Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study

Art F Y Poon, Réka Gustafson, Patricia Daly, Laura Zerr, S Ellen Demlow, Jason Wong, Conan K Woods, Robert S Hogg, Mel Krajden, David Moore, Perry Kendall, Julio S G Montaner, P Richard Harrigan

Summary

Background **HIV evolves rapidly and therefore infections with similar genetic sequences are likely linked by recent transmission events. Clusters of related infections can represent subpopulations with high rates of transmission. We describe the implementation of an automated near real-time system to monitor and characterise HIV transmission hotspots in British Columbia, Canada.**

Methods **In this implementation case study, we applied a monitoring system to the British Columbia drug treatment database, which holds more than 32 000 anonymised HIV genotypes for nearly 9000 residents of British Columbia living with HIV. On average, five to six new HIV genotypes are deposited in the database every day, which triggers an automated reanalysis of the entire database. We extracted clusters of five or more individuals with short phylogenetic distances between their respective HIV sequences. The system generated monthly reports of the growth and characteristics of clusters that were distributed to public health officers.**

Findings **In June, 2014, the monitoring system detected the expansion of a cluster by 11 new cases during 3 months, including eight cases with transmitted drug resistance. This cluster generally comprised young men who have sex with men. The subsequent report precipitated an enhanced public health follow-up to ensure linkage to care and treatment initiation in the affected subpopulation. Of the nine cases associated with this follow-up, all had already been linked to care and five cases had started treatment. Subsequent to the follow-up, three additional cases started treatment and most cases achieved suppressed viral loads. During the next 12 months, we detected 12 new cases in this cluster with reduction in the onward transmission of drug resistance.**

Interpretation **Our findings show the first application of an automated phylogenetic system monitoring a clinical database to detect a recent HIV outbreak and support the ensuing public health response. By making secondary use of routinely collected HIV genotypes, this approach is cost-effective, attains near real-time monitoring of new cases, and can be implemented in all settings in which HIV genotyping is the standard of care.**

Funding **BC Centre for Excellence in HIV/AIDS, the Canadian Institutes for Health Research, the Genome Canada-CIHR Partnership in Genomics and Personalized Health, and the US National Institute on Drug Abuse.**

Lancet HIV 2016; 3: e231–38

Published Online
April 7, 2016
http://dx.doi.org/10.1016/S2352-3018(16)00046-1

See Comment page e196

BC Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada (A F Y Poon PhD, C K Woods BSc, Prof R S Hogg PhD, D Moore MDCM, Prof J S G Montaner MD, Prof P R Harrigan PhD); **Department of Medicine** (A F Y Poon, Prof J S G Montaner, Prof P R Harrigan), **Department of Pathology and Laboratory Medicine** (Prof M Krajden), **University of British Columbia, Vancouver, BC, Canada; Vancouver Coastal Health Authority, Vancouver, BC, Canada** (R Gustafson MD, P Daly MD, L Zerr MN, S E Demlow MPH); **BC Centre for Disease Control, Vancouver, BC, Canada** (J Wong MD, Prof M Krajden MD, D Moore); **Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada** (Prof R S Hogg); **and Office of the Provincial Health Officer, Ministry of Health, Government of British Columbia, Victoria, BC, Canada** (P Kendall FRCPC)

Correspondence to:
Dr Art F Y Poon, BC Centre for Excellence in HIV/AIDS, Vancouver V6Z 1Y6, BC, Canada
apoon@cfenet.ubc.ca

Introduction

Phylogenetic clustering uses virus sequence diversity to investigate hotspots of rapid transmission.[1] For HIV, this approach has been made easier by the introduction of routine genotyping for antiretroviral resistance as standard of care for clinical management of HIV infection.[2,3] A phylogeny is a tree-based representation of how populations are related through common ancestors. For rapidly evolving viruses such as HIV,[4] the phylogeny can resemble the transmission history of the virus,[5] although they will not necessarily be congruent because of the extensive genetic diversity of the virus within each host.[6] Thus, a cluster of HIV sequences from different infections that retain a high degree of genetic similarity are likely to be related by recent transmission events.[7] Routine HIV genotypes can be sufficiently divergent for the reconstruction of phylogenies with informative clustering, even though these sequences tend to target a quite conserved *pol* region of the HIV genome that encodes protease, and reverse transcriptase, the primary targets of antiretroviral drug regimens.[8] Thus, several studies[9–12] have used phylogenetic clustering for a one-time analysis of HIV *pol* datasets to retrospectively characterise potential correlates of high transmission rates of HIV, such as treatment history, stage of infection, and routes of transmission.

Because HIV sequence-based genotypes are constantly being collected at centres of HIV treatment, there is growing interest in using this resource to inform real-time HIV prevention and control measures.[13–15] The BC Centre for Excellence in HIV/AIDS (CFE) is responsible for all HIV genotyping and distribution of antiretroviral therapy for all British Columbia residents living with HIV, provided at no cost to patients. Since 2013, the CFE has implemented and used an automated system for monitoring HIV transmission hotspots in the province by the phylogenetic clustering analysis of routine HIV genotype data. Combined with a short turnaround time for sample processing, this system can provide near real-time information about the growth of phylogenetic clusters. Here we describe our implementation of a phylogenetic monitoring system in British Columbia,

Research in context

Evidence before this study

We searched the Google Scholar database on Oct 19, 2015, for similar studies using the search terms "HIV", "genotype", "clustering", and "monitoring" with no date limits. Phylogenetic trees reconstructed from HIV sequence data bear similar shapes to recent transmission histories of the virus because of its rapid evolution, even in the quite conserved regions of its genome that encode the main targets of antiretroviral treatment. Furthermore, several studies have shown the utility of clustering methods applied to HIV sequence data to retrospectively characterise the variation in rates of HIV transmission between subpopulations in association with risk factors. These results suggest that routinely collected HIV genotype data can be used to track the progress of an epidemic in near real time.

Added value of this study

We have implemented an automated HIV monitoring system in British Columbia, Canada, which used the provincial HIV drug treatment population database to track the growth of phylogenetic clusters of closely related HIV infections in near real time. Here, we report a case study of the use of reports from this monitoring system in direct support of an HIV outbreak investigation and enhanced public health follow-up, which was associated with a marked reduction in transmitted HIV drug resistance in the affected subpopulation.

Implications of all the available evidence

Our findings and those of previous studies show that near real-time monitoring based on the clustering analysis of routinely collected HIV genotypes can become an effective and cost-effective resource for public health intervention on localised outbreaks of HIV transmission.

and present a case study of an outbreak of transmitted HIV drug resistance detected by the system, which precipitated a formal outbreak investigation and enhanced public health follow-up of the subpopulation affected by the outbreak.

Methods

Population database

All data from the British Columbia drug treatment programme at the CFE is stored on site in a secure Oracle database (enterprise version 11g). The monitoring system was developed around a cached database query that integrates de-identified clinical, demographic, epidemiological, and genetic data for all individuals enrolled in the drug treatment programme. The clinical data included: the sample collection dates, date of starting antiretroviral therapy, plasma viral loads (HIV RNA copies per mL), and HIV drug resistance levels as predicted by the VircoTYPE algorithm.[16] We imputed dates of HIV seroconversion with the method we described previously[11] (appendix p 1). Demographic data included sex, birth year, forward sortation area (first three digits of the Canadian postal code) of the physician office that requisitioned the laboratory test, and date of mortality when applicable.[17] Epidemiological data comprised the following risk factors: having ever used injection drugs, men self-identifying as homosexual or bisexual (men who have sex with men), and having received blood (transfusion) products. As of Oct 8, 2015, 32 505 HIV genotype records had been produced by the CFE laboratory programme for 8839 people with HIV—more than half the estimated prevalence of HIV in British Columbia (12 981 people in 2014). 27 850 (86%) genotypes covered 1497 bp of HIV *pol* encoding protease and the first 400 codons of the gene encoding reverse transcriptase; the remaining genotypes were derived from one or more shorter amplicons (appendix p 1). Transmitted drug resistance was inferred at the individual level when an HIV genotype sampled before starting antiretroviral therapy was resistant to one or more drug classes; this does not account for people who might have started treatment outside of British Columbia or had taken part in clinical trials.

See Online for appendix

Data analysis

The monitoring system queries the drug treatment programme database every hour. Whenever new genotype records are detected, the entire contents of the cached query are downloaded to a secure workstation. Patients' identifiers are automatically replaced with random hexadecimal strings with every download to break linkage with the database. HIV genotypes are realigned as aminoacid sequences against the HXB2 *pol* reference sequence (K03455) with a pairwise alignment algorithm.[18] Codons associated with the WHO HIV drug resistance mutations surveillance list[19] and insertions relative to HXB2 are excluded from the alignment. 100 replicate bootstrap alignments are generated by random sampling of codons with replacement and transferred to an in-house computing cluster for phylogenetic reconstruction in parallel by approximate maximum likelihood[20] (appendix p 2).

Phylogenetic clusters, defined at a minimum of five members, are assembled from pairs of individuals that fulfil the following clustering criteria in a minimum of 50 of 100 replicate phylogenies: the total path length separating tips in the phylogeny (patristic distance) is below a cutoff of 0·02 expected nucleotide substitutions

per site;[11] the tips correspond to HIV sequences from different individuals; and at least one of the sequences is derived from the individual's earliest available sample. We used a custom tree traversal algorithm implemented in Python to extract patristic distances efficiently[11] (appendix p 3). To recognise clusters between analyses in the absence of fixed patients' identifiers, we used the baseline sample collection dates of its first five members to index the cluster against a persistent register (appendix pp 3–4). Clusters were annotated with clinical and demographic variables, and visualised with a force-directed layout algorithm in GraphViz (version 2.38.0) with forces proportional to the phylogenetic distance between members. Reports were generated by populating a LaTeX template document with cluster statistics and diagrams with the Jinja2 templating engine in Python; a mock-up report is provided in the appendix (pp 7–17).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Since April, 1999, the British Columbia drug treatment programme population database has consistently accumulated an average of 5·7 HIV genotypes every day from the CFE laboratory programme. More than 32 000 genotypes have been produced for more than 8800 people with HIV in British Columbia (figure 1). The sample processing time, defined as the number of days between the date of an HIV genotype test requisition and the date that the test result was uploaded into the drug treatment programme database, has decreased significantly since the inception of the programme from a median of 14 days (IQR 11–19) before 2005, to a median of 6 days (IQR 4–7) from 2010 onwards (figure 1). Furthermore, current British Columbia treatment guidelines stipulate that an HIV genotype test is automatically done on the residual blood plasma from the baseline viral load test for every new HIV diagnosis. On the basis of data from the Seek and Treat for Optimal Prevention of HIV/AIDS (STOP) cohort in British Columbia,[21] the median delay from HIV diagnosis to viral load testing was 26 days (IQR 13–104).

Monthly and quarterly reports of the growth and characteristics of active clusters being tracked by our monitoring system have been distributed by the CFE since February, 2014, to the British Columbia Centre for Disease Control and medical health officers at the five regional health authorities of British Columbia. A phylogenetic cluster is deemed active if one or more new cases is detected in the drug treatment programme database within the reporting period. These reports were reviewed and discussed at monthly meetings between knowledge users in public health and laboratory staff from the CFE. As of Oct 8, 2015, 218 phylogenetic clusters, each comprising five or more individuals, have been detected and tracked by the monitoring system. The most active clusters (table) were typically in men who have sex with men, with substantial variation among clusters in age distributions and prevalence of transmitted drug resistance.

Diagrams automatically generated by the monitoring system for every active cluster within the reporting period are designed to inform the prioritisation of clusters for public health actions (figures 2, 3). For example, nodes in the network diagram are scaled to each individual's most recent plasma viral load. Cluster 0 was the largest phylogenetic cluster in British Columbia largely comprising people who use injection drugs in the Vancouver downtown eastside.[23] Cluster 0 has one of the most widespread distributions of all phylogenetic clusters (figure 3), partly because of the cluster's size (n=414 as of Oct 8, 2015) and long history of expansion (figure 2B).

For the **Jinja2 engine** see http://jinja2.pocoo.org

In June, 2014, the monitoring system determined that cluster 55 had grown by 11 new cases during 3 months relative to the respective sample collection dates (figure 4). Eight new cases carried the transmitted drug resistance mutation Lys103Asn in the HIV reverse transcriptase gene, which confers resistance to first generation non-nucleotide reverse transcriptase inhibitors, and a median viral load of 4·9 $\log_{10}$ copies per mL. All but one of the individuals had accessed care within Vancouver Coastal Health regional health authority. A provisional report on eight of the 11 new

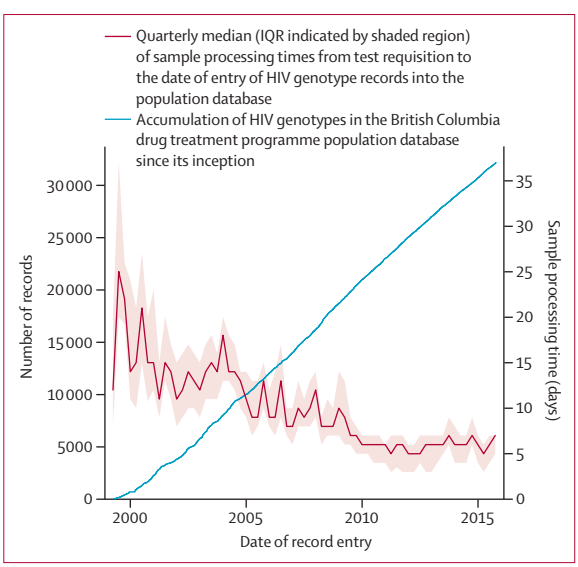

Figure 1: **Number, date, and processing time of routine HIV resistance genotyping in in the British Columbia drug treatment programme, 1999–2015**

By Oct 8, 2015, more than 32 000 HIV genotypes have been deposited in the British Columbia drug treatment programme database.

	New cases	Current size	MSM	IDU	On therapy	TDR*	PCS=0†	Age ≤30 years‡
0	11	414	10/280 (4%)	245/294 (8%)	214/325 (66%)	11/204 (4%)	30/313 (10%)	71/325 (21%)
1	7	84	48/52 (92%)	10/59 (17%)	64/78 (82%)	1/78 (1%)	22/78 (28%)	9/79 (11%)
3	9	86	45/51 (88%)	9/57 (16%)	65/79 (82%)	1/76 (1%)	25/78 (32%)	12/84 (14%)
7	23	99	45/48 (94%)	7/56 (12%)	68/88 (77%)	0/81	42/87 (48%)	31/88 (35%)
17	4	48	24/30 (80%)	6/34 (18%)	31/46 (67%)	0/43	7/44 (16%)	5/45 (11%)
49	5	26	11/12 (92%)	3/12 (25%)	21/24 (88%)	0/23	15/24 (62%)	11/24 (46%)
55	12	49	28/29 (97%)	3/31 (10%)	38/49 (78%)	17/48 (35%)	22/49 (45%)	30/49 (61%)
200	4	18	8/10 (80%)	1/13 (8%)	10/15 (67%)	0/13	1/14 (7%)	0/15
203	7	9	9/9 (100%)	0/9	8/9 (89%)	8/8 (100%)	3/8 (38%)	6/9 (67%)
217	5	8	4/5 (80%)	1/7 (14%)	8/8 (100%)	0/8	0/8	4/8 (50%)

Data are n/N (%). MSM=men who have sex with men. IDU=injecting drug users. TDR=transmitted HIV drug resistance. PCS=programmatic compliance score.
*Before treatment. †A PCS of 0 indicates complete adherence to the International Antiviral Society-USA treatment guidelines.[17] ‡At first sample collection date.

Table: **Summary statistics of the ten most active phylogenetic clusters from October, 2014, to October, 2015**

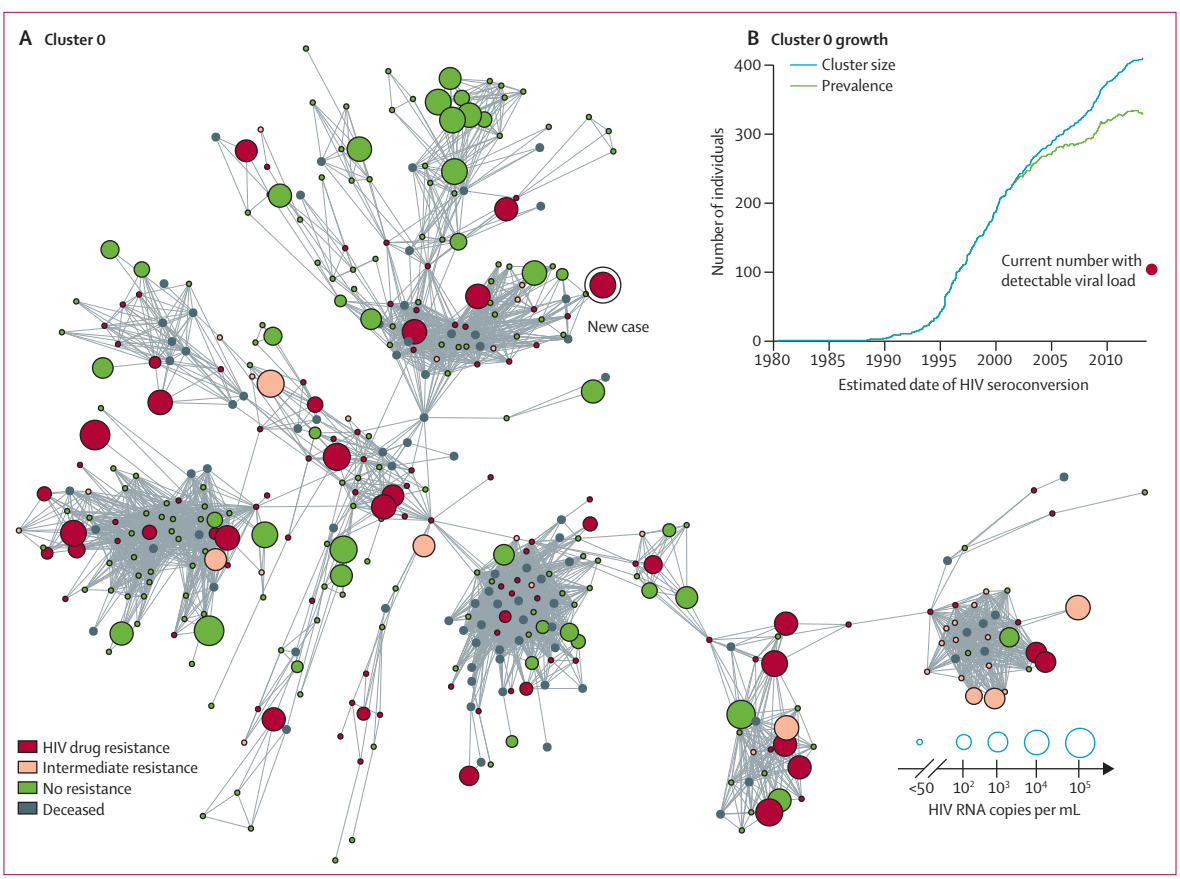

Figure 2: **Network diagram and growth trend for cluster index 0**
Diagrams were generated for the phylogenetic monitoring reports. The largest phylogenetic cluster from the British Columbia drug treatment programme database, cluster 0 is largely composed of people who use injection drugs in the Vancouver downtown eastside. (A) Each circle in the network diagram corresponds to a person living with HIV, sized in proportion to their most recent viral load, and coloured to indicate whether any HIV genotype tests for that individual have been classified with high (red) or intermediate (orange) HIV drug resistance, or mortality (grey). This annotation scheme emphasises parts of the network in which transmissions are most likely to occur, notwithstanding individuals who have not yet been sampled. New cases within the reporting period are indicated by a double outline. Lines drawn between circles indicate that the shortest distance separating HIV sequences from the respective individuals in the phylogenies fell within the clustering threshold. (B) The blue line represents the growth of cluster 0 based on the imputed dates of HIV seroconversion. Prevalence (green) was estimated by subtracting deceased cases from the growth trend. A red circle indicates the estimated number of cases for which the most recent viral loads were detectable, on the basis of the available data at the end of the reporting period.

cases in cluster 55 was issued on June 24, 2014 (figure 4), a week in advance of the scheduled quarterly report, to the British Columbia Centre for Disease Control and Vancouver Coastal Health. The baseline samples from the remaining three new cases had not yet been processed by the date of the provisional report. The rapid spread of transmitted drug resistance identified by this provisional report was deemed a sufficient public health concern that the Provincial Health Officer of British Columbia, under the authority of the Public Health Act, issued a formal request to the Chief Medical Health Officer of Vancouver Coastal Health to perform a public health investigation of all 36 individuals within cluster 55 who had accessed care within Vancouver Coastal Health.

Because HIV is a reportable disease in British Columbia, the corresponding HIV diagnoses were already known to Vancouver Coastal Health public health. For the eight new cases that appeared in the CFE database from April to June, 2014, the dates of HIV diagnoses ranged from February, 2014, to June, 2014. Of the 36 individuals, 33 unique individuals were identified (three individuals had laboratory tests requisitioned under two different names), of which 27 were already known to Vancouver Coastal Health public health, including all new cases with transmitted drug resistance, either in the sexually transmitted infection surveillance database or in the primary access regional information system. All of the known cases were men who have sex with men, with a median age at diagnosis of 27 years (range 19–56), of whom 12 (44%) had been diagnosed with acute stage disease, as defined by a laboratory testing algorithm and testing history,[24] substantially greater than the average proportion of acute stage diagnoses in the community since 2012.

During this investigation, an additional new case with transmitted drug resistance mutation was reported in cluster 55. An enhanced public health follow-up was initiated on July 2, 2014, for the nine new cases in cluster 55 (figure 4). Two additional cases without HIV drug resistance subsequently were reported in cluster 55 (July 2 and July 7) that were not included in this follow-up. The objective of the enhanced follow-up was to ensure linkage to care and initiation of antiretroviral therapy, and to complete any outstanding partner notification, testing, and linkage to care (appendix p 4). Before July 2, only five of the nine individuals were on treatment, of which only one had attained a suppressed viral load; the median viral load was 38 109 copies per mL. By Sept 5, 2014, eight of the individuals had started antiretroviral therapy and six had attained undetectable viral loads, with a median viral load of less than 40 copies per mL (below the limit of detection; figure 4). All nine individuals had received partner counselling and referral services at the time of their HIV diagnosis. During the enhanced follow-up, seven individuals were willing to re-engage with public health for further partner counselling and notification. This re-engagement resulted in 12 additional contacts (five known, seven anonymous) in addition to the 37 contacts (17 known, 20 anonymous) from the initial partner counselling and referral services. The number of known contacts who were known to be previously HIV positive increased from two to six individuals. Of these four previously positive contacts, one individual who was named as a contact of an acute case of HIV had been disengaged from care and involved in behaviours with high risk of onward HIV transmission. This individual was re-engaged in care as a result of the enhanced follow-up. Subsequent to the start of the enhanced follow-up, cluster 55 was quite inactive until the end of 2014 with only one additional case (Aug 14, 2014; figure 4); however, 12 new cases gradually came to light in the next year, of which three carried drug resistance (figure 4).

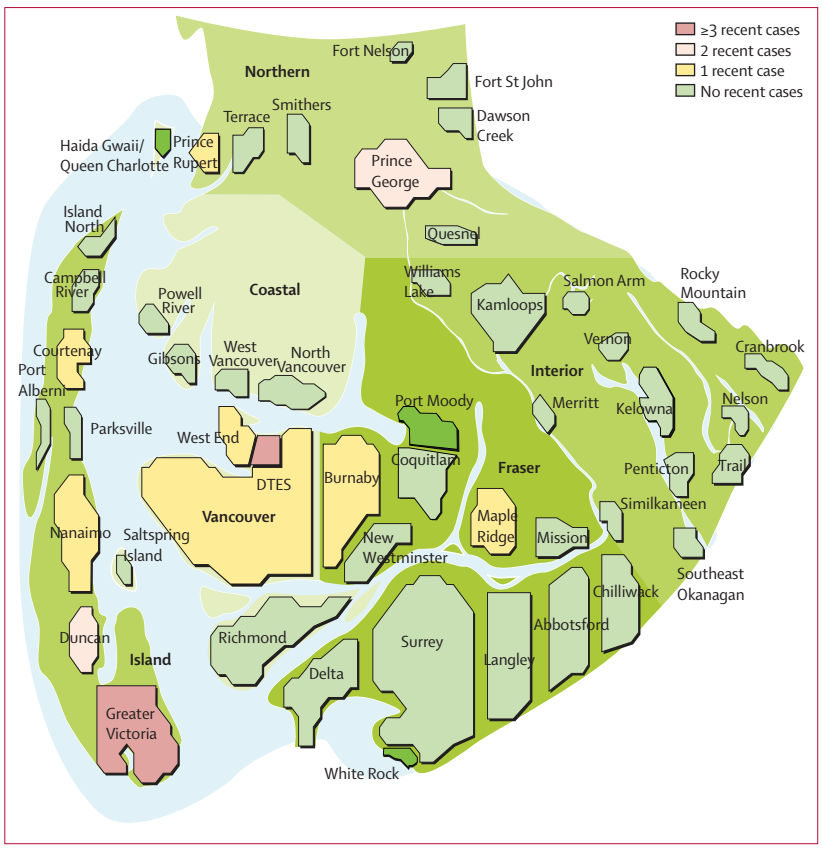

Figure 3: **Geographic distribution of cluster 0**
This simplified map of the province of British Columbia was generated automatically by the monitoring system for every active cluster within the reporting period. The map is distorted to emphasise regions with higher population densities. In this example, cases in cluster 0 were defined as recent when their earliest database entry date was in 2014 or later. The distribution of recent cases among regions, on the basis of the forward sortation areas of the physicians' offices where individuals have most recently accessed primary care, is indicated by the coloration of the respective polygons. Bold labels indicate the five regional health authorities of British Columbia. Adapted with permission from the Human Early Learning Partnership and the Western Geographical Press.[22]

Discussion

The enhanced public health follow-up of the cluster 55 outbreak was followed by suppression of viral loads in most affected individuals and several months without

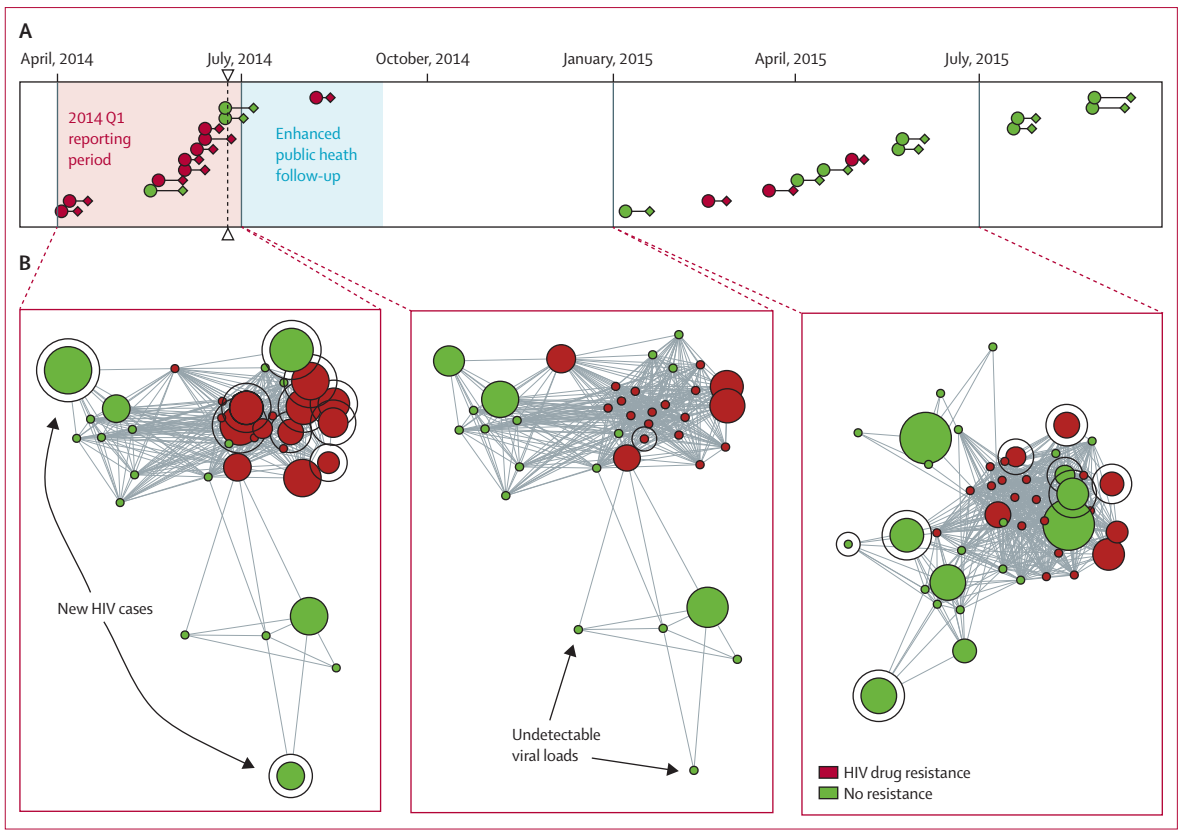

Figure 4: **Timeline of cluster 55 outbreak**
New cases are mapped to their respective dates of sample collection (circle) and HIV genotype database entry (diamond), and coloured with respect to HIV drug resistance; dashed line indicates the issuing of a provisional report on an outbreak of eight new cases in the cluster 55. The network diagrams summarise the phylogenetic relationships among the new cases (encircled) and their most recent viral loads by the end of the respective reporting periods. The HIV genotype connecting an individual to others in the network and their most recent viral load were not necessarily derived from the same sample.

new cases. It is difficult to attribute a causal effect to this intervention because no control group was undergoing a similar outbreak in which public health action was withheld. However, nine of the 12 new HIV cases that were reported in cluster 55 since January, 2015, were non-resistant strains mapping to parts of the HIV phylogeny that were not targeted by the enhanced public health follow-up. This finding suggests that the follow-up partly fulfilled its primary objective of preventing the onward transmission of HIV drug resistance, although it was not completely successful in preventing further transmissions in the cluster. Although this follow-up concluded in 2014, cluster 55 has remained a focus of public health actions.

The use of routinely collected clinical HIV data to inform public health interventions raises important ethical considerations. We have used the ethics framework of public health to establish how this information should be collected and used for HIV prevention measures. The guiding rule of this framework is to use the least intrusive but most effective intervention available. Because the use of phylogenetics is an emerging area in public health, there are no best practice guidelines for using this information. HIV is a reportable disease in British Columbia. Under the British Columbia Public Health Act Communicable Disease Regulation, a medical health officer can request personally identifying information from a physician when there is reasonable likelihood of HIV transmission. The primary objective of public health management is to prevent onward transmission of HIV by reaching populations at risk, not to ascribe transmission to specific individuals. Accordingly, we used the phylogenetic cluster defined at a minimum size of five individuals as an operative unit for public health surveillance. When cluster 55 was deemed a significant public health concern, all individuals in the cluster were identified in accordance with the provisions of the communicable disease regulation. Minimum information for all individuals in the cluster was securely transmitted in an encrypted format to Vancouver Coastal Health public health, which was already aware of all people diagnosed with HIV in the region. The identification of groups to offer counselling, testing, and treatment in rapidly growing phylogenetic clusters is consistent with the aims of HIV prevention and discourages the attribution of fault with any one individual.

One of the general limitations to monitoring HIV transmission hotspots from routine genotyping is that it is restricted to populations with members who present for HIV testing and have a plasma viral load test. Under current provincial guidelines, a genotype test is automatically done with the residual blood plasma from the baseline viral load tests for all new HIV diagnoses. We estimate that more than half of all residents of British Columbia living with HIV have HIV genotype test records in the drug treatment programme database. This sampling density is well within recommended levels for the reproducible identification of clusters.[25] However, individuals who are untested or disengaged from care are likely under-represented in the database population. Findings from a meta-analysis suggested that the undiagnosed population can contribute disproportionately to the onward transmission of HIV;[26] hence, the effect of real-time monitoring might be contingent on optimising all stages in HIV care. Furthermore, the time between HIV infection and diagnosis inevitably varies substantially between individuals. Despite these unavoidable limitations, we have shown that phylogenetic monitoring can supplement standard epidemiological methods and inform public health actions on a timescale that is sufficient to potentially alter the course of a localised outbreak.

Another challenge for the implementation of HIV phylogenetic monitoring is finding the balance between protecting an individual's right to privacy and right to refuse medical care, and the public health responsibility to prevent the onward transmission of HIV. This dilemma is exacerbated by the criminalisation of HIV exposure[27] or transmission,[28] because the same phylogenetic methods used for characterising transmission hotspots have also been misused to prosecute individuals for the transmission of HIV.[29] It is further compounded by the widespread use of the terms "transmission network" or "transmission cluster" to refer to genetically similar virus populations, which implicitly equates a phylogenetic relationship with a transmission event. On the contrary, a molecular phylogeny cannot establish whether the virus was transmitted directly from one individual to another because of the extensive diversity of HIV within hosts[6] and the potential for transmissions through unknown third parties. Furthermore, a phylogeny cannot establish the directionality of HIV transmission without extensive clonal sequencing of the virus populations.[30]

The ubiquity of routine HIV genotyping in developed settings, where the comparably low but highly heterogeneous prevalence of HIV poses significant challenges for the cost-effective deployment of HIV prevention resources, is strong motivation for near real-time monitoring through the secondary analysis of data already being collected as standard of care. We have presented a case in which our monitoring system prioritised a specific outbreak of transmitted HIV drug resistance for public health intervention. Such actions will ultimately become important for supporting targeted HIV prevention efforts and preserving treatment options for the population, and could translate to other areas of infectious disease.

Contributors

AFYP developed the monitoring system, analysed the monitoring data, generated all figures, and wrote the report. RG and PD led the formal outbreak investigation. LZ did the enhanced public health follow-up. SED analysed data from the follow-up. AFYP, RG, JW, MK, PK, DM, JSGM, and PRH reviewed monitoring reports and developed the reporting template. RSH directed the drug treatment programme. CKW did database programming. JSGM and PRH designed the study.

Declaration of interests

We declare no competing interests.

Acknowledgments

This work was supported by the BC Centre for Excellence in HIV/AIDS and by grants from the Canadian Institutes for Health Research (CIHR HOP-111406, HOP-107544), the Genome Canada-CIHR Partnership in Genomics and Personalized Health (Large-Scale Applied Research Project HIV142 contract to PRH, JSGM, and AFYP), and by the US National Institute on Drug Abuse (1-R01-DA036307-01, 5-R01-031055-02, R01-DA021525-06, and R01-DA011691). JSGM is supported with grants paid to his institution by the British Columbia Ministry of Health. He has also received limited unrestricted funding, paid to his institution, from Abbvie, Bristol-Myers Squibb, Gilead Sciences, Janssen, Merck, and ViiV Healthcare. AFYP is supported by a Career Investigator Award from the Michael Smith Foundation for Health Research, in partnership with the Providence Health Care Research Institute and the St Paul's Hospital Foundation, and by a CIHR New Investigator award. PRH is supported by a CIHR/GlaxoSmithKline Research Chair in clinical virology.

References

1 Yirrell DL, Robertson P, Goldberg DJ, McMenamin J, Cameron S, Leigh Brown AJ. Molecular investigation into outbreak of HIV in a Scottish prison. *BMJ* 1997; **314:** 1446–50.

2 Günthard HF, Aberg JA, Eron JJ, et al, and the International Antiviral Society-USA Panel. Antiretroviral treatment of adult HIV infection: 2014 recommendations of the International Antiviral Society-USA Panel. *JAMA* 2014; **312:** 410–25.

3 Tural C, Ruiz L, Holtzer C, et al, and the Havana Study Group. Clinical utility of HIV-1 genotyping and expert advice: the Havana trial. *AIDS* 2002; **16:** 209–18.

4 Williamson S. Adaptation in the *env* gene of HIV-1 and evolutionary theories of disease progression. *Mol Biol Evol* 2003; **20:** 1318–25.

5 Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc Natl Acad Sci USA* 1996; **93:** 10864–69.

6 Romero-Severson E, Skar H, Bulla I, Albert J, Leitner T. Timing and order of transmission events is not directly reflected in a pathogen phylogeny. *Mol Biol Evol* 2014; **31:** 2472–82.

7 Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 2008; **5:** e50.

8 Hué S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS* 2004; **18:** 719–28.

9 Brenner BG, Roger M, Routy JP, et al, and the Quebec Primary HIV Infection Study Group. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis* 2007; **195:** 951–59.

10 Drescher SM, von Wyl V, Yang WL, et al, and the Swiss HIV Cohort Study. Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who have sex with men in the Swiss HIV Cohort Study. *Clin Infect Dis* 2014; **58:** 285–94.

11 Poon AF, Joy JB, Woods CK, et al. The impact of clinical, demographic and risk factors on rates of HIV transmission: a population-based phylogenetic analysis in British Columbia, Canada. *J Infect Dis* 2015; **211:** 926–35.

12 Wertheim JO, Leigh Brown AJ, Hepler NL, et al. The global transmission network of HIV-1. *J Infect Dis* 2014; **209:** 304–13.

13 Little SJ, Kosakovsky Pond SL, Anderson CM, et al. Using HIV networks to inform real time prevention interventions. *PLoS One* 2014; **9:** e98443.

14 Smith DM, May SJ, Tweeten S, et al. A public health model for the molecular surveillance of HIV transmission in San Diego, California. *AIDS* 2009; **23:** 225–32.

15 Ragonnet-Cronin M, Ofner-Agostini M, Merks H, et al. Longitudinal phylogenetic surveillance identifies distinct patterns of cluster dynamics. *J Acquir Immune Defic Syndr* 2010; **55:** 102–08.

16 Vermeiren H, Van Craenenbroeck E, Alen P, Bacheler L, Picchio G, Lecocq P, and the Virco Clinical Response Collaborative Team. Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. *J Virol Methods* 2007; **145:** 47–55.

17 Lima VD, Le A, Nosyk B, et al. Development and validation of a composite programmatic assessment tool for HIV therapy. *PLoS One* 2012; **7:** e47859.

18 Gotoh O. An improved algorithm for matching biological sequences. *J Mol Biol* 1982; **162:** 705–08.

19 Shafer RW, Rhee SY, Pillay D, et al. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS* 2007; **21:** 215–23.

20 Price MN, Dehal PS, Arkin AP. FastTree 2: approximately maximum-likelihood trees for large alignments. *PLoS One* 2010; **5:** e9490.

21 Heath K, Samji H, Nosyk B, et al, and the STOP HIV/AIDS Study Group. Cohort profile: Seek and treat for the optimal prevention of HIV/AIDS in British Columbia (STOP HIV/AIDS BC). *Int J Epidemiol* 2014; **43:** 1073–81.

22 Kershaw PW, and the Human Early Learning Partnership. (Vancouver BC). The British Columbia atlas of child development, 1st edn. British Columbia: Human Early Learning Partnership; 2005.

23 Harvey E, Strathdee SA, Patrick DM, et al. A qualitative investigation into an HIV outbreak among injection drug users in Vancouver, British Columbia. *AIDS Care* 1998; **10:** 313–21.

24 Selik RM, Mokotoff ED, Branson B, Owen SM, Whitmore S, Hall IH, and the Centers for Disease Control and Prevention (CDC). Revised surveillance case definition for HIV infection—United States, 2014. *MMWR Recomm Rep* 2014; **63:** 1–10.

25 Novitsky V, Moyo S, Lei Q, DeGruttola V, Essex M. Impact of sampling density on the extent of HIV clustering. *AIDS Res Hum Retroviruses* 2014; **30:** 1226–35.

26 Hall HI, Holtgrave DR, Maulsby C. HIV transmission rates from persons living with HIV who are aware and unaware of their infection. *AIDS* 2012; **26:** 893–96.

27 Galletly C, Lazzarini Z, Sanders C, Pinkerton SD. Criminal HIV exposure laws: moving forward. *AIDS Behav* 2014; **18:** 1011–13.

28 Burris S, Cameron E. The case against criminalization of HIV transmission. *JAMA* 2008; **300:** 578–81.

29 Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992; **256:** 1165–71.

30 Rachinger A, Groeneveld PH, van Assen S, Lemey P, Schuitemaker H. Time-measured phylogenies of gag, pol and env sequence data reveal the direction and time interval of HIV-1 transmission. *AIDS* 2011; **25:** 1035–39.