HIV transmission networks among transgender women in Los Angeles County, CA, USA: a phylogenetic analysis of surveillance data

Manon Ragonnet-Cronin, Yunyin W Hu, Sheldon R Morris, Zhijuan Sheng, Kathleen Poortinga, Joel O Wertheim

Summary

Background Transgender women are among the groups at highest risk for HIV infection, with a prevalence of 27·7% in the USA; and despite this known high risk, undiagnosed infection is common in this population. We set out to identify transgender women and their partners in a molecular transmission network to prioritise public health activities.

Methods Since 2006, HIV protease and reverse transcriptase gene (*pol*) sequences from drug resistance testing have been reported to the Los Angeles County Department of Public Health and linked to demographic data, gender, and HIV transmission risk factor data for each case in the enhanced HIV/AIDS Reporting System. We reconstructed a molecular transmission network by use of HIV-TRAnsmission Cluster Engine (with a pairwise genetic distance threshold of 0·015 substitutions per site) from the earliest *pol* sequences from 22 398 unique individuals, including 412 (2%) self-identified transgender women. We examined the possible predictors of clustering with multivariate logistic regression. We characterised the genetically linked partners of transgender women and calculated assortativity (the tendency for people to link to other people with the same attributes) for each transmission risk group.

Findings 8133 (36·3%) of 22 398 individuals clustered in the network across 1722 molecular transmission clusters. Transgender women who indicated a sexual risk factor clustered at the highest frequency in the network, with 147 (43%) of 345 being linked to at least one other person (adjusted odds ratio [aOR] 2·0, p=0·0002). Transgender women were assortative in the network (assortativity 0·06, p<0·001), indicating that they tended to link to other transgender women. Transgender women were more likely than expected to link to other transgender women (OR 4·65, p<0·001) and cisgender men who did not identify as men who have sex with men (MSM; OR 1·53, p<0·001). Transgender women were less likely than expected to link to MSM (OR 0·75, p<0·001), despite the high prevalence of HIV among MSM. Transgender women were distributed across 126 clusters, and cisgender individuals linked to one transgender woman were 9·2 times more likely to link to a second transgender woman than other individuals in the surveillance database. Reconstruction of the transmission network is limited by sample availability, but sequences were available for more than 40% of diagnoses.

Interpretation Clustering of transgender women and the observed tendency for linkage with cisgender men who did not identify as MSM, shows the potential to use molecular epidemiology both to identify clusters that are likely to include undiagnosed transgender women with HIV and to improve the targeting of public health prevention and treatment services to transgender women.

Funding California HIV and AIDS Research Program and National Institutes of Health–National Institute of Allergy and Infectious Diseases.

Lancet HIV 2019

Published Online
February 11, 2019
http://dx.doi.org/10.1016/S2352-3018(18)30359-X

See Online/Comment
http://dx.doi.org/10.1016/S2352-3018(18)30365-5

Department of Medicine, University of California, San Diego, CA, USA (M Ragonnet-Cronin PhD, S R Morris MD, J O Wertheim PhD); **MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, UK** (M Ragonnet-Cronin); **and Division of HIV and STD Programs, Department of Public Health, Los Angeles, CA, USA** (Y W Hu MPH, Z Sheng MSc, K Poortinga MPH)

Correspondence to:
Dr Manon Ragonnet-Cronin, MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London SW7 2AZ, UK
manonragonnet@imperial.ac.uk

Introduction

The global HIV-1 pandemic is driven by geographical, gender, and socioeconomic disparities.[1] In Europe and North America, HIV burden is concentrated among marginalised and stigmatised populations, including sexual minorities and communities of colour. In the USA, men who have sex with men (MSM) are disproportionately affected, comprising 62% of HIV-1 diagnoses each year,[2] and African Americans account for 44% of infections despite making up only 14% of the population.[3] Transgender women (ie, individuals assigned male sex at birth but who identify as women) are estimated to have an HIV prevalence of 27·7%,[4] even higher than the 25% prevalence estimated for MSM.[5] Of concern, African American transgender women have even higher prevalence, reaching 56·3%.[4] In parallel, cross-sectional HIV testing in Los Angeles, Miami, and San Francisco found HIV prevalence of 12% among transgender women with no previous test result, indicating a high frequency of undiagnosed infection.[4–6]

Analysis of viral genetic sequences provides a method for uncovering transmission dynamics.[7] HIV is particularly amenable to phylogenetic analysis because its rapid evolution makes identification of genetic networks (ie, molecular transmission clusters) of densely connected subpopulations possible.[8] These transmission clusters are

Research in context

Evidence before this study

We searched Google Scholar for the terms "HIV" and "genetic" and "transmission" and "networks" on June 4, 2018, then added the term "transgender" to the search, with no date limits. HIV is spread through contacts within a sexual network. The virus accumulates genetic mutations within the same timeframe as transmission events. The transmission history of the virus is a subset of the network that can be reconstructed from HIV genetic sequences (although some transmission events might be missed for various reasons). The structure of the reconstructed transmission network can be informative in terms of risk factors associated with transmission and for interventions. Furthermore, several studies have shown that gaining insights into transmission of HIV among groups that are difficult to investigate using traditional epidemiological tools, such as contact tracing, is possible. We found no molecular epidemiological analyses specific to transgender women, despite these women being one of the groups with the highest prevalence of HIV in the USA. The reasons for this shortfall might include transgender statistics not being collected by departments of public health and small numbers of recorded transgender people in many jurisdictions.

Added value of this study

We reconstructed the HIV transmission network using all HIV sequences available from the Los Angeles County. We identified transgender women within these networks and looked at how they were connected to other risk groups in the network. We found that transgender women were more connected to each other and to heterosexual men and less connected to men who have sex with men than expected. This is the first analysis of genetic transmission networks among transgender women, and the first statistically rigorous examination of the partners of transgender women.

Implications of all the available evidence

The way in which people are connected through the genetic transmission network provides information on transmission patterns within the population. Transmission clusters comprising at least one transgender woman are attractive targets for interventions aimed at finding additional undiagnosed and at-risk transgender women, because individuals within that cluster are more likely to have other transgender women among their sexual or social contact networks. This study highlights the potential for molecular epidemiology to guide interventions towards subpopulations with high HIV prevalence but low diagnosis.

presumed to comprise people at increased risk of HIV transmission or evidence of recent transmission events; however, densely sampled or sequenced subpopulations can also form clusters, even in the absence of increased transmission.[9,10] These clustering approaches can reveal patterns hidden from traditional epidemiological approaches (eg, obscured transmission risk behaviours, such as self-reported heterosexual men whose viruses cluster only with those from MSM).[11,12]

With 119 589 cases, California had the second largest number of people with HIV-1 in the USA as of 2014 (second to New York State with 130 753 cases).[3] Within California, Los Angeles County (LAC) had the greatest number of HIV diagnoses in 2015 and has the largest burden of people with HIV in the state, at 60 000 individuals. The LAC HIV epidemic is dominated by infections in MSM, who account for 83% of recent diagnoses.[2] Since 2006, HIV-1 genetic sequences, generated for routine antiretroviral resistance genotyping, have been reported to the LAC Department of Public Health.

Here, we reconstructed the HIV-1 genetic transmission network from the LAC surveillance database, with a focus on transmission risk among transgender women.

Methods

Data sources

Since 2006, HIV-1 protease and reverse transcriptase (*pol*) genetic sequences generated during routine antiretroviral drug resistance testing have been reported to the LAC Department of Public Health. As of 2016, LAC HIV surveillance had received HIV-1 genetic sequences from 22 398 individuals residing or receiving care in LAC. Of 60 000 people estimated to have HIV, 49 976 had been diagnosed by 2015; thus 44·8% of diagnosed individuals had sequences available. We used the first genotype available for each individual. Information on treatment has been collected since 2006, and 69% of new cases since 2006 were treatment naive at the time of their first genotype. Deduplication of cases is done within the LAC database via a comprehensive procedure based on name, date of birth, address, and social security number.

For each case reported to the local HIV surveillance system, additional clinical and demographic data are available in the enhanced HIV/AIDS Reporting System (eHARS). We define transgender women as people who were assigned male sex at birth but identify as women. In the LAC HIV surveillance database, transgender information was initially collected in a combined sex and gender field (male, female, male-to-female transgender, and female-to-male transgender) starting in the late 1990s. From 2009 onwards, a two-step method was implemented in the HIV/AIDS adult case report form to identify transgender individuals, recording sex at birth alongside current gender identity. There are various data sources for sex and gender information that might include provider reports (as abstracted from medical charts, physician's notes, and self-administered patient intake sheets), laboratory test reports, the Ryan White

Program client registry, and public health investigation by surveillance and partner services staff. Other data available in eHARS include race and ethnicity (Native American or Alaska Native, Asian or Pacific Islander, black or African American, Latino, white, or mixed race), transmission risk factor (MSM, people who inject drugs [PWID], MSM who inject drugs [MSM-PWID], heterosexuality, perinatal exposure to HIV, and other or unknown factors), age at diagnosis, date of diagnosis, CD4 count at diagnosis, and date of last negative test. We treated age at diagnosis as a categorical variable (0–12 years, 13–19 years, 20–29 years, 30–39 years, 40–49 years, 50–59 years, and ≥60 years) and analysed date of diagnosis as a continuous variable using month and year only. When date of last negative test was available and was less than 6 months before a positive HIV test, we classified individuals as having early diagnoses. As a proxy for time since infection for other cases, we used CD4 count (>500 cells per µL, 200–500 cells per µL, and <200 cells per µL).

A transgender woman who reports having had sex with cisgender men might be classified as heterosexual (corresponding to their gender identity) or MSM (corresponding to their birth sex but disregarding their gender identity). Therefore, for transgender women, we collapsed transmission risk factor into two categories: transgender women who reported injection drug use and the remaining transgender women who did not report injecting drugs and were likely to have been infected through sex. We classified this latter group as having a sexual risk factor. To permit meaningful comparison with cisgender men and women (individuals who identify with the sex they were assigned at birth), we categorised all cisgender individuals who reported injection drug use as PWID and those who reported heterosexual risk or no risk were classified as having sexual transmission risk. Individuals who reported perinatal exposure or other transmission risk factors were included in a category termed other. As such, the final risk categories differed from those assigned by HIV surveillance.

The study was approved by both the University of California and San Diego and LAC Department of Public Health institutional review boards.

Phylogenetic analysis

A molecular transmission network was constructed from genetic sequences using HIV-TRAnsmission Cluster Engine (HIV-TRACE).[13] In brief, we aligned HIV *pol* sequences to an HXB2 reference sequence and calculated pairwise genetic distances under the Tamura-Nei 93 model. We did not remove codons associated with drug resistance, given that their removal does not affect clustering when using HIV-TRACE in similar datasets.[14,15] Each individual in the network is represented by a node, and we linked nodes to each other if their pairwise genetic distance was up to 0·015 substitutions per site. This threshold is in line with the expected divergence between sequences within an individual[16] and in accordance with the genetic distance seen between named HIV risk partners.[15] We further tested the sensitivity of our epidemiological inference at distance thresholds of up to 0·01 and up to 0·02 substitutions per site. Nodes linked to at least one other node are classed as clustered in the transmission network. 97% of sequences were subtype B, but HIV-TRACE can create a single network regardless of subtype.

Statistical analysis

Clustered sequences are closely related genetically, indicating that they are likely to be part of the same transmission chain, and high clustering within a population suggests increased transmission. Therefore, we assessed the correlates of clustering using multivariate and univariate logistic regression. We included date of HIV diagnosis, transmission risk group, age at diagnosis, race and ethnicity, CD4 count and early infection, and country of birth (USA or US territories *vs* foreign born) as covariates in the multivariate regression models. Individuals for whom information was missing for one or more of these categories were categorised as unknown. For the purpose of the logistic regression, gender and transmission risk categories were combined into a single variable. As such, our final transmission risk groups were cisgender women with a sexual risk factor, cisgender women who inject drugs, transgender women who inject drugs, transgender women with a sexual risk factor, cisgender men with a sexual risk factor (not including men who reported sex with men, but could include non-disclosed MSM), MSM, MSM-PWID, cisgender men who inject drugs, transgender men, and a further category termed other (appendix p 6). The other category comprised people with perinatal HIV infections and recipients of blood products infected with HIV. Using the same method and covariates, we then assessed the correlates of non-transgender women clustering with transgender women in the transmission network.

See Online for appendix

Assortativity is a network metric that describes for a given characteristic (eg, transmission risk factor) the tendency for nodes to link to other nodes with the same trait (ie, do PWID link to PWID?).[17] Assortativity varies between –1 (completely disassortative) and 1 (completely assortative) and was calculated with the function available in the R igraph package (version 1.2.1).[18] For this analysis, one category included all cisgender people with a sexual risk factor, and another included all cisgender people who report injecting drugs, given that we would expect them to interact with each other.

In parallel, we counted links in the network between each pair of transmission risk groups to estimate mixing patterns between transgender women and other groups.[17] To adjust for degree (ie, the total number of links connecting a given node), the number of links for each individual was divided by that individual's degree. We did this correction because some individuals have

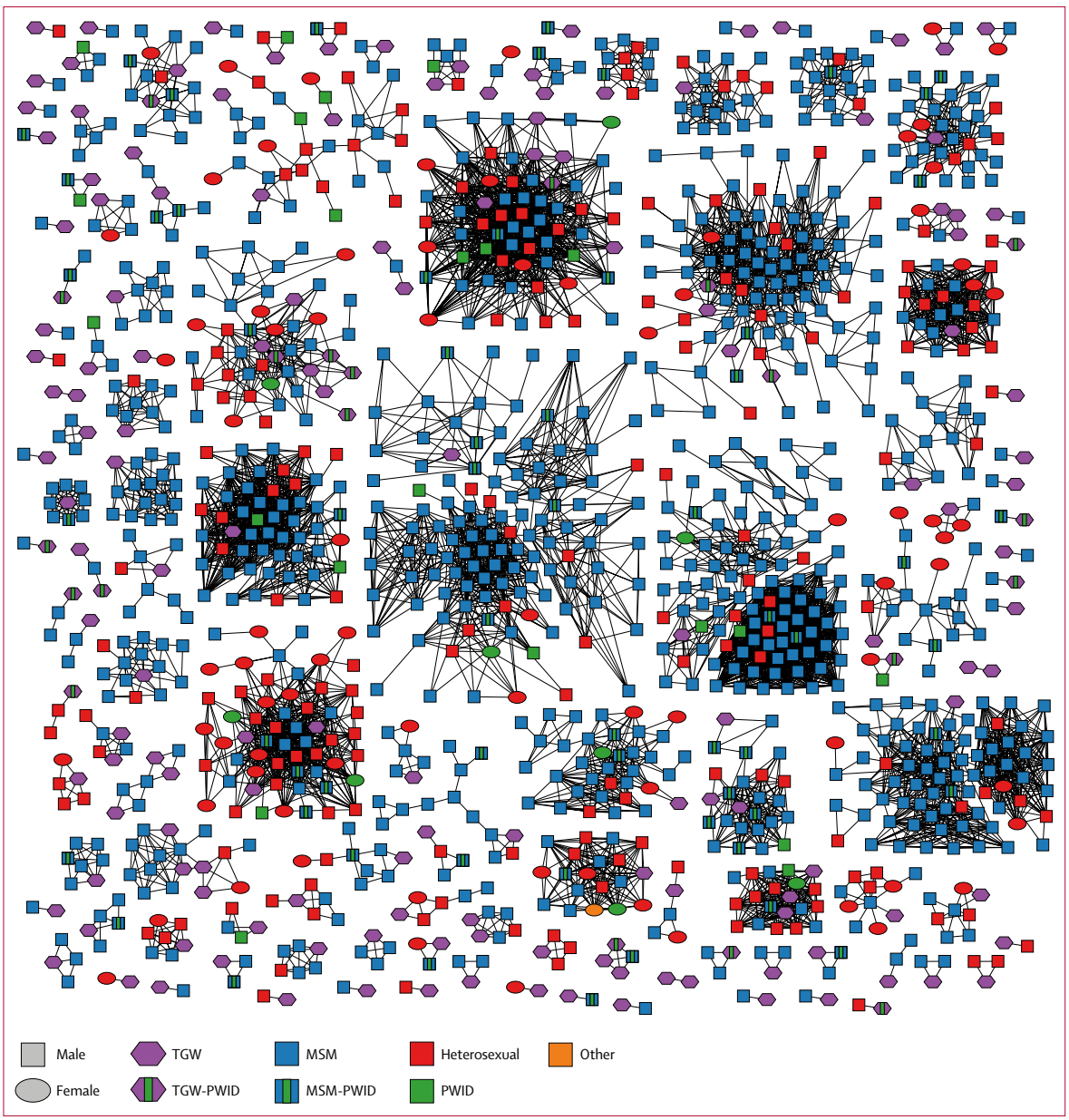

Figure 1: **Molecular transmission clusters in Los Angeles County with at least one transgender woman**
Node shape denotes gender and colour denotes transmission risk factor. Edges represent genetic distance of up to 0·015 substitutions per site. MSM=men who have sex with men. MSM-PWID=men who have sex with men and who inject drugs. PWID=people who inject drugs. TGW=transgender women. TGW-PWID=transgender women who inject drugs.

far more links than others, but we do not wish to overcount those individuals' contribution to mixing between transmission risk groups. Assortativity is influenced by the ratio of node labels (eg, PWID or MSM). To assess the statistical significance of observed patterns of mixing and assortativity given the relative representation of each transmission risk group in these clusters, we generated expected distributions for parameters by randomly permuting transmission risk group labels on the static network 1000 times in R (version 3.4.1).

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 this 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
In the LAC transmission network, 8133 (36·3%) of 22 398 unique individuals were clustered at 0·015 substitutions per site. The network was composed of

1722 clusters comprising between two and 116 nodes (figure 1; appendix p 7). 14 932 (67·0%) of 22 398 sequences in the surveillance database and 5999 (73·7%) of 8133 clustered sequences were from MSM. The LAC dataset contained sequences and demographic data from 412 transgender women, including 67 transgender women who inject drugs. The mean age of transgender women at diagnosis was 29 years (SD 8·7 years) and their mean current age was 40·4 years (10·3 years). Transgender women were less likely to be white than were non-transgender women in the dataset (Fisher's exact test p=0·0004; figure 2). The number of sequences collected and the proportion of sequences clustering each year have increased overall for transgender women with sexual risk factor specifically, but not for transgender women who inject drugs, given that diagnoses among PWID have decreased over time in LAC[2] (appendix p 9).

We sought to establish which demographic and risk characteristics were associated with clustering to identify subpopulations with higher transmission. Transgender women with sexual risk clustered at the highest frequency in the network (42·6% *vs* 40·1% for MSM; figure 3) and had the highest odds of clustering in the univariate analyses (p<0·0001; appendix p 2). In the multivariate analysis, the adjusted odds ratio (aOR) for clustering was even higher for transgender women who inject drugs than for other transgender women and MSM (figure 3). However, the aOR for clustering of transgender women who inject drugs was affected by the date of HIV diagnosis, with 90% of diagnoses having taken place before 2007 (appendix p 9), and consequently their odds of clustering were lower than for transgender women with sexual risk in the univariate analysis (appendix p 2). Individuals diagnosed with a higher CD4 count, who were likely to have been diagnosed closer to the time of infection, were more likely to cluster, but the effect was modest. Individuals with a documented negative HIV test within 6 months before diagnosis, classified as early, were more likely to cluster in the univariate analysis, but this effect was not significant in the multivariate model (appendix p 2; table 1; figure 3). An age trend was apparent, with younger individuals significantly more likely to cluster and older individuals less likely to cluster (figure 3). Individuals of Latino ethnicity were the largest racial and ethnic group (44·2% of population) and the group most likely to cluster. Individuals born outside the USA were less likely to cluster than those born in the USA or US territories. Variables associated with clustering were consistent across genetic distance thresholds (appendix p 3).

We estimated assortativity, the tendency of nodes sharing attributes to link together, by transmission risk group, across the network. The 167 transgender women were distributed across 126 clusters, with 21 clusters containing more than one transgender woman. Although only 503 (2·3%) of 21 986 non-transgender women linked to at least one transgender women in the network, 106 (21·1%)

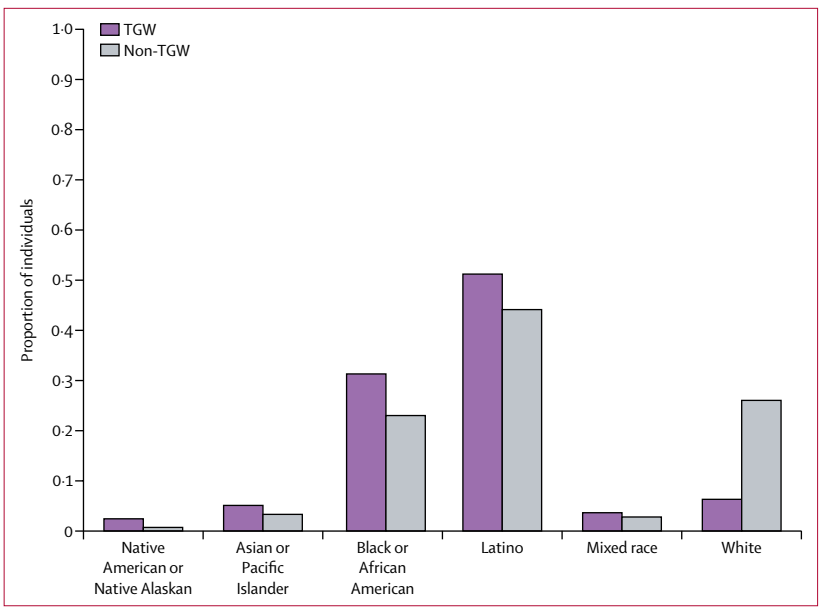

Figure 2: **Race and ethnicity of transgender women and other individuals with sequence data available in the Los Angeles County dataset**
There were 412 transgender women compared with 21 986 non-transgender women. TGW=transgender women.

of 503 of these individuals linked to a second transgender woman (figure 1). Therefore, individuals linked to one transgender woman were 9·2 times more likely to link to two transgender women than other individuals in the surveillance database. MSM, MSM-PWID, people who are cisgender with a sexual risk, and transgender women (with sexual risk or who inject drugs) were all significantly assortative in the network (figure 4). MSM were most likely to link to each other (assortativity coefficient of 0·17, p<0·001). The assortativity coefficient for transgender women with sexual risk was 0·06 (p<0·001; ie, an assortativity coefficient this extreme was not observed in any of the 1000 network permutations); however, absolute assortativity of transgender women was low relative to people who are cisgender with a sexual risk and MSM, because the total number of transgender women in the network is small. By contrast, cisgender PWID did not link assortatively, indicating that they were dispersed among other risk groups in the network. At genetic distance thresholds down to 0·01 and up to 0·02 substitutions per site, MSM, people who are cisgender with sexual risk, and transgender women with sexual risk remained significantly assortative (appendix p 10).

We characterised the subpopulations clustering with transgender women by constructing a linear regression model distinguishing between non-transgender women clustering with transgender women and those that did not. Cisgender men with sexual risk, and men who inject drugs were all more likely than MSM to be clustered with transgender women; foreign-born individuals were less likely to cluster with transgender women than were US born individuals (table 1). There were no significant differences with respect to age or race and ethnicity.

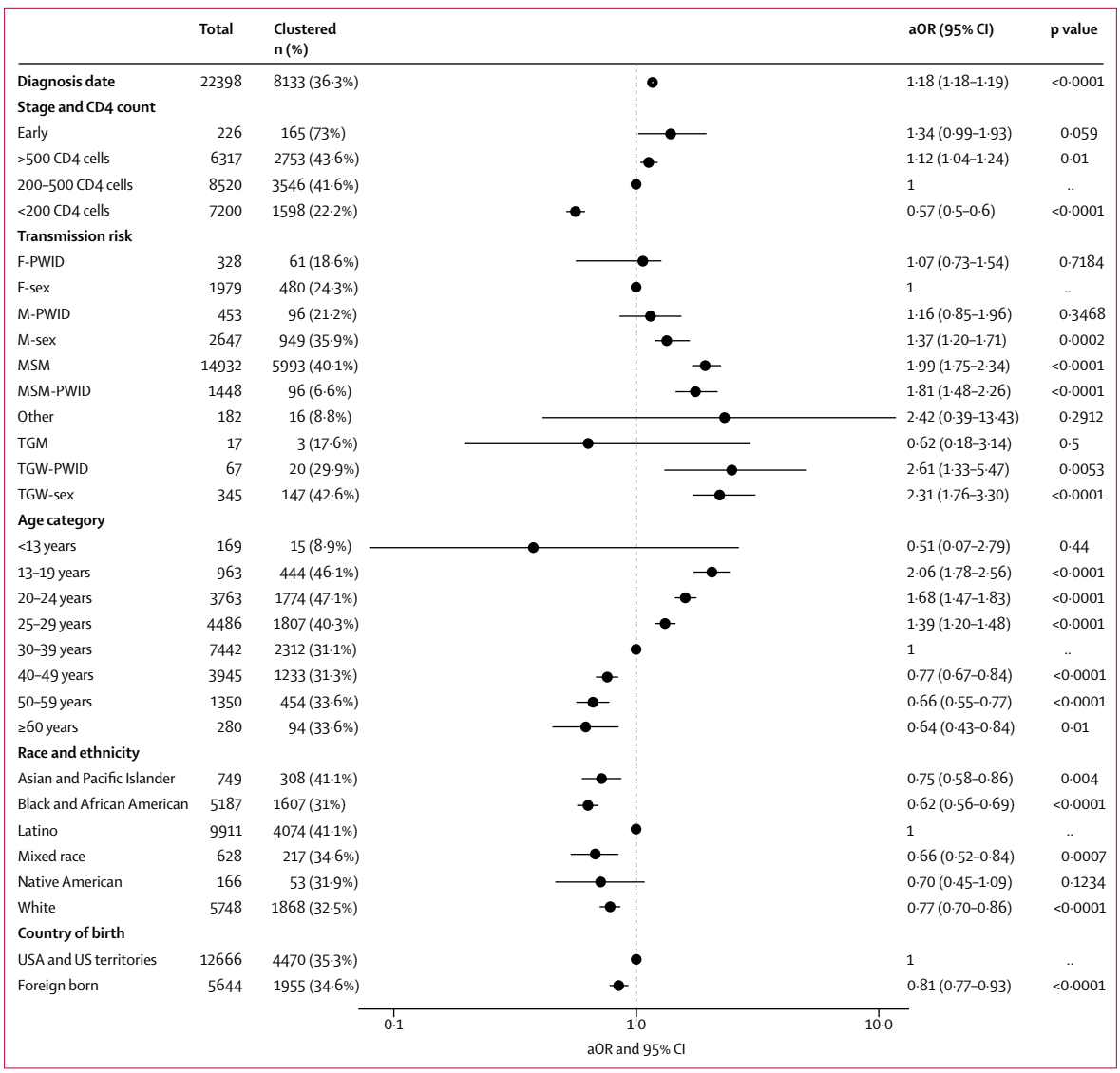

Figure 3: **Demographic breakdown of the people with HIV-1 sequence data in Los Angeles County with adjusted odds ratio for clustering**
The total column indicates the number of individuals in the Los Angeles County surveillance population in that category and the clustered column indicates the number and percentage of individuals in that category who were clustered. The aOR for diagnosis date indicates that individuals diagnosed in each year were 1·18 times more likely to be clustered than individuals sampled in the previous year. Individuals classified as early are those who tested negative for HIV within 6 months before diagnosis. aOR=adjusted odds ratio. F-PWID=cisgender women who inject drugs. F-sex=cisgender women at sexual risk. M-PWID=cisgender men who inject drugs. M-sex=cisgender men at sexual risk. MSM=men who have sex with men. MSM-PWID=men who have sex with men and who inject drugs. TGW-PWID=transgender women who inject drugs. TGW-sex=transgender women at sexual risk.

Finally, we explored the connectivity between each pair of transmission risk groups, adjusting for node degree, to establish who transgender women linked to. The network reconstruction method creates a network in which far more links are present than in the true transmission network. Because the vast majority of nodes in the network represent MSM, we expect high linkage to MSM for all transmission risk groups, and that is indeed what we observed (appendix p 5). Nonetheless, we also observed trends towards transgender women with sexual risk and transgender women who inject drugs linking to each other, as well as with cisgender men with sexual risk. To assess the statistical significance of this observation, we estimated the expected proportion of links between transgender women (either with sexual risk or those who inject drugs) and each of the other risk groups by use of the randomly permuted networks (table 2). For transgender women in either risk group, the proportion of links to other transgender women was higher than expected, whereas the proportion of links to MSM was 25–30% lower than expected. For transgender women with sexual risk, the proportion of links with cisgender men with sexual risk was higher than expected. Nonetheless most transgender

women links (75%) were with MSM. Identical mixing patterns were seen across genetic distance thresholds (appendix p 4–5).

Discussion

We found that transgender women had the highest odds of clustering in the LAC network, indicating that their risk of being in a molecular transmission cluster exceeds that of even MSM. Our findings also reveal that transgender women occupy a distinct position in the LAC transmission network. Transgender women were more likely to be the genetically linked partners of cisgender men not reporting injection drug use or sexual contact with men than expected. Furthermore, transgender women tended to cluster assortatively in the network (ie, having one transgender women in a cluster increased the odds of finding another transgender women in that same cluster).

The patterns of clustering among transgender women observed here suggest a potentially powerful strategy for using the molecular transmission network to improve public health outcomes. Assortativity indicates that non-transgender women who are genetically linked to one transgender woman are nine times more likely to be clustered with a second. On the basis of this finding, we propose that non-transgender women with a genetic link to a transgender woman might be more likely to identify additional transgender women with HIV or those at risk of infection via partner services than cisgender people or MSM who are not genetically linked to transgender women. At present, in LAC and in much of the USA, partner elicitation services are not universally offered. Furthermore, typically less than half of interviews result in the identification of a partner.[19] Molecular epidemiology could be used to prioritise these genetically linked non-transgender women for partner elicitation interviews by public health investigators, with the expectation of identifying more undiagnosed transgender women infected with HIV, transgender women who are uninfected but at high risk of HIV infection, or transgender women with HIV who are not in care. This targeted approach could lead to improved HIV diagnosis, linkage to HIV care, and pre-exposure prophylaxis access.

Clearly, named sexual partners of transgender women should also continue to be considered a high-priority group for HIV research and interventions. In interviews, 20% of the male partners of transgender women have reported being HIV positive.[20,21] Although transgender women think of themselves as a distinct community, their non-transgender partners might not, which makes them more difficult to identify.[20] Molecular epidemiology represents an approach to identify this high-risk population. Importantly, if validated, this type of network-targeted approach would be applicable to any group that clusters assortatively in a molecular transmission network. Nonetheless, we acknowledge that there is a difference between the individuals named during a partner services interview and individuals with a genetic link.[15]

Genetic clustering methods have been rightly criticised for potential bias towards identifying sub-populations with higher sampling rather than higher transmission rates.[22–24] Consistently high frequencies of clustering among MSM[14,25,26] and individuals diagnosed with acute or early-stage infection[23,27] might reflect elevated diagnosis rates rather than exceptional transmission rates; thus, clustering analyses could potentially divert public health focus from where it is most needed.[22] That being said, high clustering has been consistent with shorter transmission intervals in time-resolved analyses,[8] and the algorithm used here has shown ability to detect subpopulations with higher

	aOR (95% CI)	p-value
Diagnosis year	1·03 (1·01–1·04)	<0·00042
Risk		
MSM	1	
Cisgender men	1·74 (1·45–2·07)	<0·0001
MSM-PWID	1·39 (1·03–1·85)	0·0257
Cisgender men and PWID	1·79 (1·07–2·86)	0·0196
Birth country		
USA and US territories	1	
Foreign born	0·81 (0·67–0·97)	0·0250

Only variables with significant association in the multivariate regression model are shown. aOR=adjusted odds ratio. MSM=men who have sex with men. MSM-PWID=men who have sex with men and who inject drugs. PWID=people who inject drugs.

Table 1: **Correlates of clustering with transgender women**

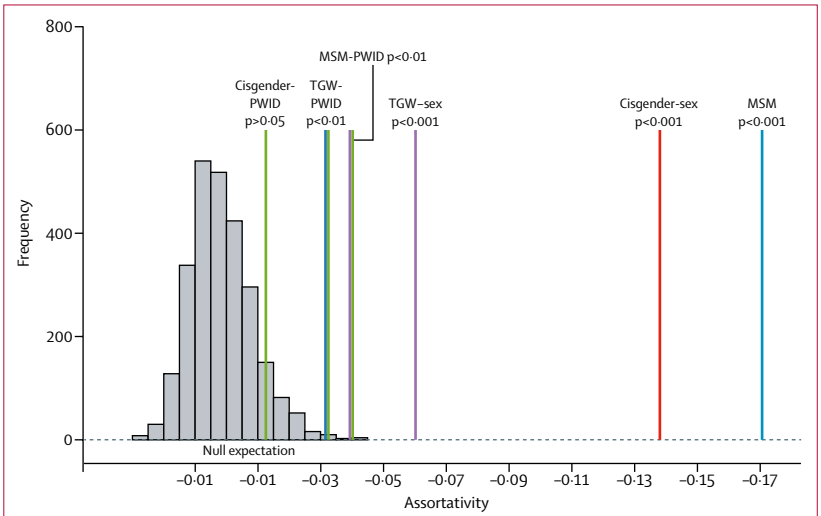

Figure 4: **Assortativity broken down by self-reported risk group**
The null distribution of expected assortativity is shown in grey and the observed assortativity for each risk group is displayed in a different colour. Cisgender-PWID=cisgender people who inject drugs. Cisgender-sex=cisgender people with a sexual risk factor. MSM=men who have sex with men. MSM-PWID=men who have sex with men and inject drugs. TGW-PWID=transgender women who inject drugs. TGW-sex=transgender women with a sexual risk factor.

	F-sex		F-PWID		TGW-PWID		TGW-sex		M-sex		MSM		MSM-PWID		M-PWID	
	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value	OR (95% CI)	p value
TGW-sex	1·09 (0·81–1·62)	0·138	0·73 (0·25–∞)	0·266	6·55 (1·5–∞)	0·017	4·65 (2·08–∞)	<0·0001	1·53 (1·21–2·01)	0·020	0·75 (0·71–0·8)	<0·0001	1·68 (1·11–2·6)	0·009	1·82 (1·05–∞)	0·027
TGW-PWID	0·72 (0·37–4·4)	0·626	0	0·99	11·99 (2·2–∞)	0·049	6·9 (1·97–∞)	0·010	1·38 (0·78–5·2)	0·209	0·69 (0·59–0·85)	0·008	2·43 (1·07–9·6)	0·0187	2·6 (0·48–∞)	0·236

Data are OR (95% CI). Ratios higher than 1 indicate an over-representation of those relationships in the true network compared with random expectation, and ratios of less than 1 indicate an under-representation of those relationships. F-sex=cisgender women at sexual risk. F-PWID=cisgender women who inject drugs. TGW-PWID=transgender women who inject drugs. TGW-sex=transgender women at sexual risk. M-sex=cisgender men at sexual risk. MSM=men who have sex with men. MSM-PWID=men who have sex with men and who inject drugs. M-PWID=cisgender men who inject drugs. OR=odds ratio.

Table 2: **Ratio of the observed proportion of pairwise links compared with the mean of the simulated proportion of pairwise links**

transmission in simulations.[22] Strikingly, in our analysis, the highest clustering was seen among transgender women, a group documented to have low diagnosis rates,[4,6,28] suggesting that in this instance, a genetic clustering approach works well to identify a hidden high-risk population in the absence of increased sampling. However, the proportion of people recently diagnosed with HIV-1 in LAC who have a reported *pol* sequence is only 40–50%, suggesting the potential for sampling bias. Furthermore, transgender women are more likely than other risk groups to engage in care after HIV diagnosis,[29] increasing the likelihood of having an HIV sequence in the surveillance database. To address this potential bias, we used CD4 count at diagnosis as a proxy for time since infection in our multivariate regression. Although a higher CD4 count (suggesting a shorter time between infection and diagnosis than a lower CD4 count) was indeed associated with clustering, the effect was weak, and our main finding was robust including this covariate. Furthermore, all genetic network analyses, such as this one, are limited because they are geographically constricted and affected by sampling, and we cannot account for migration or transmission events occurring outside of LAC. Nonetheless, given that we found that individuals from outside the USA were less likely to cluster than those from within the USA, this migration should not bias our results.

Importantly, a limitation of our clustering analysis is that HIV-TRACE does not infer directionality, and we cannot distinguish between transmitters and recipients in our clusters. However, our inference is not unduly influenced by this limitation, because identifying genetically linked partners is sufficient for deciding whether to prioritise individuals for public health interventions. We find that transgender women are more likely to be involved in HIV transmission events, but we cannot state whether they are more frequently the transmitter or recipient. This finding highlights the importance of allocating public health and other services towards the HIV-infected and at-risk transgender community.

Our finding that transgender women link preferentially to cisgender men with sexual risk (who will be composed mainly of male heterosexuals) is particularly meaningful given that MSM have far higher HIV prevalence than male heterosexuals and are expected to be the source of most infections. This finding is in agreement with interviews of transgender women[4] and their partners.[20] In a study of male partners of transgender women in San Francisco, USA, half the transgender women described themselves as straight, and only 10% identified as gay.[20] Although the genetic transmission network alone does not conclusively reveal source of infection for transgender women (sex with cisgender men with sexual risk, sex with MSM, or shared needles), traits-based phylogenetic analysis on these clusters might further elucidate transmission risk for transgender women. Nonetheless, reliance on self-reporting of transmission risk can be influenced by MSM who do not disclose their risk factors.[11,12] Reliable estimates of transgender women and diagnosis in US populations are unfortunately scarce, but would be helpful for assessing the effect of public health services provided to transgender women and their partners.

In conclusion, we report that transgender women in LAC were more likely to cluster in a molecular transmission network than other risk groups, suggesting high transmission rates, despite low representation in the database. Transgender women were genetically linked to cisgender men with sexual risk more than expected and to MSM less than expected. Transgender women tended to be part of the same clusters, indicating linkage either directly or through shared partners. This assortativity highlights the potential to use molecular epidemiology both to identify transmission clusters that are likely to include undiagnosed or undisclosed HIV-infected transgender women and to improve public health prevention and treatment activities.

Contributors

YWH, ZS, and KP collected and prepared the data. JOW and MR-C conceived of the study and did all analyses. SRM and JOW provided support and advice. YWH reviewed the statistical analyses. MR-C wrote the manuscript and all authors revised the final draft.

Declaration of interests

JOW receives funding from the Centers for Disease Control and Prevention and Gilead Sciences through grants and contracts awarded to the University of California San Diego. We declare no other competing interests.

Acknowledgments

We wish to thank Brooke Sullivan for providing feedback on the manuscript, Ying Ou for his assistance in collecting, processing, and managing HIV-1 sequence data, and Reilly Hostager for help in preparing the tables and figures. This work was supported by a California HIV and AIDS Research Program Innovative, Developmental, Exploratory Award (ID15-SD-052), a National Institutes of Health (NIH)–National Institute of Allergy and Infectious Diseases (NIAID) K01 Career Development Award (K01AI110181), and an NIH-NIAID R01 (AI135992).

References

1 Richardson ET, Collins SE, Kung T, et al. Gender inequality and HIV transmission: a global analysis. *J Int AIDS Soc* 2014; **17:** 19035.

2 County of Los Angeles Department of Public Health. 2015 annual HIV/STD surveillance report. 2018. http://publichealth.lacounty.gov/dhsp/Reports/2015HIV-STDSurveillanceReport.pdf (accessed May 6, 2018).

3 Centers for Disease Control and Prevention. Diagnoses of HIV infection in the United States and dependent areas, 2016. 2017. https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2016-vol-28.pdf (accessed Nov 4, 2017).

4 Herbst JH, Jacobs ED, Finlayson TJ, et al. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. *AIDS Behav* 2008; **12:** 1–17.

5 Centers for Disease Control and Prevention. HIV prevalence, unrecognized infection, and HIV testing among men who have sex with men—five US cities, June 2004–April 2005. *Morb Mortal Wkly Rep* 2005; **54:** 597–601.

6 Schulden JD, Song B, Barros A, et al. Rapid HIV testing in transgender communities by community-based organizations in three cities. *Public Health Rep* 2008; **123** (suppl 3)**:** 101–14.

7 Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 2004; **303:** 327–32.

8 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.

9 Poon AF, Gustafson R, Daly P, et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV* 2016; **3:** e231–38.

10 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.

11 Hue S, Brown AE, Ragonnet-Cronin M, et al. Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS* 2014; **28:** 1967–75.

12 Ragonnet-Cronin M, Hue S, Hodcroft EB, et al. Non-disclosed men who have sex with men in UK HIV transmission networks: phylogenetic analysis of surveillance data. *Lancet HIV* 2018; **5:** e309–16.

13 Kosakovsky Pond SL, Weaver S, Leigh Brown AJ, Wertheim JO. HIV-TRACE (TRAnsmission Cluster Engine): a tool for large scale molecular epidemiology of HIV-1 and other rapidly evolving pathogens. *Mol Biol Evol* 2018; **35:** 1812–19.

14 Oster AM, Wertheim JO, Hernandez AL, Ocfemia MC, Saduvala N, Hall HI. Using molecular HIV surveillance data to understand transmission between subpopulations in the United States. *J Acquir Immune Defic Syndr* 2015; **70:** 444–51.

15 Wertheim JO, Kosakovsky Pond SL, Forgione LA, et al. Social and genetic networks of HIV-1 transmission in New York City. *PLoS Pathog* 2017; **13:** e1006000.

16 Hue 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.

17 Newman ME. Mixing patterns in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2003; **67:** 026126.

18 Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006; 1695.

19 Katz DA, Hogben M, Dooley SW Jr, Golden MR. Increasing public health partner services for human immunodeficiency virus: results of a second national survey. *Sex Transm Dis* 2010; **37:** 469–75.

20 Operario D, Burton J, Underhill K, Sevelius J. Men who have sex with transgender women: challenges to category-based HIV prevention. *AIDS Behav* 2008; **12:** 18–26.

21 Operario D, Nemoto T, Iwamoto M, Moore T. Risk for HIV and unprotected sexual behavior in male primary partners of transgender women. *Arch Sex Behav* 2011; **40:** 1255–61.

22 Poon AF. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. *Virus Evol* 2016; **2:** vew031.

23 Volz EM, Koopman JS, Ward MJ, Brown AL, Frost SD. Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection. *PLoS Comput Biol* 2012; **8:** e1002552.

24 Dearlove BL, Xiang F, Frost SDW. Biased phylodynamic inferences from analysing clusters of viral sequences. *Virus Evol* 2017; **3:** vex020.

25 Ragonnet-Cronin M, Lycett SJ, Hodcroft EB, et al. Transmission of non-B HIV subtypes in the United Kingdom is increasingly driven by large non-heterosexual transmission clusters. *J Infect Dis* 2016; **213:** 1410–18.

26 Kouyos RD, von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J Infect Dis* 2010; **201:** 1488–97.

27 Pao D, Fisher M, Hue S, et al. Transmission of HIV-1 during primary infection: relationship to sexual risk and sexually transmitted infections. *AIDS* 2005; **19:** 85–90.

28 Garofalo R, Kuhns LM, Reisner SL, Mimiaga MJ. Behavioral interventions to prevent HIV transmission and acquisition for transgender women: a critical review. *J Acquir Immune Defic Syndr* 2016; **72** (suppl 3)**:** S220–25.

29 County of Los Angeles Public Health. HIV surveillance reports: 2014 Los Angeles county HIV cascades and PLWH estimate 2016. http://publichealth.lacounty.gov/dhsp/Reports/HIV-STDsurveillanceReport2014.pdf (accessed Nov 5, 2017).