighted in the recent Ebola and Zika outbreaks.

9

At this hospital, during the 2016 influenza season, WGS allowed early identification of the predominant circulating subtype as H1N1, rather than H3N2, and led to a local policy change from the recommendation of oseltamivir as first line treatment to Zanamivir. PHE altered the national guideline in a similar way, 1 month later(5). Moving forward, the use of WGS would allow a more timely and accurate shift in treatment recommendation, and individualsised antiviral treatment for influenza would be possible.

We have described an outbreak of influenza in a haematology/oncology ward in which real-time use of WGS may have mitigated propagation of the outbreak, and certainly would have increased understanding of it. The use of WGS during the influenza season either for all hospitalised patients, or targeted to patients who develop influenza during an inpatient stay, may identify transmission where this was not suspected, and allow rapid implementation of infection control procedures.

10

<u>Figure titles</u>

Figure 1

Map showing the layout of Ward. Rectangles are side rooms and squares indicate bays with 4 beds. Letters indicate patients. Patient letters are coloured according to whether influenza sequences are considered part of the same transmission cluster (green or purple). Patient D is coloured black since the genome coverage was considered of sufficient depth.

Figure 2

Maximum-likelihood tree derived from a genomic alignment of sequences generated during the influenza outbreak on Ward 1. Tips are coloured according to location within the hospital and whether the patients are considered part of the same transmission cluster (green: Ward 1, linked; blue: Ward 1, unlinked; red: elsewhere in the same hospital, unlinked). Tips in black indicate insufficient genome coverage (only NS-1). Bootstrap support values (%) from 1000 replicates are shown for each node. OPD is Out Patient Department. A and E is Accident and Emergency Department.

11

References

1. Louie JK, Winter K, Jean C, Glaser CA, Samuel MC, Rosenberg J, et al. Factors Associated With Death or Hospitalization Due to Pandemic 2009 Influenza A (H1N1) Infection in California. **2016**; 302(17).

2. PHE. PHE Weekly National Influenza Report [Internet]. 2015. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/fil e/490370/Weekly_report_current_wk_1_7Jan2016.pdf

3. Lehners N, Tabatabai J, Prifert C, Wedde M, Puthenparambil J, Weissbrich B, et al. Long-Term Shedding of Influenza Virus, Parainfluenza Virus, Respiratory Syncytial Virus and Nosocomial Epidemiology in Patients with Hematological Disorders. PLoS One [Internet]. **2016**; 11(2):e0148258. Available from: http://dx.plos.org/10.1371/journal.pone.0148258

4. Jonges M, Rahamat-Langendoen J, Meijer A, Niesters HG, Koopmans M. Sequence-based identification and characterization of nosocomial influenza A(H1N1)pdm09 virus infections. J Hosp Infect [Internet]. Elsevier Ltd; **2012**; 82(3):187–193. Available from: http://dx.doi.org/10.1016/j.jhin.2012.08.004

5. Public Health England. Antiviral prescribing when A(H1N1) pdm09 influenza virus is the dominant circulating strain. Update letter, January 2016. 2016.

6. Watson SJ, Welkers MRA, Depledge DP, Coulter E, Breuer JM, Jong MD De, et al. Viral population analysis and minority-variant detection using short read next-generation sequencing. Phylosophical Trans R Soc. **2013**; 368.

7. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol [Internet]. Oxford

12

University Press; **2013** [cited 2018 Feb 12]; 30(4):772. Available from:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3603318/

8. Larsson A. AliView: A fast and lightweight alignment viewer and editor for large

datasets. Bioinformatics. **2014**; 30(22):3276–3278.

9. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis

of large phylogenies. Bioinformatics [Internet]. Oxford University Press; **2014**

[cited 2018 Feb 12]; 30(9):1312–1313. Available from:

https://academic.oup.com/bioinformatics/article-

lookup/doi/10.1093/bioinformatics/btu033

10. Balin SJ, Cascalho M. The rate of mutation of a single gene. Nucleic Acids Res.

2009; 38(5):1575–1582.

11. Jefferson T, Ma J, Doshi P, Cb DM, Hama R, Mj T, et al. Neuraminidase inhibitors

for preventing and treating influenza in adults and children (Review). Cochrane.

2015; (4).

12. Chambers C, Skowronski D, Sabaiduc S, Winter A, Dickinson J. Interim estimates

of 2015/16 vaccine ectiveness against influenza A(H1N1)pdm09, Canada,

February 2016. Eurosurveillance. **2016**; 21(11).

13. Seong M, Cho SI, Park H, Seo SH, Lee SJ, Kim E, et al. Genotyping Influenza Virus

by Next-Generation Deep Sequencing in Clinical Specimens. **2016**; :255–258.

14. MacFadden DR, McGeer A, Athey T, Perusini S, Olsha R, Li A, et al. Use of Genome

Sequencing to Define Institutional Influenza Outbreaks, Toronto, Ontario, Canada,

2014–15. Emerg Infect Dis [Internet]. **2018** [cited 2018 May 14]; 24(3):492–497.

Available from: http://wwwnc.cdc.gov/eid/article/24/3/17-1499_article.htm

15. Meinel DM, Heinzinger S, Eberle U, Ackermann N, Schönberger K, Sing · Andreas.

13

Whole genome sequencing identifies influenza A H3N2 transmission and offers superior resolution to classical typing methods. Infection [Internet]. **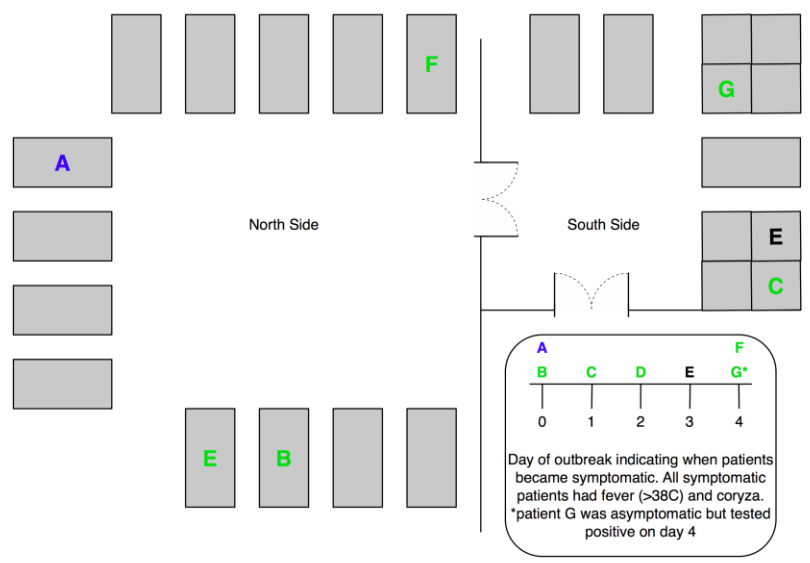123AD** [cited 2018 May 14]; 46(46):69–76. Available from: https://link.springer.com/content/pdf/10.1007%2Fs15010-017-1091-3.pdf

Conflict of interest

All authors declare that they have no conflicts of interest.

14

Figure 1.

North Side | South Side

A | B | C | D | E | F | G*
0 | 1 | 2 | 3 | 4

Day of outbreak indicating when patients became symptomatic. All symptomatic patients had fever (>38C) and coryza. *patient G was asymptomatic but tested positive on day 4

Figure 2.

2.0E-4