

# Molecular network-based intervention brings us closer to ending the HIV pandemic

Xiaoxu Han<sup>1,2,3</sup>, Bin Zhao<sup>1,2,3</sup>, Minghui An<sup>1,2,3</sup>, Ping Zhong<sup>1,4</sup>, Hong Shang (✉)<sup>1,2,3</sup>

<sup>1</sup>Key Laboratory of AIDS Immunology of National Health Commission, Department of Laboratory Medicine, The First Affiliated Hospital, China Medical University, Shenyang 110001, China; <sup>2</sup>Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, Shenyang 110001, China; <sup>3</sup>Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou 310003, China; <sup>4</sup>Department of AIDS and STD, Shanghai Municipal Center for Disease Control and Prevention, Shanghai 200336, China

© The Author(s) 2020. This article is published with open access at link.springer.com and journal.hep.com.cn

**Abstract** Precise identification of HIV transmission among populations is a key step in public health responses. However, the HIV transmission network is usually difficult to determine. HIV molecular networks can be determined by phylogenetic approach, genetic distance-based approach, and a combination of both approaches. These approaches are increasingly used to identify transmission networks among populations, reconstruct the history of HIV spread, monitor the dynamics of HIV transmission, guide targeted intervention on key subpopulations, and assess the effects of interventions. Simulation and retrospective studies have demonstrated that these molecular network-based interventions are more cost-effective than random or traditional interventions. However, we still need to address several challenges to improve the practice of molecular network-guided targeting interventions to finally end the HIV epidemic. The data remain limited or difficult to obtain, and more automatic real-time tools are required. In addition, molecular and social networks must be combined, and technical parameters and ethnic issues warrant further studies.

**Keywords** human immunodeficiency virus type 1; molecular cluster; transmission cluster; risk network; targeted intervention; evaluation

## Introduction

By the end of 2017, the WHO estimated that 36.9 million people worldwide were living with HIV, 59% of which were receiving antiretroviral treatment (ART). Furthermore, 1.8 million people were newly infected with HIV, and 940 000 died from AIDS-related illnesses in 2017 (<https://www.who.int/hiv/data/en/>). Although ART can effectively control HIV replication, it cannot completely cure HIV infection and lifelong treatment is needed. As neither a feasible cure regimen nor a prophylactic vaccine for AIDS is currently available, the Joint United Nations Program on HIV/AIDS (UNAIDS) released a 90-90-90 target as part of the global push to end the AIDS epidemic by 2030. The following are the targets: 90% of people living with HIV will know their status, 90% of people diagnosed as HIV-infected will receive sustained ART, and 90% of people receiving ART will have controlled viral

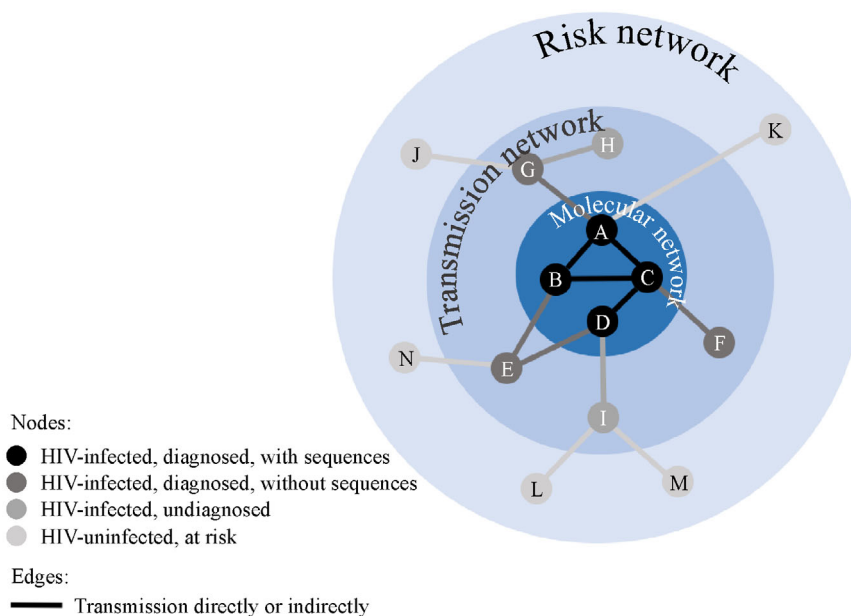
replication [1]. According to the latest data from UNAIDS, a huge gap remained between reality and the target at the end of 2017, as 75% of people living with HIV knew acquired, 79% of diagnosed HIV-infected cases were accessing treatment, and 81% of people under treatment were virally suppressed (<https://www.unaids.org/en/resources/fact-sheet>). Therefore, discovering undiagnosed HIV-infected cases and linking all diagnosed cases to medical care remain as major challenges. Nevertheless, many prevention strategies have been proved to be effective in controlling the spread of HIV among different high-risk populations. These strategies include syringe service programs for injection drug users, condom use for female sex workers and men who have sex with men (MSM), maternal and child block for HIV-infected pregnant women, and pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis for all populations. Unfortunately, implementing these interventions to all high-risk populations is often infeasible in various settings. An increasing number of studies support that molecular network analysis could contribute to disclose among whom and to where the HIV infection is spreading and

to estimate the speed of HIV transmission. This information is usually difficult to obtain via traditional epidemiology surveys because of some biological factors (such as asymptomatic periods of contagiousness), moral framing (such as the stigma that deters people from testing), and epidemiological issues (such as difficulty to track contacts in private settings) [2]. Phylogenetic analysis has long been used to infer potential transmission chain or network among HIV-infected patients [3–5]. In recent years, a new simplified genetic distance (GD)-based approach has been developed to infer potential transmission network in real time for proper interventions [6] and is increasingly used in large sequence datasets in both USA [7,8] and Europe [9]. In 2018, the US Centers for Disease Control and Prevention (CDC) affirmed the significance of molecular network analyses [10]. When combined with epidemiologic investigations and public health action plans, a molecular network-based strategy can identify more undiagnosed infections and more HIV-negative network members at high risk of infection and allow targeted prevention efforts. Accordingly, the USA and China have published guidelines on detecting and responding to HIV molecular network [11] and implemented this molecular network-based intervention strategy as key tools to HIV/AIDS control. The US Department of Health and Human Services announced a project titled “Ending the HIV Epidemic: A Plan for America” in 2019, which aimed to reduce new HIV infections in the USA by 75% in 5 years and by 90% by 2030; in this plan, rapid detection and response to expanding HIV clusters and further reduction of new transmissions were proposed as one of the four

pillars of a strategic initiative [12]. In the present review, we introduce the relationships among molecular, transmission, and risk networks, as well as the principal methods of molecular network construction, by focusing on the recent progress in HIV molecular network application and the major challenges to improve the molecular network surveillance and responses in different HIV epidemics.

## HIV molecular cluster infers transmission cluster and risk networks

HIV is characterized by high genetic variability. Individuals carrying genetically similar viral strains appear to be closely related by transmission either directly or indirectly [11]. An HIV molecular cluster is a group of HIV-infected individuals having genetically similar HIV strains, in which nodes represent an HIV-infected individual or a fragment of the HIV sequence, and edges represent the potential transmission link between cases [11]. By comparison, an HIV transmission cluster is a group of HIV-infected individuals having a direct or indirect epidemiological connection, which includes both HIV-infected patients in a molecular network and diagnosed or undiagnosed HIV-infected patients who do not appear in the molecular cluster because of unavailable sequences [11]. A risk network includes both HIV-infected cases in the transmission cluster and individuals who have not been infected with HIV but have come in contact with infected cases in the transmission clusters (Fig. 1). The identification of cases in HIV molecular clusters helps to elucidate



**Fig. 1** Molecular network and underlying transmission and risk networks.

the rapidly growing HIV transmission clusters so that more concerns and priorities could be given to the transmission clusters and risk networks that can be intervened preferentially. Through investigating the cases in the prioritized molecular clusters, we can obtain more clinical and social behavioral data and identify factors associated with HIV transmission and key characteristics of the underlying risk network, including high-risk ongoing transmission, poor outcomes, particularly vulnerable or underserved population, transmission of drug resistance, injection drug use, sexually transmitted diseases, and hepatitis coinfection [11]. Feature-specific interventions can be provided to the most associated risk networks.

Several key points deserve consideration with regard to the interpretation and application of molecular networks. First, depending on the sampling depth, a molecular network represents only a subset of what is likely a larger underlying transmission network. Second, if two individuals have highly similar HIV strains, they could be either directly or indirectly linked through transmission. Third, the link (edge) between two cases in a molecular network does not suggest the direction of infection (who transmitted HIV to whom). Forth, rich data sources of viral sequences are fundamental for molecular network analysis to obtain an overall view of HIV transmission with public health significance. Finally, although drug resistance testing is recommended for all persons with diagnosed HIV infection, actually not all persons could receive a drug resistance test, especially in resource-constrained settings.

## Methods of molecular network construction

The *pol* gene of the HIV-1 genome, the target of HIV drug-resistance testing, is most commonly used in molecular cluster construction. All patients are suggested to undergo HIV drug-resistance test before ART in developed countries. When patients experience virological failure during ART treatment in developing countries, huge volumes of relevant data can be obtained without extra expenses. Nevertheless, the *pol* gene is considered as less informative and has a relatively low substitution rate in the HIV genome [13]. Whole genome sequence or the *env* gene sequence of HIV-1 is thought to better reflect the real transmission relationship [5,14]. In an HIV-1 transmission chain consisting of nine patients, the evolutionary history inferred by phylogenetic tree with the *pol* gene sequences was not fully compatible with the known transmission history, and the multidrug-resistant viruses were incorrectly clustered; by contrast, the *env* phylogenetic tree was fully compatible with the known transmission history [15]. However, the use of whole genome sequence or the *env* gene sequence is not applicable to public health practice because of the strict technical requirements, high costs, and high length polymorphisms.

No standard approach is presently available to define molecular networks. Two general categories of approaches have been commonly used independently or combined to identify HIV molecular clusters. The first is phylogeny-based approach, in which sequences sharing a common ancestor are defined as a cluster. Several phylogenetic methods can be used to infer a phylogenetic tree, such as a neighbor-joining (NJ) tree, maximum likelihood (ML) tree, or maximum clade credibility (MCC) tree, supported by the bootstrap value, likelihood-ratio test, zero-branch length test, or posterior probability. The NJ tree is based on a distance model and can be constructed faster than the ML and MCC trees. Therefore, it has been commonly used to construct phylogenetic trees in earlier studies. The ML and MCC trees both use site substitution models to evaluate the relative likelihood of different phylogenetic topologies, which cause high computational burden. The MCC tree could also allow for the molecular clock type and demographic model to estimate the time to most recent common ancestor (tMRCA), the evolutionary rate, and past effective population size (the number of individuals in a population who contribute offspring to the next generation) through time, which can reflect the growing or declining demographic history of the viral epidemic [13,15,16]. The nucleotide substitute, molecular clock, and population dynamic models should be tested to determine which would best fit to the target sequence dataset before reconstructing evolutionary history [17,18]. Software packages (BEAST 1 and BEAST 2) are widely used for phylodynamic and phylogeographic inference [19,20]. Several recent studies also used viral sequences with spatiotemporal characteristics to infer the origin and spread of transmission cluster or network through phylodynamic and phylogeographic approaches [21,22]. The basic idea of molecular network is to classify viral sequences according to genetic similarities. However, with the phylogeny-based approach, a highly divergent descendant sequence cannot be excluded from the others with a common ancestor [23], which might imply that sample collection long after transmission does not infer a recent active transmission network.

The other approach is GD-based cluster definitions. Pairwise GD is usually calculated using the TN 93 substitution model. Individuals with a pairwise distance below the predefined GD threshold are assigned to the same clusters [24]. Various GD thresholds are recommended on the basis of the goal of the analysis. A genetic threshold of 0.5%, with approximately five different nucleotides for sequences 1000 nucleotides long, is suggested to identify cases related to recent and rapid spread. This threshold corresponds to approximately 2–3 years of independent viral evolution [11]. If the goal is to identify all possible cases potentially related to a given case, a larger GD threshold of 1.5% corresponding to a maximum of 7–8 years of viral evolution separating strains

**Table 1** Tools for GD-based and phylogeny-based approaches

Method	Model/techniques	Cutoff values	Objects	Source/method	References
<b>GD-based method</b>					
Threshold bootstrap clustering (TBC)	Resampling techniques and models of sequence evolution, such as the LogDet distance	Need to be set for research purposes	HIV group M ( $n = 38$ ) /HIV- <i>pol</i> gene (subtype B, $n = 356$ ) /HCV	Source code attached to study	[97]
Gap Procedure	Adjusted versions of the K80 distance	No user-specific threshold values	HIV- <i>pol</i> gene (subtype B, $n = 1571$ )	GapProcedure package ( <a href="https://github.com/vrbiki/GapProcedure">https://github.com/vrbiki/GapProcedure</a> )	[98]
TRAnsmiSSion Cluster Engine (HIV-TRACE)	TN <sup>a</sup> 93 distances	$\leq 1.5\%$ genetic distance for subtype B	HIV- <i>pol</i> gene (subtype B, $n = 605$ )	<a href="http://www.hivtrace.org">www.hivtrace.org</a> / <a href="https://github.com/veg/hivtrace">www.github.com/veg/hivtrace</a>	[8,99]
<b>Phylogeny-based method</b>					
Bootstrap subtrees	NJ <sup>b</sup> tree with the bootstrap value	Bootstrap ( $>85\%$ )	HIV- <i>pol</i> gene (subtype B, $n = 191$ )	PHYMLIP package	[15]
Bootstrap and branch-lengths	NJ <sup>b</sup> tree constructed from HKY <sup>c</sup> 85 distances	Bootstrap $>99\%$ , mean branch length $<1.5\%$	HIV- <i>pol</i> gene (subtype B, $n = 193$ )	Paup software	[100]
PhyloPart	ML <sup>d</sup> tree was constructed by FastTree with support values based on the SH <sup>e</sup> test; the median patristic distances within a given subtree was computed	SH tested support value ( $>90\%$ ); genetic distances (4%–8% nucleotide substitutions per site)	HIV- <i>pol</i> gene (subtype B, $n = 11\ 541$ )	Attached to the study	[28]
Cluster Picker	ML <sup>d</sup> tree was constructed by FastTree2 with support values based on the SH <sup>e</sup> test; the maximum pairwise distance within the subtree was computed	SH <sup>e</sup> tested support value (70%–99%); genetic distances (1.5%–4.5% nucleotide substitutions per site)	HIV- <i>pol</i> gene (subtype B, $n = 1381$ ) /HCV/Flu sequences	<a href="http://hiv.bio.ed.ac.uk/software.html">http://hiv.bio.ed.ac.uk/software.html</a>	[29]

<sup>a</sup>TN, Tamura–Nei; <sup>b</sup>NJ, neighbor-joining; <sup>c</sup>HKY, Hasegawa–Kishino–Yano; <sup>d</sup>ML, maximum likelihood; <sup>e</sup>SH, Shimodaira–Hasegawa test.

is suggested by the US CDC guidelines [11]. The HIV-TRACE is a distance-based visual software used for molecular network construction (<http://hivtrace.datamonkey.org/hivtrace>) and has been applied in the US and several Asian countries [25,26].

The GD-based and phylogeny-based methods are neither good nor bad. However, choosing the appropriate method on the basis of sequence characteristics and research objectives is necessary. The HIV-TRACE software tends to detect larger and fewer clusters than the Cluster Picker, which detects more clusters that contain only two sequences [27]. When the goal is to detect larger networks in a deep sampling area, the HIV-TRACE may perform more favorably and is expected to identify more transmission chains [27]. The GD-based approach can be used for more rapid computation but cannot distinguish different evolutionary rates and may underestimate the divergence time of the virus. Moreover, GD is closely associated with potential evolutionary distance, which makes the approach popular for the reconstruction of molecular networks in real time and monitoring of dynamic trends [8]. As for the phylogeny approach for molecular network construction, obtaining a high node

support value and steady topologies is difficult when processing a large number of sequences. Specific software, such as the PhyloPart [28] and Cluster Picker [29], can combine phylogenetic tree bootstraps and GDs to identify transmission clusters. Whatever inherent bias the genetic clustering method may have, the rapid succession of newly infected individuals in a predefined cluster indicates a local outbreak of HIV infection [30]. The tools commonly used for both GD-based and phylogeny-based methods are listed in Table 1.

## Molecular networks reconstruct the history of HIV spread

Phylogenetic analysis has long been used to identify HIV linkage and infer putative network among populations. The known HIV-1 transmission history could be accurately reconstructed through phylogenetic tree analysis [4,31]. Therefore, molecular investigation of HIV-1 transmission is widely used to infer HIV transmission among populations [5,32], such as in a Dutch criminal case [3] and presumed transmission pairs in a heterosexual cohort of

discordant couples in Zambia [5]. Phylogenetic analysis has also been used to infer HIV transmissions in countries with large datasets [33–36]. With the combination of demographic, sociological, and epidemiological information, phylogenetic analysis can also help to characterize the source population of HIV infection. Paraskevis *et al.* analyzed nearly 9000 HIV-1 sequences collected from one Canadian and nine European HIV cohorts and found that the sub-type B viruses that spread within MSM networks appeared to be the major driving force responsible for the HIV epidemic dispersal [37]. A large-scale study in KwaZulu-Natal, South Africa combined molecular network and demographic information to identify the key mode of sexual networks driving local HIV transmission. In that study, older men were found to acquire HIV from women of similar age and transmit HIV to younger women [38]. Faria *et al.* used the HIV-1 sequence data from Central Africa and reconstructed the early stage of HIV-1 transmission history; they emphasized that both social changes and transport networks played important roles in the viral establishment in human populations [39]. The geospatial viral migration patterns and temporal dynamics of HIV-1 transmission can be further reconstructed when molecular network analysis is combined with both geographical and temporal information [35,40,41]. A study on the HIV-1 epidemic in the Nordic countries found both different HIV-1 transmission patterns between countries and linkages in a large geographical region; Denmark and Sweden showed the strongest geographical link, and Denmark had a great part of heterosexual domestic spread of HIV-1 subtype B [32]. A phylogeographic study in Uganda detected viral migration from the general population to fishing communities, suggesting that these communities were a reservoir for, and not the source of, viral strains from the general population [35]. A recent study from Europe applied phylogeographic analyses and successfully identified the HIV transmission hotspots in Cologne-Bonn (Germany). The authors found that clustered individuals tended to live closer to one another compared with other individuals without any linkage [42]. A study from Europe used Bayesian coalescent-based methods to analyze the HIV sequences of primary HIV-1-infected individuals from the ANRS PRIMO C06 cohort for over 15 years and determined that Paris was the spread center of both subtype B and CRF02\_AG epidemics [43].

In 2014, Wertheim *et al.* developed a computationally efficient GD-based approach [44], which increased the speed of analysis to a level high enough for both large-scale sequence data and dynamic monitoring of transmission clusters in a near real-time manner. They used more than 80 000 published sequences from 141 countries and regions worldwide to construct molecular networks [44] and revealed a contemporary picture of HIV-1 transmission within and between countries, including well-characterized transmission clusters, unrecognized transmission clusters

across international borders, and other previously undescribed transmission clusters [45–47]. In subsequent studies, this approach was used to analyze cross-border transmission. Mehta *et al.* constructed HIV transmission networks in the San Diego–Tijuana border region and found five clusters consisting of individuals residing on both sides of the border [48]. Using the GD method, the US CDC studied 40 950 HIV-*pol* sequences, along with demographic and epidemiological data, and found that heterosexual women were more likely linked to MSM than to heterosexual men, especially older and African–American MSM. This study underlined the key role of MSM in the HIV epidemic in the USA [49]. Another recent study from the Los Angeles County Department of Public Health used a combination of molecular network and epidemiological data and found that transgender women with sexual risk factor tended to be clustered and linked both to other transgender women and cisgender men but were less likely linked to MSM [50].

Compared with the above European and American countries, some countries like China have complex HIV subtypes and multiple distinct viral lineages. Numerous phylogenetic studies have revealed the origins and routes of transmission of major HIV subtypes prevalent in China [51–58]. In recent years, some Chinese scholars have also used genetic surveillance or drug-resistance testing data to construct national or regional molecular networks for major HIV subtypes in China with both the above two approaches; these studies were the first attempts and explorations of molecular networks and transmission networks [59–63]. The 4th National HIV Molecular Epidemiological Survey revealed a full picture of the main epidemic clusters among different high-risk populations. Multiple clusters were identified from the MSM population, including two CRF01\_AE clusters, one CRF07\_BC cluster, and a small CRF55\_01B cluster. A greater number of MSM were observed within clusters and linked with other high-risk populations [64]. Several other large-scale studies have also demonstrated multiple epidemic clusters responsible for 85% of the total CRF01\_AE infections in China [65], among which two large clusters were found with high prevalence among MSM [57,58]. These findings all suggest that MSM are at higher risk than others in the population and highlight the importance of MSM-focused interventions for the control of HIV in China.

## Molecular networks monitor the dynamics of local HIV transmission

Molecular network analysis can also be used to monitor epidemic trends among populations. The determination of molecular cluster growth provides a quantitative evaluation of relevant transmission clusters and risk networks that

enable targeted interventions or effect evaluation. Dennis *et al.* revealed the transmission cluster expansion in North Carolina using densely sampled data and characterized active transmission clusters with phylodynamic analysis. They used effective reproductive number ( $Re$ ) to monitor active clusters, which demonstrated the propensity of steady onward propagations [21]. Another phylodynamic analysis study in Serbia revealed that the  $Re$  remained over one during the complete period of the investigation; an MSM transmission group of subtype B was determined to have a recent tMRCA and steep growth curve until 2030, whereas heterosexuals with both subtypes B and C displayed minor growth and stagnation [34]. Valverde *et al.*'s study, which focused on HIV transmission dynamics among US immigrants who were disproportionately affected by HIV, revealed that the majority of new HIV infections appeared subsequent to the immigration of related individuals to the USA. Consequently, the transmission network information regarding HIV acquisition and transmission routes among these individuals are required to improve HIV prevention among such populations [66]. In a phylodynamic study in South Africa using a large sequence dataset, the date of origin of 18 clades fell between 1979 and 1992 and a strong growth was found to have occurred in the 1990s. A decreasing growth rate in four of the clades was detected since the advent of interventions but not in the other 14 clades [22]. Effective intervention depends on the timely monitoring of the dynamic of HIV epidemic. Therefore, real-time monitoring of the growth of molecular clusters is important. Two studies demonstrated that molecular network analysis is sensitive enough to record the process of an outbreak and control of HIV infection. One study on injection drug users in rural Indiana identified a molecular cluster supported by both epidemiological and viral genetic similarity, which first arose in 2011 then had an outbreak in mid-2014, and subsequently waned after the declaration of a public health emergency and intervention [67]. Another study in British Columbia, Canada conducted in 2016 implemented an automated near real-time monitoring system, which could generate monthly reports to public health officers, and showed the growth of HIV molecular clusters. This system demonstrated the ability to monitor the outbreak of HIV drug-resistance clusters and assess the enhancement of public health action, resulting in a remarkable decrease in the transmission of those clusters in the affected sub-population [68].

## Molecular networks guide targeted intervention

HIV transmission networks elucidate the spread of HIV among population and offer the opportunity of intervention, which are traditionally identified through HIV

surveillance, partner services, and contact investigations. Molecular network analysis is complementary to the existing partnerships that underlies the social or HIV transmission networks and contributes in partner notification promotion [6], thereby bridging the previously unrecognized partner notification network component. This analysis provides more reliable evidence than partner naming for the identification of potential transmission links [69–72]. In 2009, Smith *et al.* introduced a scientific model designed for studying the molecular surveillance of HIV transmission using public health information [6]. They found that molecular epidemiology coupled with partner contact tracing can be used to identify individuals within a population that belong to highly related HIV transmission groups. These methods could be used to implement selectively targeted preventive interventions. Some parameters have been used to quantify the risk of individuals among molecular networks and to guide targeting intervention. One indicator is called link or degree. The more links the individuals have in the network, the more potential transmission partners they could have and the higher possible communication risk will be. Oster *et al.* used the link as an indicator in a study on the groups with high risks and among groups with different racial/ethnic statuses in US National HIV Surveillance System. The HIV-infected heterosexual women were predominantly linked to MSM. The interventions that were able to reduce HIV transmissions among individuals in the MSM population showed great possibilities to reduce HIV acquisition, as well as among other groups with high risks [49]. Leigh Brown *et al.* used degrees to categorize HIV infection in a phylodynamic analysis on MSM in the United Kingdom (UK) and showed the preferential association of UK MSM and called for intervention targeting high-degree individuals [73]. A San Diego research group developed a parameter called transmission network score (TNS) to estimate the HIV transmission risk from a newly diagnosed individual to his partner. In this retrospective simulation analysis, they found that compared with the clustered individuals from a randomly selected subset, the ART targeting individuals with the highest TNS showed a substantially reduced HIV transmission level within the network [8]. The effect of the TNS-based targeted ART was further supported by another study on Chinese MSM. High levels of TNS-based ART were simulated and compared with CD4 T cell counts, a viral load-based strategy, and a treat-all strategy in a primary HIV-infected MSM cohort in Beijing. The results showed that the prevention efficiency of high TNS-based ART was between 30% and 42%, which was considerably higher than the other three strategies evaluated. This study implied that TNS-based strategies may be an efficient way to provide preventive interventions [26]. In addition to targeted interventions based on individual HIV transmission risk, the transmission rate has

been used to identify recent and rapidly growing clusters and prioritize public health responses. A study from the US CDC identified 60 clusters from 27 surveillance jurisdictions in the National HIV Surveillance System from 2013 to 2017; the transmission rate of 11 clusters was eleven times higher than that of the national estimates of 4/100 person-years and should be prioritized for public health intervention [74]. Except for the parameters of molecular cluster, high viral load might also act as an indicator of targeted interventions. According to the latest research on the United States National HIV Surveillance System, the frequently transmitted strains were found in large molecular clusters and had significantly higher viral loads and increased network connectivity, and that these clusters should be afforded the highest priority for public health interventions to interrupt the transmission [75].

Investigations of molecular clusters might also help to reveal diagnosed or undiagnosed HIV-infected cases without links to medical care and uninfected cases at a very high risk of infection. An excellent example of molecular networks guiding targeted intervention in real-world settings was implemented in Canada. In this study, an automated phylogenetic system detected a recent HIV outbreak and supported an enhanced public health follow-up to ensure linkage to care and treatment initiation in the affected subpopulation; a reduction in the onward transmission of drug resistance was observed during the follow-up [68]. One retrospective study conducted in San Antonio, Texas from 2013 to 2015 identified a cluster of 27 individuals, which expanded rapidly on subsequent monitoring. Further investigation of the partner services and interview records identified 87 HIV-infected persons who were sex partners, needle-sharing partners, or social network contacts of confirmed cases; therefore, these 87 individuals were highly likely to belong to the same transmission cluster as the aforementioned 27 individuals. However, these individuals failed to receive appropriate medical care and their HIV sequencing data were not available for molecular network analysis; as a result, they remained at a high HIV transmission risk [10]. Another example is an HIV outbreak among injection drug users in Indiana. Through molecular networks and high-risk sex, needle sharing, and both sex and needle sharing contact self-reports, more than 200 HIV-negative individuals were identified with close social links to clusters of HIV-infected members, among which the outbreak occurred [67,76]. This study demonstrated that investigating actively growing molecular clusters provides opportunities for prioritizing persons associated with these clusters for linkage to care and PrEP referral [77]. A study from the UK also demonstrated that network-based approaches can guide targeted prevention efforts for individuals who were currently HIV-negative but with very high infection risks at a cost-effective manner. Simulated interventions indicated that focusing PrEP on young MSM can prevent four

times more infections over 5 years than random allocation [9]. Although PrEP is widely accepted in developed countries, its use in China is controversial. The estimated 1.2 million MSMs in China with the highest HIV incidence [78] are the potential targeted population for PrEP. Molecular network analysis helps to provide PrEP to individuals in priority network and might substantially improve the effect of PrEP in this situation.

## Molecular networks evaluate the effectiveness of interventions

Molecular networks can also be used to evaluate the effects of intervention. Several methods have been developed to evaluate whether the intervention strategy under investigation can interrupt transmission at the population level. A recent retrospective study on New York HIV transmission network demonstrated that previous growth dynamics of clusters can predict the future growth of the clusters. Therefore, the prioritization schemes at the cluster-level, with a consideration of the relationship between the previous cluster growth and size, may contribute to improve the final outcomes in public health [79]. A study from San Diego evaluated the effects of HIV control in a nucleic acid testing-based early test program using molecular cluster monitoring. The authors found that with the early test program, about 100 less HIV infections occurred compared with the number expected in the central region of San Diego in 2012. Genetic analysis also suggested that the HIV transmission chains are more likely to end in areas with marketed early testing [80]. The reproduction number ( $R$ ), a parameter reflecting how efficiently the infectious agents are transmitted, is usually used for modeling infection dynamics.  $R > 1$  represents the infectious agents can continue to spread. Two main estimators are used: the basic reproductive number ( $R_0$ ) and the effective reproductive number ( $R_e$ ).  $R_0$  and  $R_e$  are the average numbers of secondary infections caused by a typical infected individual in an entirely susceptible population and in only a part of the population that is susceptible, respectively [81]. For low prevalent epidemic diseases like HIV,  $R_e$  is equal to  $R_0$  [44]. In 2012, a group from Switzerland developed a new Bayesian phylogenetic method based on a birth–death model to estimate the  $R_0$  directly by using the viral sequence data, in which the transmission and death rates were estimated independently to substantially improve the accuracy compared with other coalescent estimates [82]. In 2017, the same group estimated the  $R_0$  of HIV epidemics among a heterosexual population in Switzerland using the population-based phylogenetic cluster analysis and found that the  $R_0$  of the population was far below the epidemic threshold [83]. This method might be able to assess the effects of currently implemented preventive measures. Another recent study

incorporated phylodynamics into a molecular cluster analysis to monitor the cluster dynamics with densely sampled sequences from North Carolina. The estimated  $Re$  of the active clusters was remarkably higher than that of the historical clusters. Determining actively growing clusters is crucial to optimize the approaches for public health responses, and an effective intervention would be expected to reduce the  $Re$  [21,84].

## Challenges and recommendations

An increasing number of studies support the essential role of molecular network analysis in HIV epidemic monitoring and targeted intervention. Molecular network-based intervention strategy has been implemented as a key tool for HIV/AIDS control in many countries, including China. However, we still need to overcome several challenges to improve the practice of molecular network-guided targeted interventions and to finally end the HIV epidemic. First, the sequence data for molecular network construction are still limited. In-depth sampling is important for more accurate characterization of HIV transmission networks [85,86]. The latest US CDC guidelines for detecting and responding to HIV transmission clusters recommended the collection of sequences from more than 60% of individuals diagnosed with HIV infection among the target population [11]. However, in most resource-limited countries, drug-resistance testing immediately after diagnosis or even before ART is still unavailable. More financial support would help to increase regular drug-resistance testing, either through government support or broadening medical insurance coverage. In other instances, sequences of drug-resistance testing and surveillance data from hospitals cannot be transmitted efficiently to public health departments. The establishment of a procedure of the *pol* gene sequence collection and analysis is pressing need to ensure that the *pol* gene sequence data are transmitted to the public health department for molecular network monitoring. Second, real-time molecular network monitoring is required to reveal the rapidly growing transmission clusters and guide timely investigation and response. This real-time monitoring requires a platform that provides a setting on which the following objectives can be achieved automatically and efficiently: collection of the *pol* sequences generated from decentralized hospitals and institutes, identification of molecular clusters, evaluation of time-space dynamics of clusters, and prioritization of clusters for investigation. Although some GD-based or phylogenetic-based software packages, such as HIV-TRACE [87] and Cluster Picker [29], are publicly available, the construction of molecular networks remains a highly technical and labor-intensive task. More automatic data transfer and analysis tools must be developed in the future. Third, molecular cluster inference must be

combined with cluster surveillance, investigation, and intervention to complement and support one another. Therefore, staff members concerned with data analysis, surveillance, and intervention, together with HIV care providers and community-based organizations, must collaborate with one another to provide timely information on transmission networks during investigations, which can help them to focus on prevention efforts [88,89]. The real-time molecular network study in Canada provides an excellent example of close cooperation between the CDC and hospitals: the study identified the clinical, demographic, epidemiological, and HIV sequence data from all HIV infections and efficiently integrated those data into a cached database that was automatically processed to support public health responses [68]. Supplementary national or regional guidelines on molecular network monitoring and response are required to meet the specific condition. Fourth, various parameters are the key determinants of the sensitivity and specificity of molecular cluster inference. However, the studies on HIV evolution support the key parameter selection and molecular network-guided applications, which were mainly focused on subtype B HIV-1 [71,90]. Subtype B is responsible for only 12.1% of HIV infections globally [91], whereas multiple non-B HIV-1 strains are being transmitted at various rates in different countries worldwide, including China [92]. Nearly all molecular network studies on non-B HIV have followed the parameters of subtype B [26,93–95]. However, whether the parameters for subtype B are also appropriate for non-B viral strains and more complicated epidemic conditions have not been fully explored. Evolutionary studies on non-B HIV must be strengthened to provide a stronger theoretical basis for non-B HIV molecular network studies. Finally, ethical, legal, and social issues remain for HIV molecular network analyses aimed at targeted intervention, as reviewed by Mehta *et al.* recently [96]. The legal provisions and public opinion on disclosure vary greatly from country to country. Further studies are warranted to develop privacy protection data-sharing techniques and strengthen the understanding between the researcher and public health officials.

In more than 30 years of practice in HIV/AIDS prevention and control, scholars from China have also gained rich experience in HIV molecular epidemiology research, epidemiological surveys, and social network investigations. Various research teams in China are presently engaged in the study of HIV-1 molecular network-based targeted interventions aligned with the HIV epidemics in China. These studies are expected to provide new strategies to deal with HIV transmission among high-risk populations in China.

## Acknowledgements

This work was supported in part by the Mega-Projects of the



National Science Research for the 13th Five-Year Plan (No. 2017ZX10201101), Innovation Team Development Program of the Ministry of Education (No. IRT\_16R70), the National Natural Science Foundation of China (No. 81871637), and Central Public-interest Scientific Institution Basal Research Fund (No. 2018PT-31042).

## Compliance with ethics guidelines

Xiaoxu Han, Bin Zhao, Minghui An, Ping Zhong, and Hong Shang declare no conflicts of interest. This manuscript is a review article and does not entail a research protocol requiring approval by the relevant institutional review board or ethics committee.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>.

## References

- McMahon JH, Medland N. 90-90-90: how do we get there? *Lancet HIV* 2014; 1(1): e10–e11
- Kazanjan P. UNAIDS 90-90-90 campaign to end the AIDS epidemic in historic perspective. *Milbank Q* 2017; 95(2): 408–439
- Machuca R, Jørgensen LB, Theilade P, Nielsen C. Molecular investigation of transmission of human immunodeficiency virus type 1 in a criminal case. *Clin Diagn Lab Immunol* 2001; 8(5): 884–890
- 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(20): 10864–10869
- Trask SA, Derdeyn CA, Fideli U, Chen Y, Meleth S, Kasolo F, Musonda R, Hunter E, Gao F, Allen S, Hahn BH. Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. *J Virol* 2002; 76(1): 397–405
- Smith DM, May SJ, Tweeten S, Drumright L, Pacold ME, Kosakovsky Pond SL, Pesano RL, Lie YS, Richman DD, Frost SD, Woelk CH, Little SJ. A public health model for the molecular surveillance of HIV transmission in San Diego, California. *AIDS* 2009; 23(2): 225–232
- Aldous JL, Pond SK, Poon A, Jain S, Qin H, Kahn JS, Kitahata M, Rodriguez B, Dennis AM, Boswell SL, Haubrich R, Smith DM. Characterizing HIV transmission networks across the United States. *Clin Infect Dis* 2012; 55(8): 1135–1143
- Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, May S, Smith DM. Using HIV networks to inform real time prevention interventions. *PLoS One* 2014; 9(6): e98443
- Volz EM, Le Vu S, Ratmann O, Tostevin A, Dunn D, Orkin C, O'Shea S, Delpech V, Brown A, Gill N, Fraser C; UK HIV Drug Resistance Database. Molecular epidemiology of HIV-1 subtype B reveals heterogeneous transmission risk: implications for intervention and control. *J Infect Dis* 2018; 217(10): 1522–1529
- Oster AM, France AM, Mermin J. Molecular epidemiology and the transformation of HIV prevention. *JAMA* 2018; 319(16): 1657–1658
- National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention; Division of HIV/AIDS. Prevention Detecting and Responding to HIV Transmission Clusters: A Guide for Health Department. 2018. <https://www.cdc.gov/hiv/pdf/funding/announcements/ps18-1802/CDC-HIV-PS18-1802-AttachmentE-Detecting-Investigating-and-Responding-to-HIV-Transmission-Clusters.pdf>
- Fauci AS, Redfield RR, Sigounas G, Weahkee MD, Giroir BP. Ending the HIV epidemic: a plan for the United States. *JAMA* 2019; 321(9): 844–845
- 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(5): 719–728
- Yebra G, Hodcroft EB, Ragonnet-Cronin ML, Pillay D, Brown AJ; PANGEA\_HIV Consortium; ICONIC Project. Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. *Sci Rep* 2016; 6(1): 39489
- Yerly S, Vora S, Rizzardi P, Chave JP, Vernazza PL, Flepp M, Telenti A, Battegay M, Veuthey AL, Bru JP, Rickenbach M, Hirschel B, Perrin L; Swiss HIV Cohort Study. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. *AIDS* 2001; 15(17): 2287–2292
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987; 4(4): 406–425
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 2012; 29(9): 2157–2167
- Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Mol Biol Evol* 2013; 30(2): 239–243
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Poppinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2019; 15(4): e1006650
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 2018; 4(1): vey016

21. Dennis AM, Hué S, Billock R, Levintow S, Sebastian J, Miller WC, Eron JJ. Human immunodeficiency virus type 1 phylogenomics to detect and characterize active transmission clusters in North Carolina. *J Infect Dis* 2019 Apr 27. [Epub ahead of print] doi: 10.1093/infdis/jiz176
22. Wilkinson E, Junqueira DM, Lessells R, Engelbrecht S, van Zyl G, de Oliveira T, Salemi M. The effect of interventions on the transmission and spread of HIV in South Africa: a phylodynamic analysis. *Sci Rep* 2019; 9(1): 2640
23. Poon AF. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. *Virus Evol* 2016; 2(2): vew031
24. Aldous JL, Pond SK, Poon A, Jain S, Qin H, Kahn JS, Kitahata M, Rodriguez B, Dennis AM, Boswell SL, Haubrich R, Smith DM. Characterizing HIV transmission networks across the United States. *Clin Infect Dis* 2012; 55(8): 1135–1143
25. Chin BS, Chaillon A, Mehta SR, Wertheim JO, Kim G, Shin HS, Smith DM. Molecular epidemiology identifies HIV transmission networks associated with younger age and heterosexual exposure among Korean individuals. *J Med Virol* 2016; 88(10): 1832–1835
26. Wang X, Wu Y, Mao L, Xia W, Zhang W, Dai L, Mehta SR, Wertheim JO, Dong X, Zhang T, Wu H, Smith DM. Targeting HIV prevention based on molecular epidemiology among deeply sampled subnetworks of men who have sex with men. *Clin Infect Dis* 2015; 61(9): 1462–1468
27. Rose R, Lamers SL, Dollar JJ, Grabowski MK, Hodcroft EB, Ragonnet-Cronin M, Wertheim JO, Redd AD, German D, Laeyendecker O. Identifying transmission clusters with cluster picker and HIV-TRACE. *AIDS Res Hum Retroviruses* 2017; 33(3): 211–218
28. Prosperi MC, Ciccozzi M, Fanti I, Saladini F, Pecorari M, Borghi V, Di Giambenedetto S, Bruzzone B, Capetti A, Vivarelli A, Rusconi S, Re MC, Gismondo MR, Sighinolfi L, Gray RR, Salemi M, Zazzi M, De Luca A; ARCA collaborative group. A novel methodology for large-scale phylogeny partition. *Nat Commun* 2011; 2(1): 321
29. Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJ, Lycett S; UK HIV Drug Resistance Database. Automated analysis of phylogenetic clusters. *BMC Bioinformatics* 2013; 14(1): 317
30. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS* 2017; 31(9): 1211–1222
31. Paraskevis D, Magiorkinis E, Magiorkinis G, Kiosses VG, Lemey P, Vandamme AM, Rambaut A, Hatzakis A. Phylogenetic reconstruction of a known HIV-1 CRF04\_cpx transmission network using maximum likelihood and Bayesian methods. *J Mol Evol* 2004; 59(5): 709–717
32. Esbjörnsson J, Mild M, Audelin A, Fonager J, Skar H, Bruun Jørgensen L, Liitsola K, Björkman P, Bratt G, Gisslén M, Sönnnerborg A, Nielsen C; SPREAD/ESAR Programme, Medstrand P, Albert J. HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. *Virus Evol* 2016; 2(1): vew010
33. Bruhn CA, Audelin AM, Helleberg M, Bjorn-Mortensen K, Obel N, Gerstoft J, Nielsen C, Melbye M, Medstrand P, Gilbert MT, Esbjörnsson J. The origin and emergence of an HIV-1 epidemic: from introduction to endemicity. *AIDS* 2014; 28(7): 1031–1040
34. Jovanović L, Šiljić M, Ćirković V, Salemović D, Pešić-Pavlović I, Todorović M, Ranin J, Jevtović D, Stanojević M. Exploring evolutionary and transmission dynamics of HIV epidemic in Serbia: bridging socio-demographic with phylogenetic approach. *Front Microbiol* 2019; 10: 287
35. Bbosa N, Ssemwanga D, Nsubuga RN, Salazar-Gonzalez JF, Salazar MG, Nanyonjo M, Kuteesa M, Seeley J, Kiwanuka N, Bagaya BS, Yebra G, Leigh-Brown A, Kaleebu P. Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations. *Sci Rep* 2019; 9(1): 1051
36. Ragonnet-Cronin M, Jackson C, Bradley-Stewart A, Aitken C, McAuley A, Palmateer N, Gunson R, Goldberg D, Milosevic C, Leigh Brown AJ. Recent and rapid transmission of HIV among people who inject drugs in Scotland revealed through phylogenetic analysis. *J Infect Dis* 2018; 217(12): 1875–1882
37. Paraskevis D, Beloukas A, Stasinou K, Pantazis N, de Mendoza C, Bannert N, Meyer L, Zangerle R, Gill J, Prins M, d'Arminio Montforte A, Kran AB, Porter K, Touloumi G; CASCADE collaboration of EuroCoord. HIV-1 molecular transmission clusters in nine European countries and Canada: association with demographic and clinical factors. *BMC Med* 2019; 17(1): 4
38. de Oliveira T, Kharsany AB, Gräf T, Cawood C, Khanyile D, Grobler A, Puren A, Madurai S, Baxter C, Karim QA, Karim SS. Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. *Lancet HIV* 2017; 4(1): e41–e50
39. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pèpin J, Posada D, Peeters M, Pybus OG, Lemey P. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 2014; 346(6205): 56–61
40. Waruru A, Achia TNO, Tobias JL, Nganga J, Mwangi M, Wamicwe J, Zielinski-Gutierrez E, Oluoch T, Muthama E, Tylleskär T. Finding hidden HIV clusters to support geographic-oriented HIV interventions in Kenya. *J Acquir Immune Defic Syndr* 2018; 78(2): 144–154
41. Butt Z, Grady S, Wilkins M, Hamilton E, Todem D, Gardiner J, Saeed M. Spatial epidemiology of HIV-hepatitis co-infection in the State of Michigan: a cohort study. *Infect Dis (Lond)* 2015; 47(12): 852–861
42. Stecher M, Hoenigl M, Eis-Hbinger AM, Lehmann C, Fätkenheuer G, Wasmuth JC, Knops E, Vehreschild JJ, Mehta S, Chaillon A. Hotspots of transmission driving the local human immunodeficiency virus epidemic in the Cologne-Bonn Region, Germany. *Clin Infect Dis* 2019; 68(9): 1539–1546
43. Chaillon A, Essat A, Frange P, Smith DM, Delaugerre C, Barin F, Ghosn J, Pialoux G, Robineau O, Rouzioux C, Goujard C, Meyer L, Chaix ML; on behalf the ANRS PRIMO Cohort Study. Spatiotemporal dynamics of HIV-1 transmission in France (1999–2014) and impact of targeted prevention strategies. *Retrovirology* 2017; 14(1): 15
44. Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, Kosakovsky Pond SL. The global transmission network of HIV-1. *J Infect Dis* 2014; 209(2): 304–313
45. Fisher M, Pao D, Brown AE, Sudarshi D, Gill ON, Cane P, Buckton AJ, Parry JV, Johnson AM, Sabin C, Pillay D.

- Determinants of HIV-1 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic approach. *AIDS* 2010; 24(11): 1739–1747
46. Yirrell DL, Pickering H, Palmarini G, Hamilton L, Rutemberwa A, Biryahwaho B, Whitworth J, Brown AJ. Molecular epidemiological analysis of HIV in sexual networks in Uganda. *AIDS* 1998; 12(3): 285–290
  47. Ng KT, Ng KY, Chen JH, Ng OT, Kamarulzaman A, Tee KK. HIV-1 transmission networks among men who have sex with men in Asia. *Clin Infect Dis* 2014; 59(6): 910–911
  48. Mehta SR, Wertheim JO, Brouwer KC, Wagner KD, Chaillon A, Strathdee S, Patterson TL, Rangel MG, Vargas M, Murrell B, Garfein R, Little SJ, Smith DM. HIV transmission networks in the San Diego-Tijuana Border Region. *EBioMedicine* 2015; 2(10): 1456–1463
  49. 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(4): 444–451
  50. Ragonnet-Cronin M, Hu YW, Morris SR, Sheng Z, Poortinga K, Wertheim JO. HIV transmission networks among transgender women in Los Angeles County, CA, USA: a phylogenetic analysis of surveillance data. *Lancet HIV* 2019; 6(3): e164–e172
  51. Li X, Liu H, Liu L, Feng Y, Kalish ML, Ho SYW, Shao Y. Tracing the epidemic history of HIV-1 CRF01\_AE clusters using near-complete genome sequences. *Sci Rep* 2017; 7(1): 4024
  52. Li Z, He X, Wang Z, Xing H, Li F, Yang Y, Wang Q, Takebe Y, Shao Y. Tracing the origin and history of HIV-1 subtype B' epidemic by near full-length genome analyses. *AIDS* 2012; 26(7): 877–884
  53. Ye J, Lu H, Su X, Xin R, Bai L, Xu K, Yu S, Feng X, Yan H, He X, Zeng Y. Phylogenetic and temporal dynamics of human immunodeficiency virus type 1B in China: four types of B strains circulate in China. *AIDS Res Hum Retroviruses* 2014; 30(9): 920–926
  54. Meng Z, Xin R, Zhong P, Zhang C, Abubakar YF, Li J, Liu W, Zhang X, Xu J. A new migration map of HIV-1 CRF07\_BC in China: analysis of sequences from 12 provinces over a decade. *PLoS One* 2012; 7(12): e52373
  55. Chen X, Ye M, Pang W, Smith DM, Zhang C, Zheng YT. First appearance of HIV-1 CRF07\_BC and CRF08\_BC outside China. *AIDS Res Hum Retroviruses* 2017; 33(1): 74–76
  56. Han X, Takebe Y, Zhang W, An M, Zhao B, Hu Q, Xu J, Wu H, Wu J, Lu L, Chen X, Liang S, Wang Z, Yan H, Fu J, Cai W, Zhuang M, Liao C, Shang H. A large-scale survey of CRF55\_01B from men-who-have-sex-with-men in China: implying the evolutionary history and public health impact. *Sci Rep* 2015; 5(1): 18147
  57. Han X, An M, Zhang M, Zhao B, Wu H, Liang S, Chen X, Zhuang M, Yan H, Fu J, Lu L, Cai W, Takebe Y, Shang H. Identification of 3 distinct HIV-1 founding strains responsible for expanding epidemic among men who have sex with men in 9 Chinese cities. *J Acquir Immune Defic Syndr* 2013; 64(1): 16–24
  58. An M, Han X, Xu J, Chu Z, Jia M, Wu H, Lu L, Takebe Y, Shang H. Reconstituting the epidemic history of HIV strain CRF01\_AE among men who have sex with men (MSM) in Liaoning, northeastern China: implications for the expanding epidemic among MSM in China. *J Virol* 2012; 86(22): 12402–12406
  59. Li X, Xue Y, Lin Y, Gai J, Zhang L, Cheng H, Ning Z, Zhou L, Zhu K, Vanham G, Kang L, Wang Y, Zhuang M, Pan Q, Zhong P. Evolutionary dynamics and complicated genetic transmission network patterns of HIV-1 CRF01\_AE among MSM in Shanghai, China. *Sci Rep* 2016; 6(1): 34729
  60. Zhu Z, Hu Y, Xing W, Guo M, Zhao R, Han S, Wu B. Identifying symptom clusters among people living with HIV on antiretroviral therapy in China: a network analysis. *J Pain Symptom Manage* 2019; 57(3): 617–626
  61. Zhang Z, Dai L, Jiang Y, Feng K, Liu L, Xia W, Yu F, Yao J, Xing W, Sun L, Zhang T, Wu H, Su B, Qiu M. Transmission network characteristics based on env and gag sequences from MSM during acute HIV-1 infection in Beijing, China. *Arch Virol* 2017; 162(11): 3329–3338
  62. Li X, Zhu K, Xue Y, Wei F, Gao R, Duerr R, Fang K, Li W, Song Y, Du G, Yan W, Musa TH, Ge Y, Ji Y, Zhong P, Wei P. Multiple introductions and onward transmission of HIV-1 subtype B strains in Shanghai, China. *J Infect* 2017; 75(2): 160–168
  63. Chen M, Ma Y, Chen H, Dai J, Dong L, Yang C, Li Y, Luo H, Zhang R, Jin X, Yang L, Cheung AKL, Jia M, Song Z. HIV-1 genetic transmission networks among men who have sex with men in Kunming, China. *PLoS One* 2018; 13(4): e0196548
  64. Shao Y. AIDS molecular network research and AIDS precision prevention and control. 2018 National Conference on HIV and Hepatitis C Prevention and Treatment. April 19, 2018
  65. Wang X, He X, Zhong P, Liu Y, Gui T, Jia D, Li H, Wu J, Yan J, Kang D, Han Y, Li T, Yang R, Han X, Chen L, Zhao J, Xing H, Liang S, He J, Yan Y, Xue Y, Zhang J, Zhuang X, Liang S, Bao Z, Li T, Zhuang D, Liu S, Han J, Jia L, Li J, Li L. Phylodynamics of major CRF01\_AE epidemic clusters circulating in mainland of China. *Sci Rep* 2017; 7(1): 6330
  66. Valverde EE, Oster AM, Xu S, Wertheim JO, Hernandez AL. HIV transmission dynamics among foreign-born persons in the United States. *J Acquir Immune Defic Syndr* 2017; 76(5): 445–452
  67. Campbell EM, Jia H, Shankar A, Hanson D, Luo W, Masciotra S, Owen SM, Oster AM, Galang RR, Spiller MW, Blosser SJ, Chapman E, Roseberry JC, Gentry J, Pontones P, Duwve J, Peyrani P, Kagan RM, Whitcomb JM, Peters PJ, Heneine W, Brooks JT, Switzer WM. Detailed transmission network analysis of a large opiate-driven outbreak of HIV infection in the United States. *J Infect Dis* 2017; 216(9): 1053–1062
  68. Poon AFY, Gustafson R, Daly P, Zerr L, Demlow SE, Wong J, Woods CK, Hogg RS, Krajden M, Moore D, Kendall P, Montaner JSG, Harrigan PR. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV* 2016; 3(5): e231–e238
  69. Pasquale DK, Doherty IA, Sampson LA, Hué S, Leone PA, Sebastian J, Ledford SL, Eron JJ, Miller WC, Dennis AM. Leveraging phylogenetics to understand HIV transmission and partner notification networks. *J Acquir Immune Defic Syndr* 2018; 78(4): 367–375
  70. Avila D, Keiser O, Egger M, Kouyos R, Böni J, Yerly S, Klimkait T, Vernazza PL, Aubert V, Rauch A, Bonhoeffer S, Günthard HF, Stadler T, Spycher BD; Swiss HIV Cohort Study. Social meets molecular: combining phylogenetic and latent class analyses to understand HIV-1 transmission in Switzerland. *Am J Epidemiol* 2014; 179(12): 1514–1525

71. Wertheim JO, Kosakovsky Pond SL, Forgiione LA, Mehta SR, Murrell B, Shah S, Smith DM, Scheffler K, Torian LV. Social and genetic networks of HIV-1 transmission in New York City. *PLoS Pathog* 2017; 13(1): e1006000
72. Kostaki EG, Nikolopoulos GK, Pavlitina E, Williams L, Magiorkinis G, Schneider J, Skaathun B, Morgan E, Psychogiou M, Daikos GL, Sypsa V, Smyrnov P, Korobchuk A, Malliori M, Hatzakis A, Friedman SR, Paraskevis D. Molecular analysis of human immunodeficiency virus type 1 (HIV-1)-infected individuals in a network-based intervention (Transmission Reduction Intervention Project): phylogenetics identify HIV-1-infected individuals with social links. *J Infect Dis* 2018; 218(5): 707–715
73. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT; UK HIV Drug Resistance Collaboration. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. *J Infect Dis* 2011; 204(9): 1463–1469
74. France AM, Panneer N, Ocfemia CB, Saduvala N, Campbell E, Switzer WM, Wertheim J, Oster AM. Rapidly growing HIV transmission clusters in the United States, 2013–2016. 2018 Conference on Retroviruses and Opportunistic Infections. March 4–7, 2018
75. Wertheim JO, Oster AM, Switzer WM, Zhang C, Panneer N, Campbell E, Saduvala N, Johnson JA, Heneine W. Natural selection favoring more transmissible HIV detected in United States molecular transmission network. *Nat Commun* 2019; 10(1): 5788
76. Peters PJ, Pontones P, Hoover KW, Patel MR, Galang RR, Shields J, Blosser SJ, Spiller MW, Combs B, Switzer WM, Conrad C, Gentry J, Khudyakov Y, Waterhouse D, Owen SM, Chapman E, Roseberry JC, McCants V, Weidle PJ, Broz D, Samandari T, Mermin J, Walthall J, Brooks JT, Duwve JM; Indiana HIV Outbreak Investigation Team. HIV infection linked to injection use of oxycodone in Indiana, 2014–2015. *N Engl J Med* 2016; 375(3): 229–239
77. Monterosso A, Minnerly S, Goings S, Morris A, France AM, Dasgupta S, Oster AM, Fanning M. Identifying and investigating a rapidly growing HIV transmission cluster in Texas. Conference on Retroviruses and Opportunistic Infections. March 8, 2017. Seattle, Washington
78. Shang H, Xu J, Han X, Spero Li J, Arledge KC, Zhang L. HIV prevention: bring safe sex to China. *Nature* 2012; 485(7400): 576–577
79. Wertheim JO, Murrell B, Mehta SR, Forgiione LA, Kosakovsky Pond SL, Smith DM, Torian LV. Growth of HIV-1 molecular transmission clusters in New York City. *J Infect Dis* 2018; 218(12): 1943–1953
80. Mehta SR, Murrell B, Anderson CM, Kosakovsky Pond SL, Wertheim JO, Young JA, Freitas L, Richman DD, Mathews WC, Scheffler K, Little SJ, Smith DM. Using HIV sequence and epidemiologic data to assess the effect of self-referral testing for acute HIV infection on incident diagnoses in San Diego, California. *Clin Infect Dis* 2016; 63(1): 101–107
81. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, Eames KT, Edmunds WJ, Frost SD, Funk S, Hollingsworth TD, House T, Isham V, Klepac P, Lessler J, Lloyd-Smith JO, Metcalf CJ, Mollison D, Pellis L, Pulliam JR, Roberts MG, Viboud C; Isaac Newton Institute IDD Collaboration. Modeling infectious disease dynamics in the complex landscape of global health. *Science* 2015; 347(6227): aaa4339
82. Stadler T, Kouyos R, von Wyl V, Yerly S, Böni J, Bürgisser P, Klimkait T, Joos B, Rieder P, Xie D, Günthard HF, Drummond AJ, Bonhoeffer S; Swiss HIV Cohort Study. Estimating the basic reproductive number from viral sequence data. *Mol Biol Evol* 2012; 29(1): 347–357
83. Turk T, Bachmann N, Kadelka C, Büni J, Yerly S, Aubert V, Klimkait T, Battegay M, Bernasconi E, Calmy A, Cavassini M, Furrer H, Hoffmann M, Günthard HF, Kouyos RD, Aubert V, Battegay M, Bernasconi E, Böni J, Braun DL, Bucher HC, Calmy A, Cavassini M, Ciuffi A, Dollegenmaier G, Egger M, Elzi L, Fehr J, Fellay J, Furrer H, Fux CA, Günthard HF, Haerry D, Hasse B, Hirsch HH, Hoffmann M, Hösli I, Kahlert C, Kaiser L, Keiser O, Klimkait T, Kouyos RD, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Marzolini C, Metzner KJ, Müller N, Nicca D, Pantaleo G, Paioni P, Rauch A, Rudin C, Scherrer AU, Schmid P, Speck R, Stöckle M, Tarr P, Trkola A, Vernazza P, Wandeler G, Weber R, Yerly S. Assessing the danger of self-sustained HIV epidemics in heterosexuals by population based phylogenetic cluster analysis. *eLife* 2017; 6: e28721
84. France AM, Oster AM. The promise and complexities of detecting and monitoring HIV transmission clusters. *J Infect Dis* 2019 Apr 27. [Epub ahead of print] doi: 10.1093/infdis/jiz177
85. 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(12): 1226–1235
86. Chaillon A, Delaugerre C, Brenner B, Armero A, Capitant C, Nere ML, Leturque N, Pialoux G, Cua E, Tremblay C, Smith DM, Goujard C, Meyer L, Molina JM, Chaux ML. In-depth sampling of high-risk populations to characterize HIV transmission epidemics among young MSM using PrEP in France and Quebec. *Open Forum Infect Dis* 2019; 6(3): ofz080
87. 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(7): 1812–1819
88. Dennis AM, Pasquale DK, Billock R, Beagle S, Mobley V, Cope A, Kuruc J, Sebastian J, Walworth C, Leone PA. Integration of contact tracing and phylogenetics in an investigation of acute HIV infection. *Sex Transm Dis* 2018; 45(4): 222–228
89. Fitzmaurice AG, Linley L, Zhang C, Watson M, France AM, Oster AM. Novel method for rapid detection of spatiotemporal HIV clusters potentially warranting intervention. *Emerg Infect Dis* 2019; 25(5): 988–991
90. Hightower GK, May SJ, Pérez-Santiago J, Pacold ME, Wagner GA, Little SJ, Richman DD, Mehta SR, Smith DM, Pond SL. HIV-1 clade B pol evolution following primary infection. *PLoS One* 2013; 8(6): e68188
91. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, Williams B, Gouws-Williams E, Ghys PD; WHO–UNAIDS Network for HIV Isolation Characterisation. Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis. *Lancet Infect Dis* 2019; 19(2): 143–155
92. Bbosa N, Kaleebu P, Ssemwanga D. HIV subtype diversity worldwide. *Curr Opin HIV AIDS* 2019; 14(3): 153–160

93. Fabeni L, Alteri C, Berno G, Scutari R, Orchi N, De Carli G, Bertoli A, Carioti L, Gori C, Forbici F, Salpini R, Vergori A, Gagliardini R, Cicalini S, Mondì A, Pinnetti C, Mazzuti L, Turriziani O, Colafigli M, Borghi V, Montella F, Pennica A, Lichtner M, Girardi E, Andreoni M, Mussini C, Antinori A, Ceccherini-Silberstein F, Perno CF, Santoro MM4; SENDIH Study group. Characterisation of HIV-1 molecular transmission clusters among newly diagnosed individuals infected with non-B subtypes in Italy. *Sex Transm Infect* 2019; 95 (8): 619–625
94. Bon I, Ciccozzi M, Zehender G, Biagetti C, Verrucchi G, Lai A, Lo Presti A, Gibellini D, Re MC. HIV-1 subtype C transmission network: the phylogenetic reconstruction strongly supports the epidemiological data. *J Clin Virol* 2010; 48(3): 212–214
95. Parczewski M, Leszczyszyn-Pynka M, Bander D, Urbanska A, Boroń-Kaczmarek A. HIV-1 subtype D infections among Caucasians from Northwestern Poland—phylogenetic and clinical analysis. *PLoS One* 2012; 7(2): e31674
96. Mehta SR, Schairer C, Little S. Ethical issues in HIV phylogenetics and molecular epidemiology. *Curr Opin HIV AIDS* 2019; 14(3): 221–226
97. Prospero MC, De Luca A, Di Giambenedetto S, Bracciale L, Fabbiani M, Cauda R, Salemi M. The threshold bootstrap clustering: a new approach to find families or transmission clusters within molecular quasispecies. *PLoS One* 2010; 5(10): e13619
98. Vrbik I, Stephens DA, Roger M, Brenner BG. The Gap Procedure: for the identification of phylogenetic clusters in HIV-1 sequence data. *BMC Bioinformatics* 2015; 16(1): 355
99. 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(7): 1812–1819
100. 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(5): 719–728