Exploring Dynamic Changes in HIV-1 Molecular Transmission Networks and Its Key Influencing Factors: A Cross-Sectional Study

Yan He, Ying Tang, Qun Hua, Xin Li, You Ge, Yangyang Liu, Rong Tang, Ye Tian, Wei Li

Submitted to: JMIR Public Health and Surveillance
on: January 21, 2024

Table of Contents

Exploring Dynamic Changes in HIV-1 Molecular Transmission Networks and Its Key Influencing Factors: A Cross-Sectional Study

Yan He[1]; Ying Tang[2]; Qun Hua[2]; Xin Li[3]; You Ge[4]; Yangyang Liu[4]; Rong Tang[5]; Ye Tian[2]; Wei Li[2]

[1]Department of Infection Management, Nanjing Drum Tower Hospital Nanjing CN
[2]Children's Hospital of Nanjing Medical University Nanjing CN
[3]Nanjing Municipal Center for Disease Control and Prevention Nanjing CN
[4]School of Public Health, Southeast University Nanjing CN
[5]Nanjing Qixia District Center for Disease Control and Prevention Nanjing CN

Corresponding Author:
Wei Li
Children's Hospital of Nanjing Medical University
72 Guangzhou Rd, Gulou District
Nanjing
CN

Abstract

Background: These dynamic of HIV-1 molecular networks are closely linked to the personal characteristics and behaviors of infected individuals. The real-time analysis of HIV-1 molecular network characteristics allows for the identification and comprehension of the ongoing transformations within molecular networks, thereby facilitating prompt interventions and response strategies

Objective: To identify the dynamic changes occurring in HIV-1 molecular transmission networks and analyze the primary influencing factors driving the dynamics of HIV molecular networks.

Methods: From September 1, 2015, to June 30, 2019, this research was carried out in five districts of Nanjing, with a focus on individuals who were newly diagnosed with HIV infection during this period. The HIV-1 molecular network in Nanjing was constructed by utilizing a combination of phylogenetic tree and gene distance methods. The dynamic changes in the molecular network over a specific time period were analyzed and compared between the baseline and observed endpoint molecular network.

Results: In this study, 955 HIV-1 pol fragments were successfully amplified from 1013 specimens. Subtype identification revealed that CRF01_AE and CRF07_BC were the predominant subtypes, accounting for 40.84% and 33.61% respectively. Through the analysis and comparison of the basic and terminal molecular networks, it was discovered that 144 sequences constituted static molecular network, and 487 sequences contributing to the formation of dynamic molecular networks. The findings of the multivariate analysis indicate that the factors, including occupation as a student, migrant status, Han ethnicity, engagement in occasional or multiple sexual partnerships, participation in anal sex, and being single, are independent risk factors for the dynamic changes observed in the HIV-1 molecular network, and the OR value (95%CI) were 2.63 (1.54~4.47), 1.83 (1.17~2.84), 2.91 (1.09~7.79), 1.75 (1.06~2.90), 4.12 (2.48~6.87), 5.58 (2.43~12.80), 2.10 (1.25~3.54) respectively. Heterosexuality and homosexuality appear to be protective factors against such changes when compared to bisexuality, with OR values (95%CI) of 0.12 (0.05~0.32) and 0.26 (0.11~0.64). Additionally, the national eight-item score and experience with sex education are also identified as protective factors against dynamic changes in the HIV-1 molecular network, the OR values (95%CI) are 0.12 (0.05~0.32) and 0.26 (0.11~0.64), respectively.

Conclusions: The identification of HIV dynamic molecular networks has provided valuable insights into the characteristics of individuals undergoing dynamic alterations. This groundbreaking study focused on analyzing the dynamic changes and determinants of molecular networks, which uncovered independent risk factors associated with these changes.

(JMIR Preprints 21/01/2024:56593)
DOI: https://doi.org/10.2196/preprints.56593

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Exploring Dynamic Changes in HIV-1 Molecular Transmission Networks and Its Key Influencing Factors: A Cross-Sectional Study

Yan He[1†], Ying Tang[2†], Qun Hua[2†], Xin Li[3], You Ge[4], Yangyang Liu[4], Rong Tang[5], Ye Tian[6*], Wei Li[7*]
1. Department of Infection Management, Nanjing Drum Tower Hospital, Nanjing, Jiangsu, 210008, China
2. Department of Ultrasound Diagnostic, Children's Hospital of Nanjing Medical University, Nanjing 210008, China
3. Nanjing Municipal Center for Disease Control and Prevention, Nanjing, Jiangsu, 210003, China.
4. Department of Epidemiology and Health Statistics, School of Public Health, Southeast University, Nanjing, Jiangsu, 210009, China
5. Nanjing Q ixia District Center for Disease Control and Prevention, Nanjing, Jiangsu, 210004, China
6. Department of Infection, Children's Hospital of Nanjing Medical University, Nanjing, Jiangsu, 210008, China
7. Department of Clinical Research, Children's Hospital of Nanjing Medical University, Nanjing, Jiangsu, 210008, China

[*]Correspondence should be addressed to Dr. Wei Li, Email: weili126@126.com and Dr, Ye Tian, Email: medtronic@163.com

[†]These authors contributed equally to this work

Abstract

Background: The HIV-1 molecular network emerges as an innovative tool, utilizing gene sequences to understand transmission attributes and complementing social and sexual network studies. While previous research focused on static network characteristics, recent studies' emphasis on dynamic features enhances our understanding of real-time changes, offering insights for targeted interventions and efficient allocation of public health resources.

Objective: To identify the dynamic changes occurring in HIV-1 molecular transmission networks and analyze the primary influencing factors driving the dynamics of HIV-1 molecular networks.

Methods: We selected individuals newly diagnosed with HIV-1 infection in Nanjing from September 1, 2015, to June 30, 2019 as subjects. The HIV-1 molecular network was constructed using a combination of phylogenetic tree and gene distance methods. We analyzed and compared the dynamic changes in the molecular network over a specific time period between the baseline and

observed endpoint. The primary factors influencing the dynamic changes in the HIV-1 molecular network were identified through univariate analysis and multivariate analysis.
Results: A total of 955 HIV-1 pol fragments were successfully amplified from 1013 specimens, CRF01_AE and CRF07_BC were the predominant subtypes, accounting for 40.84% and 33.61%, respectively. Through the analysis and comparison of the basic and terminal molecular networks, it was discovered that 144 sequences constituted static molecular network, and 487 sequences contributed to the formation of dynamic molecular networks. The findings of the multivariate analysis indicated that the factors, including occupation as a student, Floating Population, Han ethnicity, engagement in occasional or multiple sexual partnerships, participation in anal sex, and being single, were independent risk factors for the dynamic changes observed in the HIV-1 molecular network, and the OR value (95%CI) were 2.63 (1.54~4.47), 1.83 (1.17~2.84), 2.91 (1.09~7.79), 1.75 (1.06~2.90), 4.12 (2.48~6.87), 5.58 (2.43~12.80), and 2.10 (1.25~3.54), respectively. Heterosexuality and homosexuality seem to exhibit protective effects when compared to bisexuality, with OR values (95%CI) of 0.12 (0.05~0.32) and 0.26 (0.11~0.64), respectively. Additionally, the national eight-item score and experience with sex education were also identified as protective factors against dynamic changes in the HIV-1 molecular network, the OR values (95%CI) were 0.12 (0.05~0.32) and 0.26 (0.11~0.64), respectively.
Conclusion: HIV-1 molecular network analysis showed 144 sequences in static networks and 487 in dynamic networks. Multivariate analysis revealed that occupation as a student, floating Population, ethnicity (Han) and risky sexual behavior were independent risk factors for dynamic changes, while heterosexuality and homosexuality were protective compared to bisexuality. Higher national eight-item score and experience of sex education were also protective factors. The identification of HIV dynamic molecular networks has provided valuable insights into the characteristics of individuals undergoing dynamic alterations. These findings contribute to a better understanding of HIV-1 transmission dynamics and could inform targeted prevention strategies.
Keywords: HIV, Dynamic, Molecular Transmission Network, Influence Factors

Introduction

AIDS, a sexually transmitted disease, is significantly influenced by the social and sexual network structure and characteristics of people living with HIV (PLHIV)[1, 2]. Understanding the social and sexual network structure of individuals infected with HIV is of great public health significance, as it facilitates comprehension of transmission dynamics within the population, and informs targeted prevention and control measures[3-5]. Prior research examining the social and sexual network structure of PLHIV has predominantly utilized field epidemiological techniques, including questionnaire surveys, peer tracing, and community follow-ups, to delineate the attributes of social and sexual transmission networks[6-9]. However, the extended latency period of HIV and the time gap between infection and diagnosis create a challenge for social communication networks that rely on self-reported data from infected individuals[10]. The quality of self-reported data provided by infected individuals may be influenced by various factors such as social discrimination, stigma, and privacy concerns. These factors can create barriers to accurate reporting and may result in underreporting or misrepresentation of information. Consequently, conventional epidemiological investigations, including behavioral surveys, exposure history assessments, and contact tracing, encounter difficulties in analyzing the structure and attributes of HIV transmission networks[11]. Alternative approaches and methodologies may need to be considered to overcome these difficulties and obtain a more comprehensive understanding of the HIV transmission network.

The HIV-1 molecular network has garnered extensive acknowledgement as an innovative technique that employs the gene sequences of HIV-1 infected individuals to examine its transmission attributes[12]. It functions as a valuable supplement to the study of HIV-1 social and sexual networks[13-16]. The implementation and progression of the HIV-1 molecular network have introduced novel viewpoints for the prevention and control of HIV/AIDS[17-20]. The molecular network has been utilized to assess the efficacy of antiretroviral therapy in preventing secondary HIV transmission[21]. This approach is deemed a precise mechanism for HIV/AIDS prevention and control, enabling the more effective utilization of public health resources[22]. By identifying individuals or groups who are at a higher risk of transmitting HIV-1, targeted interventions can be implemented to prevent the spread of the virus and improve treatment outcomes. Furthermore, extensive investigations can promptly identify undiagnosed infected individuals within the transmission network. Moreover, the implementation of timely interventions among uninfected individuals at risk networks can effectively impede the further dissemination and spread of HIV-1, thereby holding promising prospects for curtailing the spread of HIV and reducing the incidence of new infections. It becomes possible to optimize prevention efforts and achieve better outcomes in controlling the spread of HIV-1[23].

Prior research on the impact of molecular networks on HIV risk has predominantly concentrated on static network characteristics, assessing the structural attributes of the network at a particular moment. Factors such as network size, density, and individual characteristics (e.g. age, sex, occupation, risky sexual behavior, HIV-related

knowledge, and attitudes) have been identified as important considerations in understanding HIV transmission patterns and implementing effective prevention and control measures[24, 25].

Recent research has underscored the importance of dynamic features within the HIV molecular network[26, 27]. Identified actively growing clusters, despite demographic and risk characteristics that may diverge from the overall population, warrant focused intervention prioritization. The dynamics of cluster growth offer valuable guidance for the allocation and prioritization of public health resources, thereby bolstering the utility of networks in assessing intervention efficacy. These dynamic changes intricately intertwine with the personal characteristics and behaviors of infected individuals. Real-time analysis of HIV-1 molecular network characteristics provides a framework for resource allocation towards swiftly evolving molecular clusters. This approach facilitates the identification and comprehension of ongoing transformations within molecular networks, thus enabling prompt interventions and response strategies[26, 28]. However, prevailing research on molecular networks in China predominantly focuses on static networks at a single time point, with limited exploration of dynamic changes[29-32]. Our study specifically targets newly formed and dynamically growing molecular clusters to elucidate the dynamic changes within HIV-1 molecular transmission networks. By analyzing the primary influencing factors driving these dynamics, we aim to unravel the intricate nature of HIV molecular networks. Ultimately, this research endeavor seeks to empower public health initiatives with the capability to more effectively target HIV spread and implement tailored prevention measures through the active monitoring and analysis of molecular network dynamics.

Materials and Methods

Participants

From September 1, 2015, to June 30, 2019, this research was conducted in five districts of Nanjing: Qinhuai, Xuanwu, Qixia, Jiangning, and Gulou. The study focused on individuals who were newly diagnosed with HIV infection during this specific time frame. The Nanjing Center for Disease Control and Prevention (CDC) was responsible for conducting confirmation testing on all HIV-infected individuals included in the study.

Questionnaire Survey

A structured questionnaire was used for the one-on-one survey, which took place in a separate room. The privacy of the participant was prioritized throughout the questionnaire survey procedure, and strict measures were implemented to uphold the confidentiality of participants' personal data. The questionnaire covered various topics related to the participants. These included demographic information such as age, gender, occupation as a student or not, marital status, religion, ethnicity, transmission route, sexual orientation, geographical localities (floating population or not) and sex transmission disease history. Additionally, the questionnaire explored participants'

knowledge and behaviors related to HIV, including condom use, casual sexual partners, multiple sexual partners anal sex, use of enhancers, experiences with sex education, and HIV/AIDS knowledge was assessed using an 8-item questionnaire(National eight-item) score, which was designed by the Chinese CDC[33, 34] (**Supplemental Table 1**).

Specimen Collection and Storage

Blood samples were collected from survey participants using EDTA anticoagulant tubes and transported to the Nanjing Center for Disease Control and Prevention, and the samples were processed within 12 hours. Plasma, lymphocyte enrichment solution, and red blood cells were separated and divided into separate tubes. The aliquoted cryopreservation tubes were labeled and stored at -80°C.

HIV-1 pol Region Fragment Amplification

The amplification of the HIV-1 virus was executed through the utilization of reverse transcription polymerase chain reaction (RT-PCR) and nested polymerase chain reaction (nPCR) methodologies, in accordance with the protocols as described earlier[35]. The amplification process was directed towards a segment of the *pol* region of the specimen, with a specific focus on the 1-99 amino acids of the protease region and the 1-254 amino acids of the reverse transcriptase region. (HXB2: 2253 to 3312). The length of the amplified sequence was 1060 bp. In cases where the amplified sequence was shorter than 1060 bp or samples failed to amplify, re-amplification and sequencing were performed. HIV drug resistance was demonstrated as our previous study[35].

Identification of HIV-1 Subtypes in Nanjing

The gene sequences that were sequenced with success underwent sorting and splicing procedures, while the possibility of contamination was addressed through the utilization of the online plagiarism detection tool, ElimDupes[36]. The present study employed the HIV-BLAST online tool for the identification and genotyping of HIV-1 in Nanjing. Furthermore, gene subtype reference sequences, such as CRF01_AE, CRF07_BC, CRF08_BC, CRF5501_B, B, CRF6701_B, CRF6801_B, were procured from the HIV sequence database website, and a reference sequence dataset was constructed in conjunction with the Nanjing HIV-1 gene sequence data. The gene evolution tree was constructed using an approximately-maximum likelihood method in FastTree v 2.1 for comparative analysis, and the distribution of HIV-1 genotypes in Nanjing was determined based on the findings of the evolution tree.

The HIV-1 molecular network in Nanjing

An approximately-maximum likelihood phylogenetic tree was constructed using FastTree v2.1. The tree was built under the General Time Reversible (GTR) + Gamma distribution (G) + proportion of invariable sites (I) nucleotide substitution model. The Shimodaira-Hasegawa-like test was performed to assess the support value for each node in the phylogenetic tree. A support value of 90% was used as the threshold for considering a node well-supported. The genetic distance between pairs of sequences

was calculated using the Tamura-Nei 93 (TN93) method, a genetic distance threshold of less than or equal to 0.045 was used to identify potential transmission clusters[37, 38]. The HIV-1 molecular network in Nanjing was constructed by utilizing a combination of phylogenetic tree and gene distance (90%+0.045) methods, which was demonstrated in our previous study[39]. The visual editing of HIV molecular transmission networks was accomplished using Cytoscape 3.10.1, a widely utilized platform for constructing molecular networks[40, 41].

Identification of Dynamic Changes in HIV-1 Molecular Network

The molecular network constructed from 2015 to 2017 served as the baseline molecular network (initial time point), while the molecular network observed from 2015 to 2019 represented the observed endpoint molecular network (final time point). By comparing the baseline and observed endpoint molecular networks, the following definitions were derived: The static molecular network consists of unchanged molecular clusters and molecular clusters that do not appear in the observed endpoint molecular network. This means that the structure and members of these clusters remain constant during the observation period. The dynamic molecular network refers to the newly formed molecular clusters observed in the observed endpoint molecular network and the dynamically growing molecular clusters based on the baseline molecular network (**Figure 1**). This indicates that the clusters were formed within the observation period and may continue to grow or change over time.

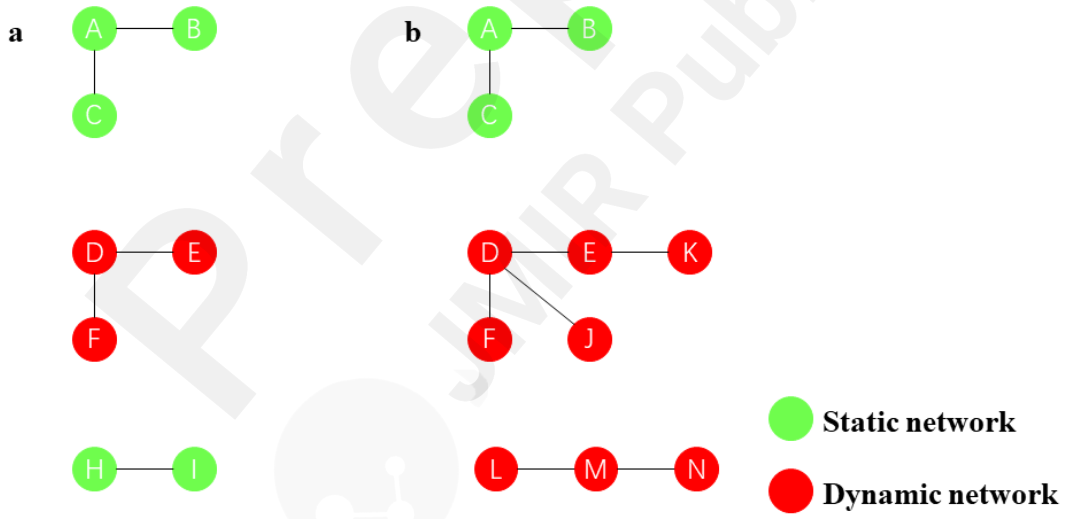


Figure 1 The dynamic changes in molecular network recognition

Figure 1-a baseline molecular network, Figure 1-b the observed endpoint molecular network. The molecular cluster composed of A, B, and C is the unchanged molecular cluster. The molecular cluster composed of H and I do not appear in the observed endpoint molecular network. The molecular cluster composed of D, E, F, J and K is a dynamically growing molecular cluster. The molecular cluster composed of L, M and N is the newly formed molecular cluster. The red spheres represent the dynamic molecular network, while the green spheres represent a statical molecular network.

Statistical Analysis

Epidata software wa used to construct a database, two individuals each with a computer input the questionnaire u, and the questionnaire's completeness were verified by means of telephone follow-ups or secondary surveys to supplement any missing information. The classification data in this study were expressed as a percentage (%), and the chi-square test was employed to compare various groups. Additionally, multivariate analysis was conducted using Logistic regression analysis, with a test level of α=0.05.

Ethics Approval

Written informed consent was obtained from all patients participating in the study , and the research protocol was approved by the Medical Ethics Committee of Zhongda Hospital of Southeast University (ID: 2017ZDKYSB045).

Results

Subtypes of HIV-1 pol Fragment Genome in Nanjing from 2015 to 2019

In this study, 955 HIV-1 pol fragments were successfully amplified from 1013 specimens, the success rate of amplification was 94.27%. Subtype identification revealed that CRF01_AE and CRF07_BC were the predominant subtypes, accounting for 40.84% and 33.61% respectively. The remaining subtypes included URF, CRF67_01B, CRF68_01B, B, CRF55_01B, CRF08_BC, CRF58_01B, CRF59_01B, CRF87_cpx, and C, which accounted for 9.01%, 4.61%, 3.66%, 3.46%, 2.30%, 1.26%, 0.63%, 0.42%, 0.10%, and 0.10% respectively.

Identification of HIV-1 Dynamic Molecular Network in Nanjing from 2015 to 2019

The baseline molecular network is the HIV-1 molecular network in Nanjing from 2015 to 2017 (**Figure 2**). A total of 295 HIV-1 sequences entered the network, with a clustering ratio of 56.84% (295/591), forming 75 molecular clusters. Among them, there were 38 CRF01_AE molecular clusters (122 sequences), 23 CRF07_BC molecular clusters (90 sequences), 2 CRF08_BC molecular clusters (5 sequences), 2 CRF6701_B molecular clusters (23 sequences), and 1 CRF6801_B molecular cluster (17 sequences), 1 CRF5501_B molecular cluster (8 sequences), 3 B subtype molecular clusters (8 sequences), 4 URF molecular clusters (19 sequences), and 1 CRF5901_B molecular cluster (3 sequences).

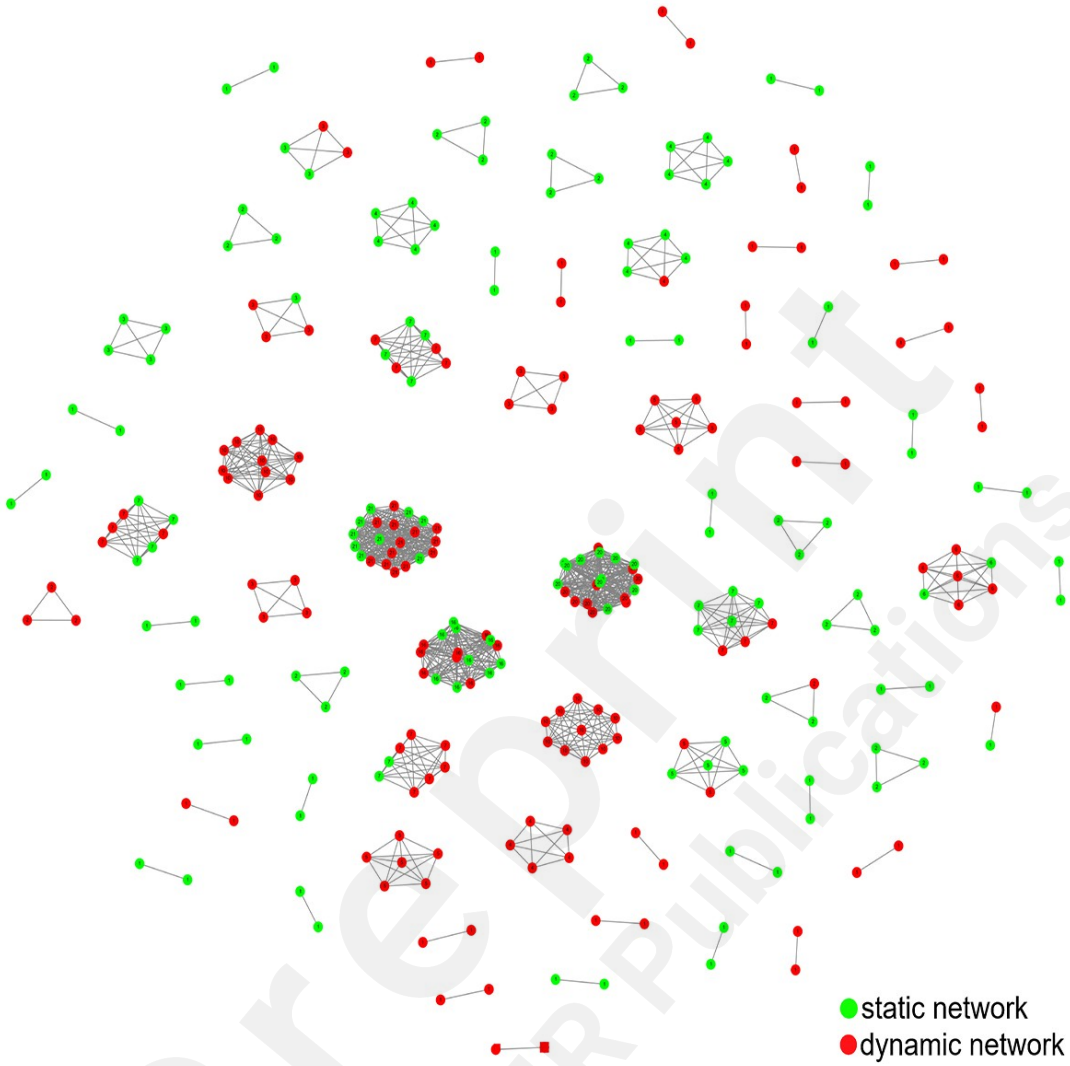


Figure 2 The HIV-1 molecular network in Nanjing from 2015 to 2017 (The baseline molecular network).

The red spheres represent a dynamic molecular network, while the green spheres represent a statical molecular network. The number in the center of each sphere represents the degree of the molecular network

The observation endpoint molecular network is the HIV-1 molecular network in Nanjing from 2015 to 2019 (**Figure 3**). A total of 565 HIV-1 sequences were included in the network, the clustering ratio was 59.16% (565/955), forming 124 molecular clusters. Among them, there were 57 CRF01_AE molecular clusters (233 sequences), 34 CRF07_BC molecular clusters (188 sequences), 2 CRF08_BC molecular clusters (6 sequences), 5 CRF6701_B molecular clusters (34 sequences), and 3 CRF6801_B molecular clusters (19 sequences), 3 CRF5501_B molecular clusters (14 sequences), 8 subtype B molecular clusters (20 sequences), 10 URF molecular clusters (46 sequences), and 2 other molecular clusters including 5 sequences (CRF5801_B 2 pieces, CRF5901_B 3 pieces).

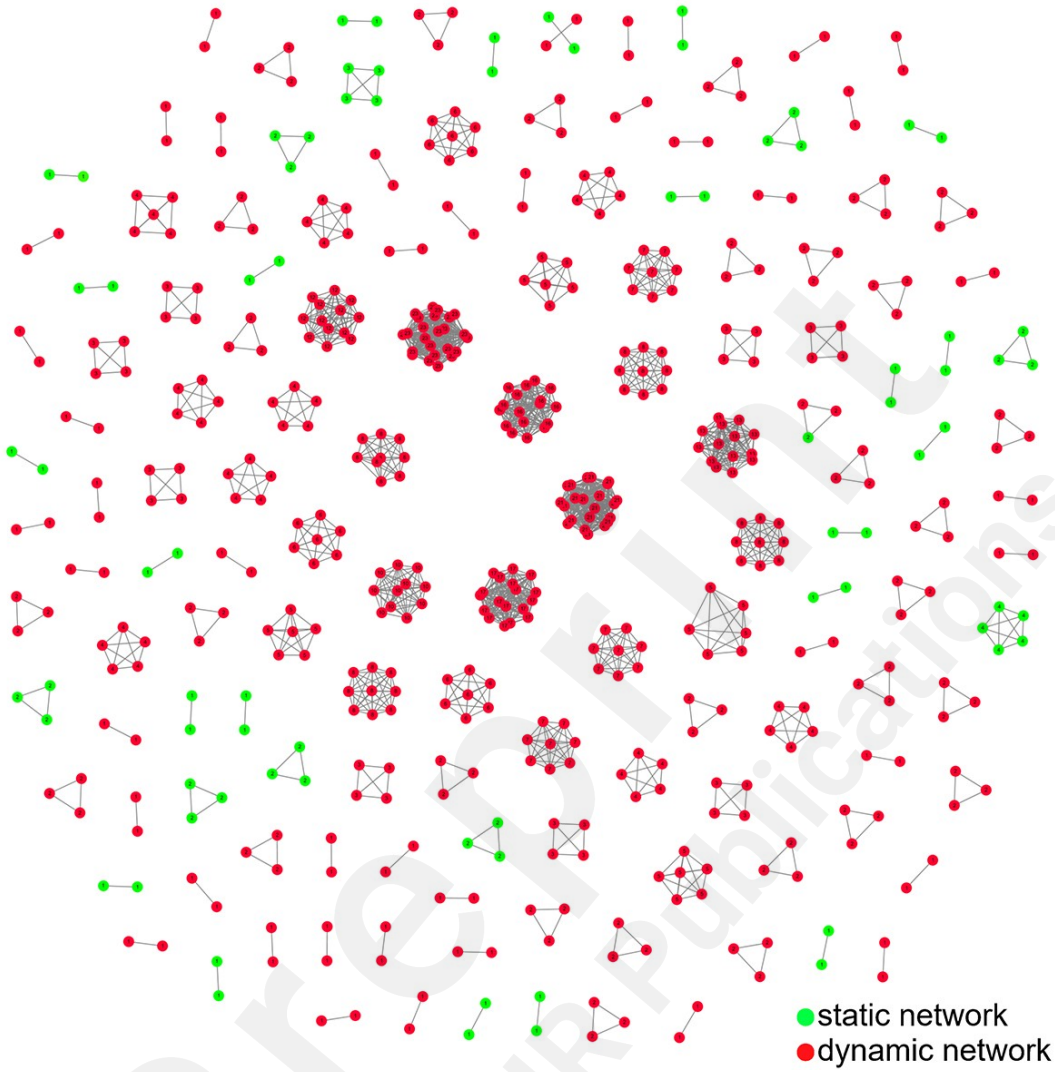


Figure 3 The HIV-1 molecular network in Nanjing from 2015 to 2019 (The observed endpoint molecular network).
The red spheres represent a dynamic molecular network, while the green spheres represent a statical molecular network. The number in the center of each sphere represents the degree of the molecular network

By comparing the baseline and observed endpoint molecular networks, it was discovered that 144 sequences remained static or disappeared, these sequences formed molecular clusters that constituted a static molecular network. On the other hand, 487 sequences were found in newly formed or dynamically growing molecular clusters, contributing to the formation of dynamic molecular networks within the clusters where these sequences were located. In addition, the remaining 324 gene sequences were excluded from the analysis of the dynamic molecular network because they were absent from both the basic molecular network and the endpoint molecular network.

The Impact of General Personal Characteristics on the Composition of Molecular Networks

The molecular network composition (dynamic and static) was taken as the dependent variable for univariate analysis. The findings indicated that the proportion of male (77.18%) infected individuals entering the dynamic molecular network was marginally higher than that of female (69.23%), although no statistical significant difference was observed. Notably, a significantly higher proportion of infected individuals aged 20 years and below entered the dynamic molecular network, with 89.23% of such individuals being represented, compared to those over 20 years old (75.40%), the difference was statistically significant (*P*=0.015). A significantly higher proportion of infected students (82.80%) were found to have entered the dynamic molecular network compared to non-student infected individuals (74.31%) (*P*=0.022). The study found significant variations in the proportion of infected individuals with different sexual orientations entering the dynamic molecular network. Specifically, the proportion of bisexual infected individuals entering the network was the highest (81.68%). Additionally, the proportion of infected individuals from the floating population was significantly higher than that of the non-migrant population (80.87% VS 73.62%), and the proportion of Han infected individuals entering the dynamic molecular network was higher than that of minority infected individuals. The study also observed a higher proportion of infection among single individuals. But there were no statistically significant differences in the proportion of infected persons with different education levels, infection routes, religion, and history of sexuality (**Table 1**).

Table1 The Impact of General Personal Characteristics on the Composition of Molecular Networks

Variables	Static molecular network（n=144）	Dynamic molecular network（n=487）	x^2	*P*
Gender			0.89	0.347
Female	8（30.77%）	18（69.23%）		
Male	136（22.82%）	460（77.18%）		
Age groups			6.26	0.015
≤20（years）	7（10.77%）	58（89.23%）		
（20（years）	137（24.60%）	420（75.40%）		
Occupation as a student			5.27	0.022
No	112（25.69%）	324（74.31%）		
Yes	32（17.20%）	154（82.80%）		
Education level			1.3	0.246

			5	
Blow college	52（26.00%）	148（74.00%）		
College and above	92（21.80%）	330（78.20%）		
			0.0	
Transmission route			1	0.993
Homosexual behavior	114（23.12%）	379（76.88%）		
Heterosexual activity	26（23.42%）	85（76.58%）		
Other	4（22.22%）	14（77.79%）		
			6.4	
Sexual orientation			0	0.041
Heterosexual	25（24.27%）	78（75.73%）		
Homosexual	71（27.63%）	186（72.37%）		
Bisexual	48（18.32%）	214（81.68%）		
Geographical			4.5	
localities			3	0.033
permanent residents	91（26.38%）	254（73.62%）		
floating population	53（19.13%）	224（80.87%）		
			6.3	
Ethnicity			8	0.012
ethnic minority	10（45.45%）	12（54.55%）		
Han (Han Chinese)	134（22.33%）	466（77.67%）		
			0.0	
Religious belief			1	0.952
No	125（23.19%）	414（76.81%）		
Yes	19（22.89%）	64（77.11%）		
			7.1	
Marital status			0	0.008
Not Single	111（26.24%）	312（73.76%）		
Single	33（16.58%）	166（83.42%）		
Sex transmission			0.2	
disease history			4	0.625
No	104（22.66%）	355（77.34%）		
Yes	40（24.54%）	123（75.46%）		

The Impact of HIV-1-related Knowledge and Risky Behaviors on the Molecular Network Composition of Infected Persons

There was no significant difference in the proportion of infected persons entering the dynamic molecular network with different condom usage conditions (*P*=0.788). However, the proportion of infected persons with casual sexual partners (79.54%) was

higher than that of patients without casual partners (67.83%) (*P*=0.004). Moreover, the proportion of individuals who reported having multiple sexual partners, and engaged in anal sex, entering the dynamic molecular network was significantly higher compared to those who did not report such behaviors, these differences were statistically significant. Also, the proportion of infected persons with qualified scores in the eight national items who had received sex education entering the dynamic molecular network was significantly lower than the proportion of infected persons with unqualified scores in the eight national items and had not received sex education. However, there was no statistical difference in the proportion of infected individuals who reported using enhancers compared to those who did not (*P*=0.962). See **Table 2** for details.

Table 2 The impact of HIV-1 Infection Knowledge and Risky Behaviors on Molecular Network Composition

Variables	Static Molecular Network (n=144)	Dynamic Molecular Network (n=487)	x^2	*P*
Condom Use			1.05	0.788
Always	20 (21.74%)	72 (78.26%)		
Frequently	54 (22.04%)	191 (77.96%)		
Occasionally	53 (25.60%)	154 (74.40%)		
Never	17 (21.79%)	61 (78.21%)		
Casual sexual Partners			8.49	0.004
No	46 (32.17%)	97 (67.83%)		
Yes	98 (20.46%)	381 (79.54%)		
Multiple Partners			19.54	□0.001
No	56□36.13%□	99□63.87%□		
Yes	88□18.84%□	379□81.16%□		
Anal Sex			8.29	0.004
No	40□33.06%□	81□66.94%□		
Yes	104□20.76%□	397□79.24%□		
Use of Enhancers			0.00	0.962
No	97□23.10%□	323□76.90%□		
Yes	47□23.27%□	155□76.73%□		
Sex Education			4.56	0.033
No	38□18.10%□	172□81.90%□		
Yes	106□25.73%□	306□74.27%□		
National Eight-point Score			6.39	0.011
Unqualified	25□15.82%□	133□84.18%□		

Qualified	119（25.65%）	345（74.35%）

Effect of Subtype, Drug Resistance and CD4 value on Molecular Network Composition

The statistical analysis revealed a significant difference in the proportion of infected individuals with CD4 value <500 cells/mm^3 entering the dynamic molecular network (79.08%) compared to those with CD4 value ≥500 cells/mm^3 (70.55%) (*P*=0.029). However, no statistically significant differences were observed in the proportion of infected individuals with different genotypes or drug resistance entering the dynamic molecular network. Further details can be found in **Table 3**.

Table 3 The effect of Subtype, Drug Resistance and CD4 value on Molecular Network Composition

Variables		Static Molecular Network (n=144)	Dynamic Molecular Network (n=487)	x^2	*P*
Genetic Subtypes				3.00	0.223
	CRF01_AE	62（25.20%）	184（74.80%）		
	CRF07_BC	40（19.05%）	170（80.95%）		
	other	42（25.30%）	124（74.70%）		
Drug Resistance				1.79	0.181
	No	120（22.26%）	419（77.74%）		
	Yes	24 （28.92%）	59（71.08%）		
CD4+(cells/mm3)				4.92	0.027
	<500	96（20.92%）	363（79.08%）		
	≥500	48（29.45%）	115（70.55%）		

Logistic Regression Multivariate Analysis

Taking the molecular network status of HIV-1 infected individuals as the dependent variable (dynamic = 1, static = 0), the variables with statistical significance in univariate analysis (occupation, age, ethnicity, sexual orientation, floating population, casual sexual partner, multiple sexual partners, national eight-item scores, anal intercourse, singleness, sex education experience, CD4) were included in the multivariate analysis. The assignments were as follows: occupation (student=1, non-student=0), age (≤20 years old=1, >20 years old=0), ethnicity (Han=1, minority=0), sexual orientation (heterosexual=1, same-sex=2, bisexual=3), floating population (yes=1, no=0), casual partner (yes =1, no=0), multiple sexual partners (yes=1, no=0), national eight-item score (qualified=1, unqualified=0), anal intercourse (yes=1, no=0), marital status (Single=1, not single=0), sex education experience (yes=1, no=0), CD4 (≥500 cells/mm3=1, <500 cells/mm3=0).

The findings of the multivariate analysis indicate that various factors, including occupation as a student, migrant status, Han ethnicity, engagement in occasional or multiple sexual partnerships, participation in anal sex, and being single, are independent risk factors for the dynamic changes observed in the HIV-1 molecular network, and the OR value (95%CI) were 2.63 (1.54~4.47), 1.83 (1.17~2.84), 2.91 (1.09~7.79), 1.75 (1.06~2.90), 4.12 (2.48~6.87), 5.58 (2.43~12.80), 2.10 (1.25~3.54) respectively. Conversely, heterosexuality and homosexuality appear to be protective factors against such changes when compared to bisexuality, with OR values (95%CI) of 0.12 (0.05~0.32) and 0.26 (0.11~0.64). Additionally, the national eight-item score and experience with sex education are also identified as protective factors against dynamic changes in the HIV-1 molecular network, the OR values (95%CI) are 0.12 (0.05~0.32) and 0.26 (0.11~0.64), respectively. Further details can be found in **Table 4.**

Table 4 Logistic Regression Analysis of Factors Influencing the Dynamic Changes in HIV-1 Molecular Networks

Variables	β	SE	Wald χ2	*P*	OR	(95%CI)
Occupation (Student)	0.97	0.27	12.64	<0.001	2.63	1.54~4.47
Heterosexual	-2.10	0.49	18.73	0.003	0.12	0.05~0.32
Homosexual	-1.35	0.46	8.74	0.002	0.26	0.11~0.64
Geographical localities (floating population)	0.60	0.23	7.14	0.008	1.83	1.17~2.84
Ethnicity (Han Chinese)	1.07	0.50	4.54	0.033	2.91	1.09~7.79
Casual Partners (Yes)	0.56	0.26	4.85	0.028	1.75	1.06~2.90
Multiple sexual Partners (Yes)	1.42	0.26	29.62	<0.001	4.12	2.48~6.87
National Eight-point Score (Qualified)	-1.58	0.33	23.7	<0.001	0.21	0.11~0.39
Anal Sex (Yes)	1.72	0.42	16.50	0.000	5.58	2.43~12.80
Marital Status (Single)	0.74	0.27	7.83	0.005	2.10	1.25~3.54
Sex Education (Yes)	-0.52	0.24	4.76	0.029	0.59	0.37~0.95
Constant	-0.26	0.59	0.19	0.659	0.77	

Discussion

The study revealed a diverse distribution of HIV-1 subtypes in Nanjing, with CRF01_AE and CRF07_BC being the predominant subtypes. This observation aligns with the global trend of HIV-1 subtype distribution, with CRF01_AE and CRF07_BC commonly found in Asian regions[42, 43]. The identification of various other subtypes, including URFs and less prevalent subtypes, emphasizes the genetic complexity of the HIV epidemic in the studied population.

The dynamic changes in the molecular networks were analyzed to understand the dynamics of HIV-1 transmission[44]. The network's dynamic changes are indicative of potential transmission associations within it. The addition of new gene sequences can disrupt the existing network and give rise to novel networks, thereby revealing the

evolving transmission patterns with greater precision. This approach facilitates the real-time monitoring of the molecular network's dynamic changes, thereby enabling the prompt identification of rapidly expanding or emerging molecular clusters. Such information can inform targeted interventions and the evaluation of prior prevention and control measures[45].

By conducting a multi-factor analysis of the HIV-1 molecular network's dynamic changes, it was observed that individuals who engage in high-risk behaviors, such as bisexuality, multiple sexual partners, casual sexual partners, and anal sex, are more likely to be present in the dynamic molecular network. MSM (men who have sex with men) individuals in China, in particular, may face social and familial pressures that lead them choose to marry with female[46], but engage in extramarital same-sex relationships, thereby acting as a bridge for HIV transmission from high-risk groups to the general population[47]. Bisexual individuals have a higher incidence of unprotected sex, lower risk perception, and weaker awareness of self-protection, making them more susceptible to HIV transmission. Furthermore, individuals constrained by family and involved in fixed marital relationships have limited opportunities to engage with fixed same-sex partners, leading them to seek commercial same-sex services when they have sexual needs. And individuals who engage in sexual activity with multiple partners are prone to inconsistent condom use in contrast to those who engage in sexual activity with only one partner. The emergence of online social platforms has facilitated casual sexual encounters, resulting in a surge of inadvertent sexual behaviors. The failure to consistently use condoms during casual sexual activity significantly heightens the risk of HIV infection and transmission. The physiological vulnerability of the anus renders anal sex a higher risk activity, and the lack of condom use during same-sex anal intercourse further exacerbates the risk of infection and transmission[48].

The findings of the analysis indicate that individuals with dynamic network connections are more likely to be affiliated with migrant, single, and student populations. The floating population serves as a significant conduit for HIV transmission across various regions. Furthermore, the population in question predominantly comprises of sexually active young adults who spend prolonged periods away from their families, resulting in a dearth of both physical and emotional support. This lack of restraint renders them susceptible to engaging in high-risk behaviors related to HIV. The unattached demographic, unencumbered by familial obligations, exhibits a greater propensity for engaging in precarious conduct, such as engaging in casual sexual liaisons and maintaining multiple sexual partners, thereby heightening their susceptibility to contracting HIV and propagating the virus. In contemporary times, college students have emerged as a crucial cohort for HIV prevention and management. Despite possessing a higher level of education, this group manifests a significant disparity between their awareness and conduct. Their inclination towards novelty-seeking, coupled with the proliferation of diverse social media and mobile applications, finding sexual partners online has become extremely convenient, further exacerbates this phenomenon. A study conducted by our research team revealed that a majority of university students infected with HIV contracted the

virus through male-to-male sexual behaviors, with mobile apps serving as the primary platform for locating male sexual partners. The male-to-male sexual contacts exhibit a lack of awareness regarding the high prevalence of HIV/AIDS in their population, and they do not consistently utilize condoms during sexual encounters[49]. Higher national eight-item scores and sex education experience also emerged as protective factors, suggesting the potential role of knowledge and awareness in reducing transmission dynamics.

In response to these findings, Nanjing has implemented targeted measures to enhance the accessibility of HIV/AIDS pre-exposure prophylaxis (PrEP) post-exposure prophylaxis (PEP) services among MSM, especially the floating population of Nanjing[50, 51]. These measures include improving the availability of PrEP and PEP medication and related healthcare services, developing educational programs to increase awareness and knowledge about PrEP and PEP, and engaging with MSM communities through community-based organizations and support groups.

Conclusion

We conducted HIV molecular network analysis in Nanjing from 2015 to 2019, CRF01_AE and CRF07_BC were identified as the predominant subtypes. Molecular network analysis revealed dynamic changes in HIV molecular network over time, with 487 sequences contributing to newly formed or dynamically growing molecular clusters. Multivariate analysis confirmed occupation as student, floating population, Han ethnicity, engaging in casual or multiple sexual partnerships, participation in anal sex, and being single as independent risk factors for dynamic network changes, while heterosexuality and homosexuality appeared to be protective factors when compared with bisexuality. Additionally, higher national eight-item scores and sex education experience were identified as protective factors against dynamic network changes. The findings emphasize the importance of targeted interventions addressing specific risk factors identified in the study. Strategies focused on education, awareness, and behavioral interventions may contribute to stabilizing or reducing the dynamic changes observed in the HIV-1 molecular network. Understanding the intricate dynamics of transmission networks is crucial for designing effective public health measures to control and prevent further spread of HIV.

Limitations

This study has the following limitations. Firstly, the definition of dynamic network in this study is only based on the comparison of two time-point networks, and multiple time-point comparisons should be used to discover nascent networks and dynamic growth networks. Secondly, although we used the HIV pol gene sequence from 2015 to 2019 to analyze the dynamic changes in the molecular network, the sample size was not large enough. Thirdly, despite conducting the investigation in a one-on-one format, there remains a possibility of recall bias attributable to HIV-related stigma, compounded by the fact that the time of infection is unknown for many individuals.

Availability of datasets

We have submitted our sequences data to GenBank（submission ID is: 2796534）, the datasets are available from the corresponding author Wei Li (weili126@126.com) after the co-authors approve of the request.

Founding

This work was supported by China Postdoctoral Science Foundation (2022M721682)

Conflict of interests

None of the authors has any conflict of interest to disclose.

Author contributions

WL and YT were responsible for the conception and design of the study; YH, YT were major contributors in writing original draft; QH and XL were responsible for the acquisition of data, YG and YL contributed to the investigation, YG and RT helped in the analysis and interpretation of data; WL and YH contributed to the final approval of the version to be submitted. All authors read and approved the final version of the manuscript.

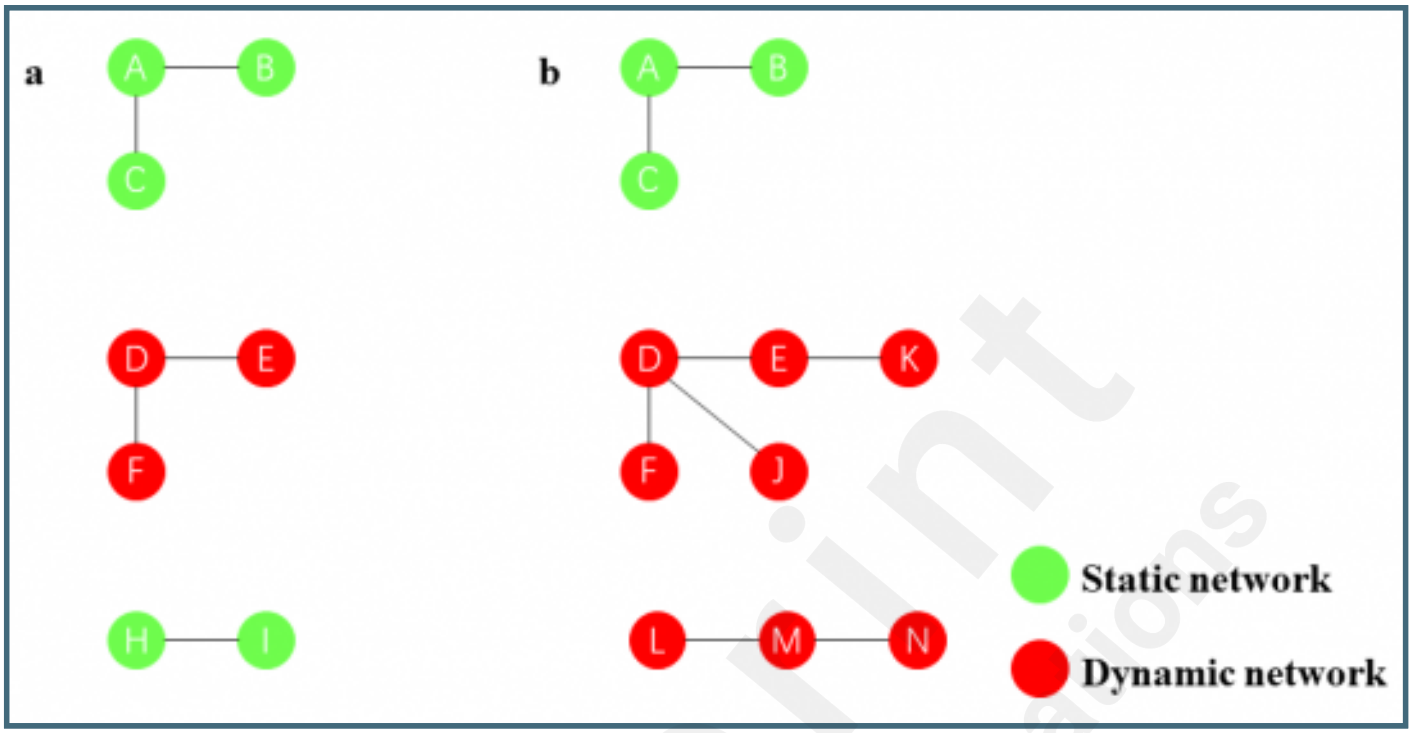
References

Figure legend

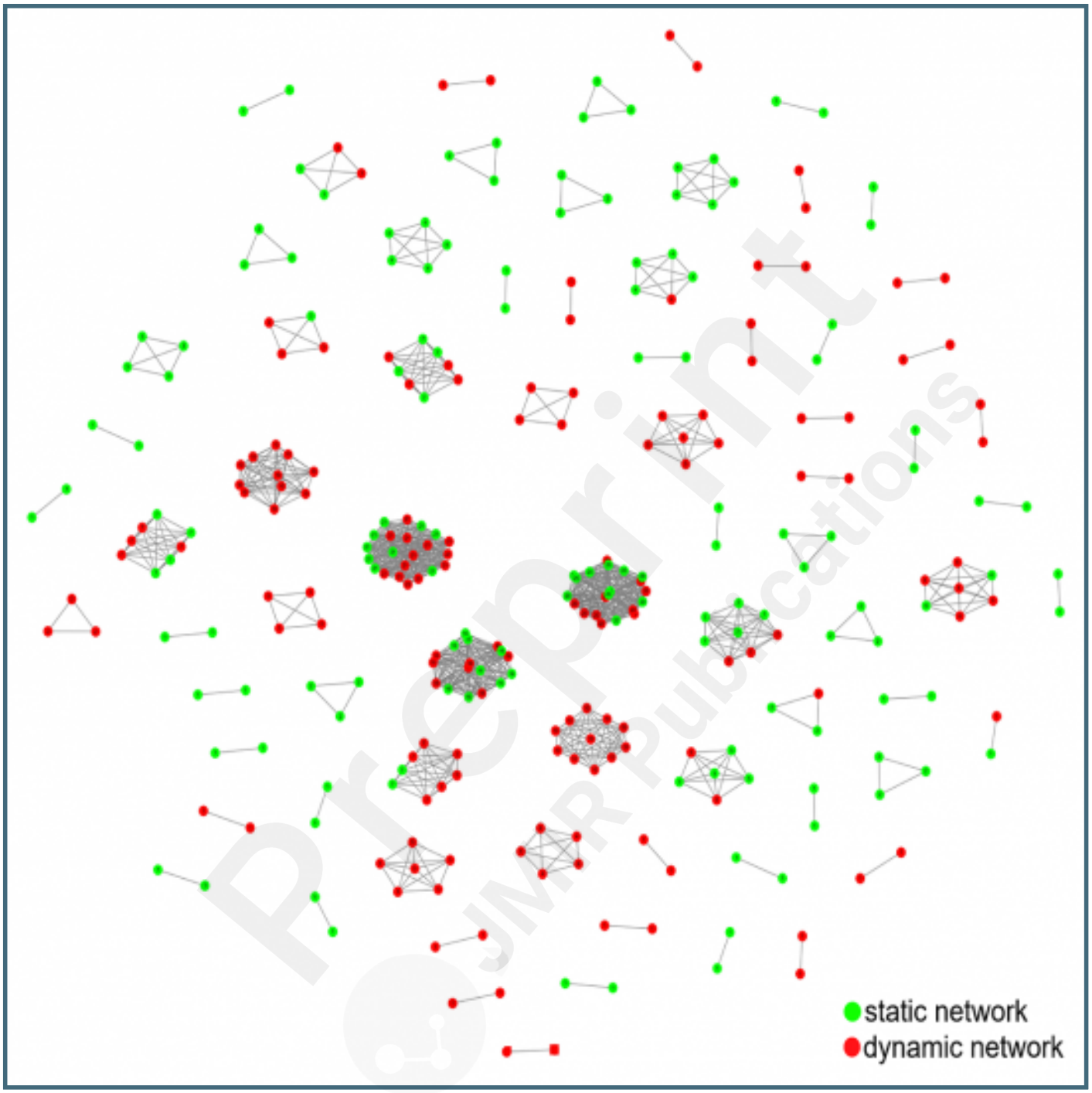
Supplementary Files

Figures

The dynamic changes in molecular networks recognition.



The baseline molecular network.



The observed endpoint molecular network.