

## Exploratory social network analysis and gene sequencing in people who inject drugs infected with hepatitis C virus

Y. WANG<sup>1</sup>, X. D. TAN<sup>1\*</sup>, C. ZHOU<sup>2</sup>, W. ZHOU<sup>3</sup>, J. S. PENG<sup>3</sup>, Y. S. REN<sup>4</sup>,  
Z. L. NI<sup>1</sup>, B. LIU<sup>1</sup>, F. YANG<sup>5</sup> AND X. D. GAO<sup>6</sup>

<sup>1</sup> *School of Public Health, Wuhan University, Wuhan, Hubei, China*

<sup>2</sup> *Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*

<sup>3</sup> *Wuhan Centers for Disease Control & Prevention, Wuhan, Hubei, China*

<sup>4</sup> *Institute of Basic Medical Sciences, Wuhan University, Wuhan, Hubei, China*

<sup>5</sup> *Departments of Nursing, Hubei University of Chinese Medicine, Wuhan, Hubei, China*

<sup>6</sup> *College of Health science and Nursing, Wuhan Polytechnic University, Wuhan, Hubei, China*

*Received 24 December 2015; Final revision 6 May 2016; Accepted 25 May 2016;  
first published online 13 July 2016*

### SUMMARY

Social networks facilitate the transmission of hepatitis C virus (HCV) in people who inject drugs (PWID). The aim of this study was to assess how certain network structural characteristics are related to HCV infections in PWID and to determine the most susceptible individuals for HCV transmission in a network of PWID. PWID ( $N = 80$ ) from central China were recruited from a previous follow-up case-control study. Demographic and behavioural information was obtained from a computerized database for each group. HCV RNA was extracted from blood specimens. Sequences were used to construct a phylogenetic tree and to determine genetic distances. Socio-metric social links were established between participants. Network measures were calculated using UCINET. Three HCV genotypes were identified, covering five subtypes. The density of the social networks for the whole sample ( $N = 80$ ), case group ( $n = 31$ ) and control group ( $n = 49$ ) was 0.038, 0.054 and 0.008, respectively. PWID infected with HCV were in frequent contact with others within their group. There were four pairs of nodes with genotypic distances of 0.000 that were identified and clustered in subtypes 6a and 1b; each subject pair was linked and found in one clique. Three of the five most active nodes were infected with HCV. These three nodes served as a bridge, contributing to the connection of other nodes. These findings identify susceptible individuals for HCV transmission in PWID based on their frequent contact with others in the network. These results provide data that could be used for modelling HCV transmission patterns and in public health policies.

**Key words:** Genetic diversity, hepatitis C virus infection, people who inject drugs, social network analysis.

### INTRODUCTION

Hepatitis C virus (HCV) is a contagious blood-borne disease that is considered a global public health threat

[1]. The World Health Organization (WHO) estimates that about 130–150 million people are infected with chronic hepatitis C globally and that about 35–50 million die from HCV-related diseases [2, 3]. There has been an increasing trend in the prevalence of HCV infection in East-Central Asia, with East Asia reported to have an anti-HCV antibody prevalence of 1.2% and a viraemic HCV prevalence of 0.7% [2, 3].

\* Author for correspondence: Dr X. D. Tan, School of Public Health, Wuhan University, Wuhan 430071, Hubei, China.  
(Email: xiaodongtan@yahoo.com)

Studies examining the genetic diversity of HCV have indicated that six genotypes and more than 80 sub-genotypes exist in the HCV-infected population [4].

People who inject drugs (PWID) are a high-risk population for HCV infection [5, 6]. The estimated prevalence of HCV in PWID in China is 61.4–70.0%, with the heavily endemic areas being the Hubei, Yunnan, Guangxi, Hunan and Xinxiang provinces. A previous study reported that China has the largest population of PWID worldwide, with about 1.33 million PWID [7]. Considering the large number of PWID in China, the challenge of controlling HCV transmission in PWID is great.

Epidemiological studies have recently been increasing their focus on the social networks of PWID [8]. Social network analysis (SNA) allows for the identification and targeting of groups of people with high-risk behaviours and the investigation of the mode of disease transmission in a scientific manner, based on traditional contact tracing using structured interviews and questionnaires [9]. The target groups can include highly active participants, high-risk behaviour groups and individuals serving as a ‘bridge’ connecting people. A number of previous studies have used the SNA approach to explore HIV transmission in different countries including China [10–12]. A study derived from a cohort of PWID confirmed that HCV phylogenetic clustering is associated with the social injecting network, highlighting the importance of the injecting network in HCV transmission [13]. Rolls *et al.* [14] created an empirical network of PWID and developed a detailed model of HCV transmission in this network, providing valuable insights for controlling high-risk groups. Additionally, researchers from Australia determined that HCV transmission clusters correlated with reported injecting relationships [15].

To our knowledge, limited information exists regarding SNA of PWID infected with HCV in China. No studies have used SNA and gene sequencing together to explore the transmission routes of HCV in Chinese patients under methadone maintenance therapy (MMT) who are infected with HCV. It was hypothesized that social network linkages would exist in PWID attending different MMT clinics in a single city (Wuhan, China), as a previous study had identified social network linkages between PWID attending different street drug markets in Melbourne, Australia (a city of slightly larger area than Wuhan) [13]. Using a cohort study performed in China in MMT users infected with HCV [16], this study aimed to: (1) explore the social networks in

PWID in central China; (2) reveal the HCV genotypes of the target groups; and (3) assess how certain network structural characteristics such as node degree, network density and betweenness centrality are related to HCV infection in PWID. This study will provide useful insights into HCV transmission in PWID, which potentially could be used by public health services to help control HCV epidemics.

## METHODS

### Study design and participants

This study comprised of three parts: a baseline survey, HCV viral sequencing of a case group (HCV seroconversion), and SNA.

The clinical records of all patients visiting the 20 MMT clinics in Wuhan (China) were reviewed. The distances of the other MMT clinics from the most central clinic ranged from 1.1 km to 23.6 km (mean  $10.3 \pm 7.5$  km). In 2013, Wuhan city had a population of ~10 220 000, and 0.27% of the population (~27 950) were PWID (according to the Public Security Bureau of Wuhan City). Around 14 000 PWID attended MMT clinics in 2013, and the number of people taking drugs each day was estimated to be 4000.

Of the 16 085 patients visiting between May 2006 and June 2011, 12 755 (79.3%) received a HCV test at MMT entry. Of the 3558 MMT patients who tested HCV-negative at MMT entry, 47.8% (1702/3558) agreed to participate in the prospective follow-up study. To ensure that the investigated HCV antibody seroconversion was the result of transmission during the MMT programme, only participants who had two HCV-negative blood specimens (at MMT entry and at least 3 months later) and who completed at least one follow-up interview were included in the final study sample; therefore, 502 (29.5%) of 1702 eligible patients were excluded. Consequently, 1200 participants were included in the prospective follow-up sample. Their characteristics have been published recently [16]. Of the 1200 participants, seroconversion was observed in 555 between May 2006 and June 2011. By the beginning of the present study (April 2014), 32.07% (178/555) of patients who showed seroconversion were still attending their MMT clinic and 31 participants agreed to supply a blood sample. Consequently, the case group included these 31 participants. Of the 645 non-seroconversion participants, only 128 continued to attend their MMT clinic, of which 49 agreed to participate in the study. Ultimately, the case

group included 31 participants and the control group included 49 participants (see Supplementary Fig. S1).

### Epidemiological investigation

The socio-demographic data (including gender, age, education, marital status, employment status, living status) and self-reported behavioural data (such as illicit drug use history, needle/equipment sharing history, drug rehabilitation, treatment history) were assessed using a questionnaire. This survey was conducted during 1–15 April 2014, with the cooperation of MMT doctors (face-to-face method). All 80 patients completed the questionnaire, as required for inclusion in the final analysis.

### Specimen collection and HCV genotyping

Fasting whole blood specimens (10 ml) were collected in K<sub>3</sub> EDTA tubes, immediately stored at 4 °C, and transferred to the laboratory of Wuhan CDC for processing on the same day. Plasma samples were stored in aliquots at –20 °C until analysed.

All specimens were tested for anti-HCV antibodies using the third-generation enzyme-linked immunosorbent assay (EIA-3) system (Kehua Biotechnology Inc., China), according to the manufacturer's instructions. Specimens found to be initially positive by EIA-3 were retested in duplicate. Repeat reactive specimens were assumed to be seropositive. Specimens shown to be reactive after repeat testing were used to conduct further HCV RNA analyses.

HCV RNA was extracted using the care HCV RT–PCR assay v. 2 kit (Qiagen, Germany), according to the manufacturer's instructions. The amount of HCV RNA was measured by UV spectrophotometry. RNA (1 µg) was used for real-time polymerase chain reaction (PCR) analysis using the 7300 real-time PCR system. The isolated RNA was used to synthesize cDNA using random primers. RNA (1 µg), the random primers and ddH<sub>2</sub>O were mixed in a reaction volume of 12 µl and incubated at 65 °C for 5 min in the PCR instrument (Eppendorf, Germany). The mixture was immediately placed on ice and dNTPs and reverse transcriptase (RT) were added. After incubation at 30 °C for 10 min and 42 °C for 20 min, cDNA was obtained. The cDNA was then amplified in a reaction volume of 20 µl using RT–PCR and the SYBR Green reaction system (Applied Biosystems, USA) using two sets of primers (Ac2 and Sc2; S7 and A7) targeting the 5'-UTR and core (335 bp) regions [17]. The sense sequencing primer was designed at nucleotide positions –12 to 8

of the 5'-UTR and core regions. The antisense primer was located at nucleotide positions 319–343 of the core region. The reaction conditions were pre-denaturation at 95 °C for 1 min, denaturation at 95 °C for 15 s, renaturation at 56 °C for 20 s, and extension at 72 °C for 45 s, for 40 cycles. The fluorescence signal was collected during the extension phase. The RT–PCR amplified DNA products were analysed by agarose gel electrophoresis (2%) containing 0.5 µg/ml ethidium bromide. The DNA PCR products were purified from the agarose gel and sequenced by a sequencing company (Sangon Biotech Co. Ltd, China).

A phylogenetic tree was constructed using the 5'-UTR/core region sequences of 31 HCV strains aligned to randomly selected published HCV sequences that were retrieved from the GenBank database. The statistical reliability of this tree was determined using a bootstrap analysis that provides a percentage relatedness score for each cluster predicted by the tree. A cluster with a score of >95% was considered as significantly related.

The resulting sequences were aligned using CLUSTAL X 2.1 (University College Dublin, Ireland). Phylogenetic trees were constructed with MEGA 5.05 software (USA) using the maximum-likelihood method, and the reliability of the clustering was assessed using the bootstrap test (1000 replicates). Genetic distances were calculated using the Poisson model with MEGA 5.05. Reference sequences were used to construct the phylogenetic tree on the basis of 15 HCV isolates selected from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). The 15 HCV isolates were D11443 (Japan), D49374 (Japan), GU814263 (Italy), NC009824 (Japan), NC009827 (Japan), EU246930 (Vietnam), FJ410172 (USA), NC004102, D10750, FJ390398, NC009826, NC009823 (USA), D10922, GU814265 (Egypt) and NC009825 (Middle East).

The obtained sequences were submitted to GenBank and given the following accession numbers: JX944384–JX944467 and KC348398–KC348438.

### SNA investigation

Social network data were obtained using a social network questionnaire (SNQ) following the SNA design principle, which stipulates that the questionnaire must include a roster (i.e. a list of the names, including given names and commonly used nicknames, of all 80 participants), anchored choice (e.g. 'face-to-face contact' and 'have never met') and rating (a scale between the two anchored choices to indicate the

frequency of contact) [18]. Participants were asked to examine the roster and indicate which participants they had contacted frequently in the past 3 months, which participants they had a close relationship with, and which participants they had never known (excluding themselves). The contact dimension was divided into two levels: regular contact (defined as ‘face-to-face contact >6 times in the past 3 months’ or ‘face-to-face contact <6 times in the past 3 months but have close relationship with the other participant’) and minimal contact (‘face-to-face contact <6 times in the past 3 months and do not have close relationship with the other participant’ or ‘have never met’).

To ensure the quality of the research, each participant was shown only the names/aliases of the other participants in the roster, in order to avoid subjective selections unrelated to the study objective. Other information (such as gender, MMT clinics attended, home address and contact information) was available only to the researchers. The collection of SNA data was conducted during 15–30 April 2014. SNA data were collected for all 80 patients.

### Statistical analysis

The epidemiological data were entered into EpiData software (www.EpiData.dk). The statistical analyses were performed using SPSS v. 17.0 (SPSS Inc., USA). Continuous data are expressed as the mean  $\pm$  standard deviation (s.d.) and were analysed using Student’s *t* test.  $P < 0.05$  was considered statistically significant.

All network analyses were conducted using UCINET software [19]. The networks were constructed with each study participant represented by a node (or vertex), and with frequent contact or familiarity between two participants represented by an edge between two nodes. The networks were described on the basis of the following properties: network density, node degree, number of cliques, network centralization index, number of components, average distance and betweenness centrality. The network density (range 0–1) describes the general level of cohesion in the network, and was calculated as the number of edges in the network (i.e. direct links between nodes representing frequent contact or familiarity) divided by the maximum number of edges possible [which is given by  $n(n-1)/2$ , where  $n$  is the total number of nodes]. The node degree for each node was defined as the number of other nodes that connected to that node (i.e. the number of incident edges). A clique in a social network describes a subset of participants

who are more closely tied to one another than to other members of the network. A clique was defined as a subset of participants such that every two vertices were adjacent (i.e. all nodes within the clique were directly connected by edges to all other nodes within the clique); the minimum size of a clique was set to three nodes. Network centralization index (range 0–100%) measures the degree of dispersion of all the node degree centrality scores in a network from the maximum degree centrality score obtained in the network, where the degree centrality score of a node is an indicator of how connected that node is to all other nodes. A high network centralization index indicates that a network contains a small number of central nodes that are highly connected to most other nodes, and a larger number of less central nodes that tend to connect only to the small number of central nodes. The average distance was calculated as the average length of all the shortest paths (geodesics) between all pairs of vertices. The components of the network were defined as subgraphs (each of which contains a group of connected nodes) that were disconnected from each other [20]. The betweenness centrality measures the extent to which a node lies on paths between other nodes, and thus measures the ability of a participant (represented by that node) to potentially influence the interaction between other participants in the network. The betweenness centrality of a particular node is calculated as the number of shortest paths from all nodes to all others that pass through that particular node. In the context of this study, a node with high betweenness centrality would be considered to have a large influence on HCV transmission through the network, under the assumption that HCV transmission follows the shortest paths.

### Ethical approval

Ethical approval was given by Institutional Review Boards (IRB) of the Wuhan Centers for Disease Control and Prevention (CDC) (reference no. WHCDCIRB-K-2014001). All participants signed an informed consent form that included a detailed description of the potential benefits of the study. The patients were aware and understood that for SNA their identity would be provided to the other participants.

## RESULTS

### Characteristics of the participants

The demographic and socioeconomic characteristics of the participants are summarized in Table 1. Of

Table 1. *Characteristics of the participants*

Category	Cases ( <i>n</i> = 31)	Controls ( <i>n</i> = 49)	<i>P</i>
Age, years, <i>n</i> (%)			0.177
<30	1 (3)	3 (6)	
30–49	28 (91)	36 (73)	
≥ 50	2 (6)	10 (21)	
Sex, <i>n</i> (%)			0.116
Male	22 (71)	33 (67)	
Female	9 (29)	16 (33)	
Education, <i>n</i> (%)			0.224
Junior high school or lower	20 (65)	40 (82)	
Senior high school	10 (32)	8 (16)	
Junior college or higher	1 (3)	1 (2)	
Marital status, <i>n</i> (%)			0.294
Married or has regular partner	15 (48)	29 (59)	
Divorced/widowed	8 (26)	6 (12)	
Single	8 (26)	14 (29)	
Employment status			0.691
Employed	7 (23)	13 (27)	
Unemployed	24 (77)	36 (73)	
Living status			0.199
With family	28 (90)	38 (78)	
With friends	0 (0)	4 (8)	
Alone	3 (10)	7 (14)	
Drug administration method*			<0.001
Injected	26 (84)	21 (43)	
Inhaled	5 (16)	28 (57)	
Drug injection in the past 30 days*			<0.001
Yes	22 (71)	23 (47)	
No	9 (29)	26 (53)	

\* *P* < 0.05.

the 31 case participants, 71% were male and 91% were in the 30–49 years age group; 67% of the controls were male and 73% of the controls were in the 30–49 years age group. All participants reported low levels of education and did not have steady or continuous employment. Within the case group, 48% were married or had a regular partner, and most (90%) lived with their family. Over half (59%) the control group was married or had a regular partner; 78% lived with their family. A comparison between the case and control groups regarding drug injection within the past 30 days revealed significant differences (71% vs. 47%, respectively, *P* < 0.05).

### HCV genotypic distribution

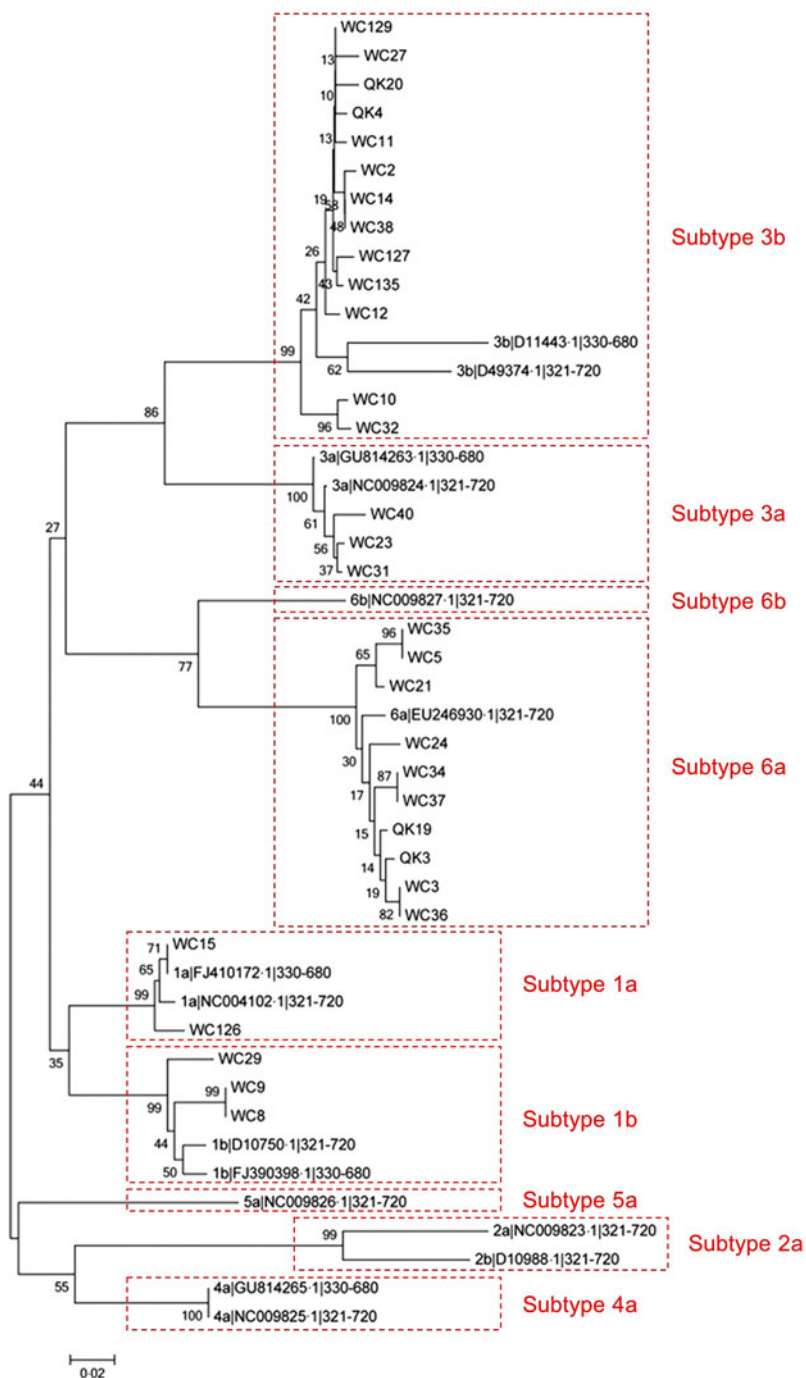
The genotypes of 31 HCV sequences were determined by phylogenetic analysis against a panel of HCV

reference sequences. In total, three HCV genotypes covering five different subtypes were identified, as shown in Figure 1. The largest group, comprising 13 sequences on the 5'-UTR/core tree, clustered with the subtype 3b reference sequence D11443. The second largest group, comprising 10 sequences on the 5'-UTR/core tree, clustered with the subtype 6a reference sequence. The other groups clustered with subtypes 3a, 1b and 1a. The proportions for each genotype were: 51.6% (16/31) for genotype 3 (three with subtype 3a and 13 with subtype 3b), 32.3% (10/31) for genotype 6 (all with subtype 6a), and 16.1% (5/31) for genotype 1 (two with subtype 1a and three with subtype 1b).

### Characteristics of the social networks

The overall descriptive statistics for the two groups are presented in Table 2. The densities of the social networks for the whole sample (*N* = 80), case group (*n* = 31) and control group (*n* = 49) were 0.038, 0.054 and 0.008, respectively. The mean node degree was  $3.025 \pm 2.525$  (range 0–17) for the whole sample,  $3.000 \pm 2.680$  (range 0–14) for the case group and  $0.776 \pm 0.789$  (range 0–3) for the control group (*P* < 0.05, Kruskal–Wallis test, Table 2). The most notable result, as shown in Table 2, was that 49 cliques between the 49 ego networks in the control group were identified. This number is 1.69 times greater than that in the case group, which suggests a structural difference between the two groups. Similarly, the network centralization index of the overall population (28.05%) was higher than that of the control group (0.04%). The network for all participants contained eight components, the largest of which contained 72 nodes.

Visualizing social networks is another way of determining the structural differences between them (Fig. 2). It was evident that the larger the node, the higher the value of betweenness for the participant. The network presented with five central points that were active in the network: participant 1 (936.26), participant 2 (411.26), participant 13 (312.97), participant 66 (456.74) and participant 70 (491.45). Additionally, clustering was prevalent for the case group, and it was particularly noticeable that PWID infected with HCV mainly maintained contacts with participants within the same group. This is clear in Figure 2, from which it is evident that participants in the case group had frequent contact with other members of their group, whereas contact was much less frequent between members of the control group.



**Fig. 1.** The genotypes of 31 HCV sequences were determined by phylogenetic analysis against a panel of HCV reference sequences. In total, three HCV genotypes covering five different subtypes were identified. Thirteen sequences (QK4, QK20, WC2, WC10, WC11, WC12, WC14, WC27, WC32, WC38, WC127, WC129, WC135) clustered with the subtype 3b reference sequence, ten sequences (QK3, QK19, WC3, WC5, WC21, WC24, WC34, WC35, WC36, WC37) clustered with the subtype 6a reference sequence, three sequences (WC23, WC31, WC40) clustered with the subtype 3a reference sequence, three sequences (WC8, WC9, WC29) clustered with the subtype 1b reference sequence, and two sequences (WC15, WC126) clustered with the subtype 1a reference sequence.

The distribution of the number of participants from each MMT clinic is shown in Supplementary Table S1. Figure 3 presents an alternative visualization to emphasize the within-clinic structure of the social network, with

lines between points representing familiarity/frequent contact between participants in the same MMT clinic. The results depicted in Figure 3 suggest that participants within the same MMT clinic have more familiarity or

Table 2. Social network analysis indicators in the two groups

Indicator	Group		
	All (N = 80)	Cases (n = 31)	Controls (n = 49)
Network density	0.038	0.054	0.008
Node degree (mean ± s.d.)	3.025 ± 2.525	3.000 ± 2.680	0.776 ± 0.789
Node degree (range)	0–17	0–14	0–3
Number of cliques	78	29	49
Network centralization index	28.05%	2.01%	0.04%
Components	8	18	43
Average distance	3.588 ± 1.337	1.858 ± 1.068	1.136 ± 0.343
Betweenness centrality	0.809	0.114	0.009

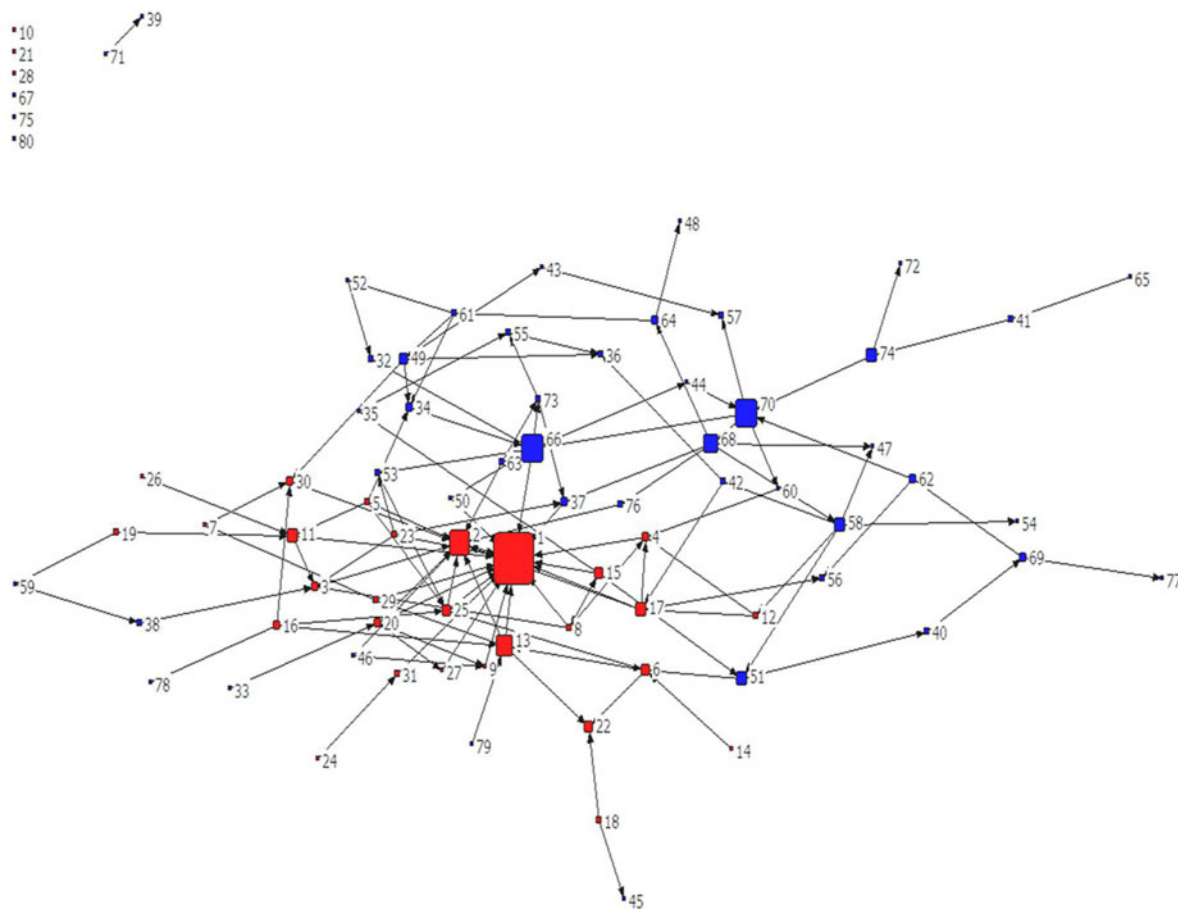


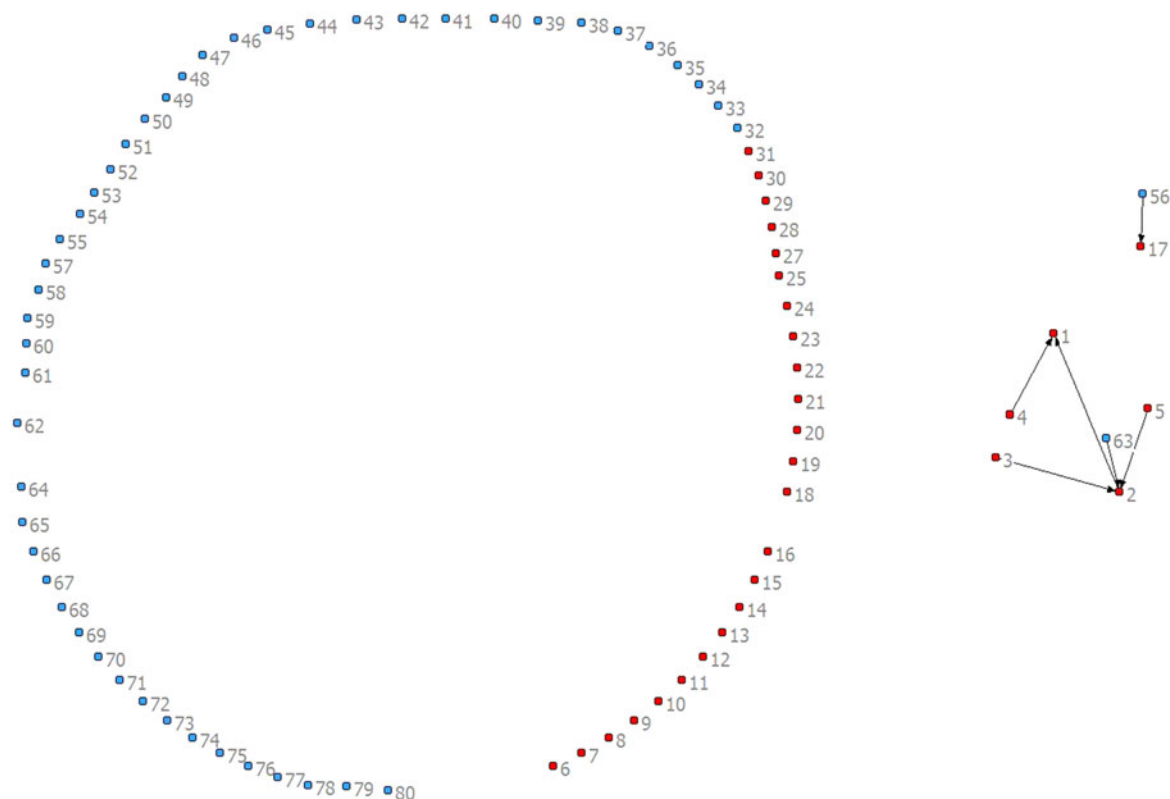
Fig. 2. Socio-metric graph of the study population. Large nodes depict higher betweenness centrality.

frequent contact, although it was not possible to determine whether this contact between participants within a clinic was initiated before or after registration at the clinic.

**Construction of the susceptible population and participants with a high-probability of transmission based on gene sequencing and SNA**

Participants 1–31 belonged to the case group, while participants 32–80 belonged to the control group.

Based on the interesting findings regarding the characteristics of the social networks described above, we attempted to determine the potentially susceptible population and the participants with a high probability of transmission. First, four pairs of nodes with a genotypic distance of 0.000 were identified, these clustered in subtypes 6a and 1b. The four pairs of nodes were participants 24 (6a) and 5 (6a), participants 7 (1b) and 6 (1b), participants 26 (6a) and 2 (6a), and



**Fig. 3.** Visualization of the within-clinic structure of the social network. Red symbols identify individuals who were HCV-positive. Lines between participants represent familiarity/frequent contact between participants in the same MMT clinic.

participants 25 (6a) and 3 (6a). Further analysis of the social network revealed that each pair of subjects was linked and existed in one clique. Second, three of the five most active nodes were infected with HCV; they were determined to be participants 1 (6a), 2 (6a) and 13 (3b). The above three nodes served as a bridge, contributing to the connection of the other nodes that were identified as high-probability transmission participants with the highest betweenness centrality. Correspondingly, the negative individuals who were in close connection with the above three participants and existed in the same clique were regarded as the principal susceptible population.

## DISCUSSION

Social networks facilitate the transmission of HCV in PWID, but no studies have used SNA and gene sequencing to explore the transmission of HCV in MMT patients infected with HCV in China. Therefore, this study aimed to assess how certain network structural characteristics are related to HCV infection in PWID and to determine which individual

PWID are susceptible to HCV transmission. Three HCV genotypes were identified, covering five subtypes. The densities of the social networks for the whole sample, case group and control group were 0.038, 0.054 and 0.008. PWID infected with HCV were in frequent contact with others within their group. There were four pairs of nodes with genotypic distances of 0.000 that were identified and clustered in subtypes 6a and 1b; each pair of subjects was linked and found in one clique. Three of the five most active nodes were infected with HCV. These three nodes served as a bridge, contributing to the connection of the other nodes.

This study showed that there are differences in injection risk behaviours between HCV-infected PWID and those not infected. Gene sequencing further showed the genetic types and distribution in HCV-positive participants. Using SNA we identified highly central individuals. The combination of SNA and gene sequencing was used in order to circumvent the recognized limitations of traditional epidemiological investigations [21]. SNA can improve case-finding in vulnerable populations using specific questionnaires that identify the



activity density, activity diameter, cliques and high-risk behaviours [14, 18].

In this exploratory study, we first determined the epidemiological characteristics of PWID in central China. Most of the participants were middle-aged men aged 30–49 years. The study population was similar to that of a previous investigation [22]. In addition, the case group had a low level of education and was mostly unemployed. In terms of injection risk behaviours, HCV-infected PWID generally reported a high rate of drug injection in the past 30 days (71%), similar to the findings of another report (68.8%) [23]. Despite the substantial decline in the incidence of acute HCV infection observed over a 25-year period (1982–2006) [2], being a PWID remains a relevant risk factor for acquiring a new HCV infection. Nonetheless, in the context of the high seroconversion rate in PWID, the injecting risk behaviours were extensive in this population [24]. Thus, it is possible that the control participants in this study who did not seroconvert were interacting with fewer PWID since they were injecting less frequently.

In this study, the HCV genes found in PWID were sequenced and genotyped. The results showed that 3b (41.94%, 13/31) and 6a (32.26%, 10/31) were the predominant subtypes, which is similar to the genotypic distribution in the PWID population in southern Chinese provinces, where the main HCV subtypes included 6a (38%) and 3b (37%) [25]. However, the genotypic distribution of HCV in patients receiving MMT in this study was not entirely consistent with previous results in PWID not receiving MMT, which showed the predominant HCV subtypes to be 6a (50%) and 3b (32.2%) [26]. This may be due, in part, to the small sample size in this study. Sequence analysis provides a powerful tool that may be used to investigate the spread of HCV within a community. Interestingly, four pairs of PWID presented genetic distances of 0; their HCV genes showed a 99% similarity. We found that the proportion of patients in transmission clusters with a genotypic distance of 0.000 accounted for up to 25.8% (8/31) of the participants.

The limitations of traditional epidemiological research preclude the exploration of whether PWID with genetically similar HCV genes knew each other and were in frequent contact, and of how this seroconverted population actually transmits HCV. However, SNA affords the ability to determine answers to the above questions [15]. The use of simulated contact networks generated from an empirically grounded

network model to obtain data regarding HCV transmission and treatment represents a recently developed statistical approach [27]. In this study, SNA of HCV infections suggests that HCV may be more easily spread in small groups of PWID, where one individual most likely acts as the main source of the spread of HCV. The case group network was much closer, numerically denser and more centralized than the control group network. PWID in the case group were more likely to resemble a small world, showing clustering within a few cliques. Highly central individuals have been targeted with prevention campaigns to become peer leaders in preventative interventions [28]. The present study also highlights the potentially dual importance of people with high betweenness centrality. First, SNA can identify the HCV-infected PWID that have the highest probability of transmitting HCV to other participants, and can recognize the population at highest risk of transmission. In a sense, the central points can be referred to as potential ‘super-spreaders’. Second, people with high betweenness centrality could be used as effective peer educators in preventative interventions in order to reach various high-risk populations. Therefore, the use of questionnaires to identify contacts and the application of SNA to these data could potentially detect HCV-infected PWID with high betweenness centrality as well as identifying their network of contacts (i.e. other PWID with whom they interact). This would facilitate the targeting of interventions to high-risk populations in order to reduce HCV transmission.

The present study has a slightly smaller sample size than that typically used in the field of epidemiology. This was due to the recruitment of a susceptible group and represents one of the limitations of the study. Some of the participants from the previous follow-up study [16] appear to have been lost to follow-up for a variety of reasons. Since drug trading is illegal in China, some PWID were in prison or in compulsory detoxification. Many PWID refused to participate because of the non-confidential nature of the SNA. Because of these limitations, the present study is merely exploratory. Another limitation is that SNA does not provide direct information about who injected with whom. Indeed, because drug injection is illegal, the response rate for this specific type of enquiry may be very low. However, since each injecting drug use territory was relatively small, the participants were likely to share needles only with people they knew. Furthermore, although the results suggest that participants from the same MMT clinic had

closer contact, it was not possible to determine whether the familiarity started before or after registration at the clinic, and hence what role the clinic played in the establishment of the contact. Another limitation is that the socio-demographic, self-reported behavioural and SNA data were collected in April 2014, whereas blood tests for seroconversion were only performed between May 2006 and June 2011; thus, it is possible that some of the participants classified as not showing seroconversion (by June 2011) may in fact have shown seroconversion by the time the questionnaires were administered (April 2014). In addition, since the sample may have had different characteristics from the larger patient population from which it was taken, further work is needed to establish the generalizability of these findings. Additional, larger-scale studies are necessary to address these issues.

## CONCLUSION

This is the first study of HCV-infected PWID in China using SNA and gene sequencing. It provides reliable and scientific information that may be useful for controlling HCV in PWID through social-network visualization and gene distance. It also motivates new research that will be able to use this innovative method of combining spatial contact with gene linkage. Due to the high prevalence of HCV in PWID compared to HIV, this study clearly describes the susceptible population and participants with high probability of transmission. These results might be of great importance to public health agencies. The findings also highlight the significance of future research on the effectiveness of interventions based on social networks in PWID, and on the importance of targeting gatekeepers and ‘bridges’.

## SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0950268816001333>.

## ACKNOWLEDGEMENTS

We thank Professor R. S. Schottenfeld and Professor M. C. Chawarski from Yale University School of Medicine for providing valuable suggestions for this programme.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. **Bruggmann P, et al.** Historical epidemiology of hepatitis C virus (HCV) in selected countries. *Journal of Viral Hepatitis* 2014; **21** (Suppl. 1): 5–33.
2. **Walsh N, et al.** WHO guidance on the prevention of viral hepatitis B and C among people who inject drugs. *International Journal of Drug Policy* 2014; **25**: 363–371.
3. **Gower E, et al.** Global epidemiology and genotype distribution of the hepatitis C virus infection. *Journal of Hepatology* 2014; **61**: S45–57.
4. **Dong ZX, et al.** Distribution of hepatitis C virus genotypes in Chinese patients with chronic hepatitis C: correlation with patients’ characteristics and clinical parameters. *Journal of Digestive Diseases* 2012; **13**: 564–570.
5. **de Vos AS, Prins M, Kretzschmar ME.** Hepatitis C virus treatment as prevention among injecting drug users: who should we cure first? *Addiction* 2015; **110**: 975–983.
6. **Hagan H, Pouget ER, Des Jarlais DC.** A systematic review and meta-analysis of interventions to prevent hepatitis C virus infection in people who inject drugs. *Journal of Infectious Diseases* 2011; **204**: 74–83.
7. **Zhuang X, et al.** Risk factors associated with HIV/HCV infection among entrants in methadone maintenance treatment clinics in China: a systematic review and meta-analysis. *Drug and Alcohol Dependence* 2012; **126**: 286–295.
8. **Puerta JL.** Network analysis and medicine: a new perspective [in Spanish]. *Medicina Clinica (Barcelona)* 2013; **140**: 273–277.
9. **Gardy JL, et al.** Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine* 2011; **364**: 730–739.
10. **Gyarmathy VA, et al.** Social network structure and HIV infection among injecting drug users in Lithuania: gatekeepers as bridges of infection. *AIDS and Behavior* 2014; **18**: 505–510.
11. **Li J, et al.** Sexual transmissibility of HIV among opiate users with concurrent sexual partnerships: an egocentric network study in Yunnan, China. *Addiction* 2011; **106**: 1780–1787; discussion 1788–1789.
12. **Rice E, et al.** Mobilizing homeless youth for HIV prevention: a social network analysis of the acceptability of a face-to-face and online social networking intervention. *Health Education Research* 2012; **27**: 226–236.
13. **Sacks-Davis R, et al.** Hepatitis C virus phylogenetic clustering is associated with the social-injecting network in a cohort of people who inject drugs. *PLoS ONE* 2012; **7**: e47335.
14. **Rolls DA, et al.** Modelling hepatitis C transmission over a social network of injecting drug users. *Journal of Theoretical Biology* 2012; **297**: 73–87.
15. **Hellard M, et al.** Hepatitis C transmission and treatment as prevention – the role of the injecting network. *International Journal of Drug Policy* 2015; **26**: 958–962.

16. **Zhou W, et al.** Hepatitis C seroconversion in methadone maintenance treatment programs in Wuhan, China. *Addiction* 2015; **110**: 796–802.
17. **Ohno O, et al.** New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. *Journal of Clinical Microbiology* 1997; **35**: 201–207.
18. **Scott J.** *Social Network Analysis: A Handbook*, 2nd edn. London: Sage Publications, 2000.
19. **Borgatti SP, Everett MG, Freeman LC.** *UCINET IV Test Version*. Columbia: Analytic Technologies, 1991.
20. **Freeman LC, Borgatti SP, White DR.** Centrality in valued graphs: a measure of betweenness based on network flow. *Social Networks* 1991; **13**: 141–154.
21. **El-Sayed AM, et al.** Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives and Innovations* 2012; **9**: 1.
22. **Hahn JA, et al.** Hepatitis C virus risk behaviors within the partnerships of young injecting drug users. *Addiction* 2010; **105**: 1254–1264.
23. **Clatts MC, et al.** Prevalence and incidence of HCV infection among Vietnam heroin users with recent onset of injection. *Journal of Urban Health* 2010; **87**: 278–291.
24. **Tsui JI, et al.** Risk behaviors after hepatitis C virus seroconversion in young injection drug users in San Francisco. *Drug and Alcohol Dependence* 2009; **105**: 160–163.
25. **Garten RJ, et al.** Coinfection with HIV and hepatitis C virus among injection drug users in southern China. *Clinical Infectious Diseases* 2005; **41** (Suppl. 1): S18–24.
26. **Peng JS, et al.** Genetic variation of hepatitis C virus in a cohort of injection heroin users in Wuhan, China. *Virus Research* 2008; **135**: 191–196.
27. **Rolls DA, et al.** Hepatitis C transmission and treatment in contact networks of people who inject drugs. *PLoS ONE* 2013; **8**: e78286.
28. **Latkin C, et al.** The dynamic relationship between social norms and behaviors: the results of an HIV prevention network intervention for injection drug users. *Addiction* 2013; **108**: 934–943.