Genetic variability and its effect on genetic network construction

Tsz Ho Kwan[1], Shui Shan Lee[2]*

[1] Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Shatin, Hong Kong
[2] Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Shatin, Hong Kong

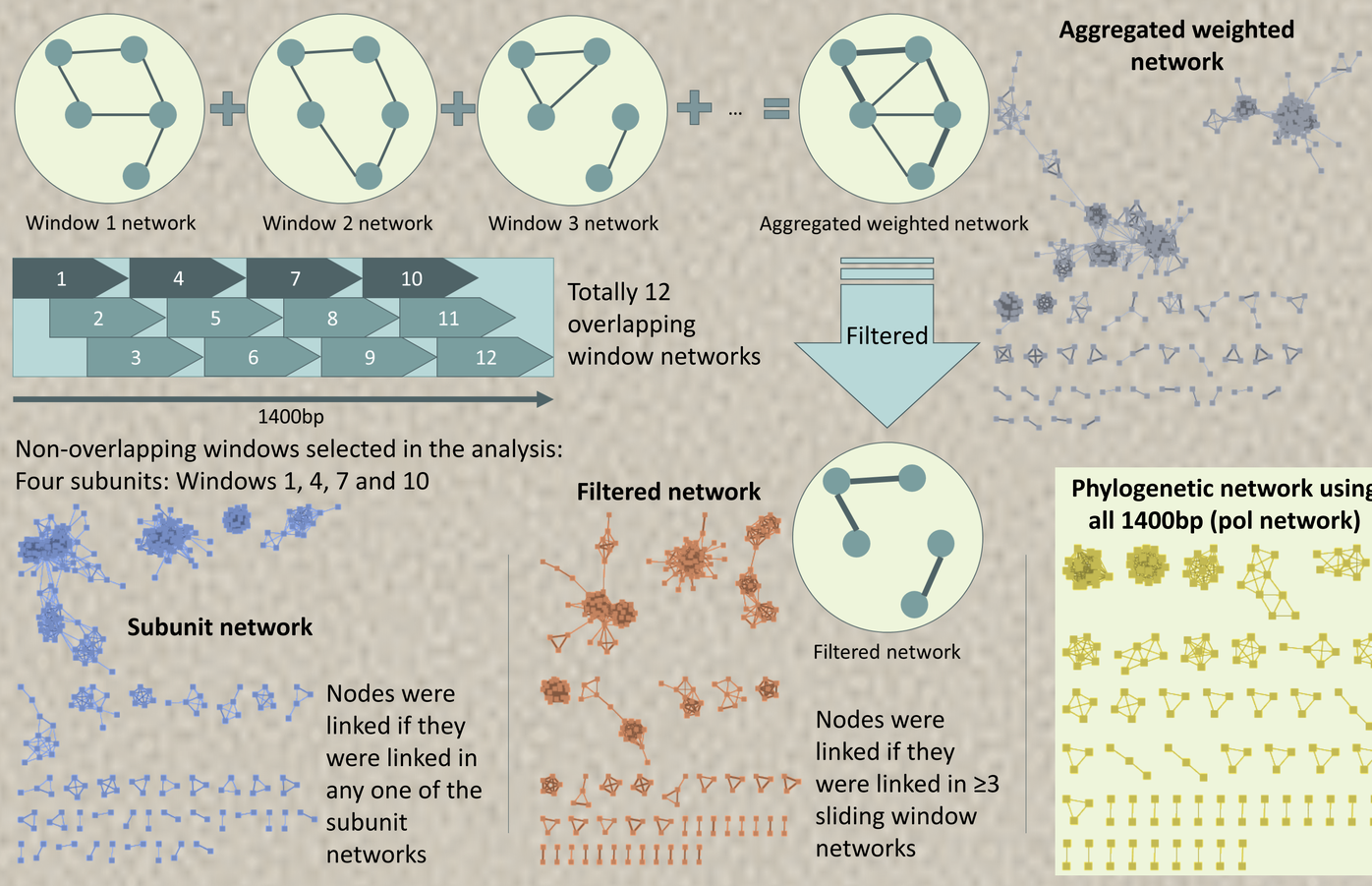
Background

Recombination of different HIV subtypes is not uncommon, especially in area where multiple subtypes are in circulation. While understanding molecular epidemiology of recombinants is important, phylogeny-based methods do not always apply, therefore novel method is warranted.

Methods

HIV partial pol sequences collected from newly diagnosed patients from all four HIV specialist clinics in Hong Kong were studied. With a sliding window of 300bp shifted by 100bp, aligned sequences were subsetted and TN93 pairwise distance was calculated for each window. Distance matrices were calculated by splitting the sequence into four subunits and entire sequence. Pairs with <1.5% distance were considered genetically linked. Pol network from whole sequence was separately differenced by networks of linked pairs for at least 3 sliding windows and for at least 1 subunit to compare edge distribution and its variation across subtypes.

The study was approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee (Ref. No.: CREC2015.232). Institutional approval was also obtained from the collaborating sites where patients were recruited: Queen Elizabeth Hospital, Princess Margaret Hospital, Prince of Wales Hospital, and Integrated Treatment Centre of Department of Health.



Results

- Some 79% of 438 sequences analysed were connected with another for at least one of twelve 300bp-window forming 2397 edges, of which 1086 (45%) were links of at least 3 windows
- There were 1717 links in subunit network, 59% of which were one-subunit links, and the second subunit (i.e. window 4) had the most linkages
- All edges in pol network were a subset of networks generated by the two comparators
- Differencing pol network by the two, links associated with recombinants accounted for 21% and 27%, respectively
- Among intra-subtype edges which accounted for over 70% in both differenced networks, CRF07_BC contributed to the most (42%;51%), followed by B (35%;30%), CRF01_AE (16%;13%) and A (1%;<1%)

Networks constructed by	# nodes	# edges
Whole sequence (1400bp)	189	528
Having links in ≥1 window network	344	2397
Having links in ≥3 window networks	279	1086
Having links in ≥1 non-overlapping subunits	324	1717

Table 1. Summary of networks constructed by different methods

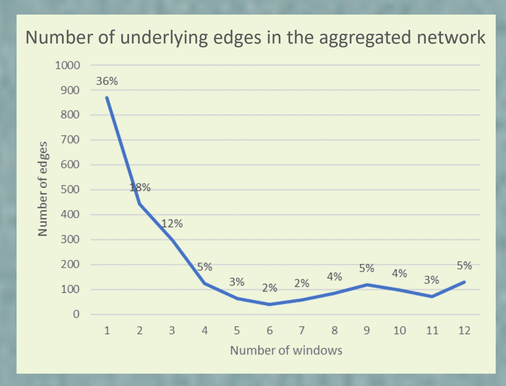


Fig 1. Number of edges at least linked in n window networks

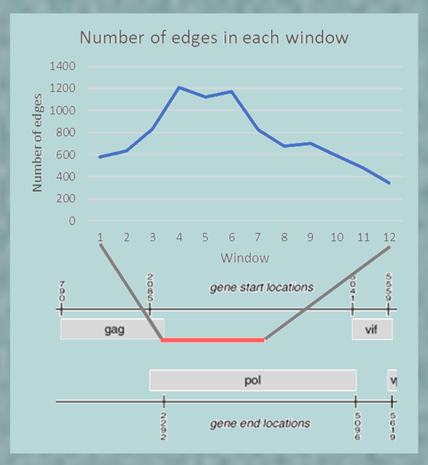


Fig 2. Number of edges in each sliding- window network

Difference of column by row (# nodes; # edges)	pol network	Filtered network	Subunit network
pol network		0; 0	0; 0
Filtered network	204; 558		30; 20
Subunit network	276; 1189	214; 651	

Table 2. Number of nodes and edges of the differenced network

Filtered – pol network	Recombinant	CRF01_AE	B	CRF07_BC	A
Recombinant	21				
CRF01_AE	24	92			
B	54	0	154		
CRF07_BC	20	0	0	183	
A	0	7	0	0	3

Table 3. Distribution of node subtypes of edges present in the filtered network but not in the pol network

Table 4. Distribution of node subtypes of edges present in the subunit network but not in the pol network

Subunit – pol network	Recombinant	CRF01_AE	B	CRF07_BC	A
Recombinant	50				
CRF01_AE	42	154			
B	94	1	259		
CRF07_BC	134	0	1	439	
A	0	12	0	0	3

Conclusion

Differential genetic variation was observed across pol gene. Highly varied region averaged out the similarity in other regions, particularly in subtype CRF07_BC. The high proportion of intra-subtype links not identified by whole gene suggested possible intra-subtype recombinations. Slicing sequences into windows could identify more genetically linked pairs yet its impact requires further exploration.

Funding

Research Grants Council of Hong Kong (14103315) and AIDS Trust Fund (MSS243R)

Acknowledgements

Technical support from Li Ka Shing Institute of Health Sciences is acknowledged

*Presented at **Infection 2019***
16th Annual Scientific Meeting of Stanley Ho Centre for Emerging Infectious Diseases
10 – 12 June 2019, Hong Kong