Electronic Supplementary Material

Large discrepancy between the two-way rNHT distances in Hemagglutinin-inhibition assay

Yousong Peng[1], Dayan Wang[2], Yuelong Shu[2✉], Taijiao Jiang[3,4✉]

1. College of Biology, Hunan University, Changsha 410082, China
2. National Institute for Viral Disease Control and Prevention, China CDC, Beijing 102206, China
3. Center of System Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China
4. Suzhou Institute of Systems Medicine, Suzhou 215123, China

Table S1. The HI dataset used in the study including the pairwise two-way rNHT distances between viruses, the HA1 protein sequences for the virus and the sources for the HI dataset for influenza A (H1N1), A (H3N2) and B viruses, respectively
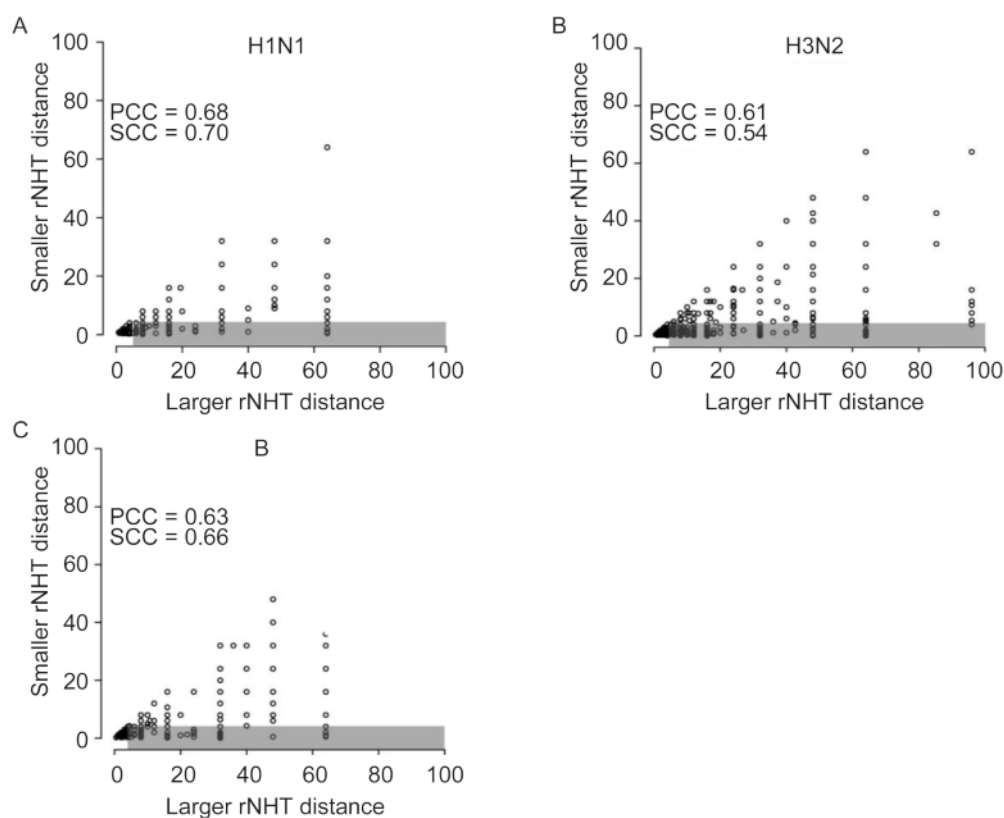

Figure S1. The pairwise two-way rNHT distances for influenza A (H1N1) (A), A (H3N2) (B) and B (C) virus, respectively. For a pair of rNHT distances, the smaller one was displayed along the y-axis, while the larger one was displayed along the x-axis. PCC and SCC refer to the Pearson and Spearman Correlation Coefficient respectively between the two-way rNHT distances. The pairs of rNHT distances in the gray rectangle region refer to those disagree with each other in determining the antigenic variant with the cutoff 4.
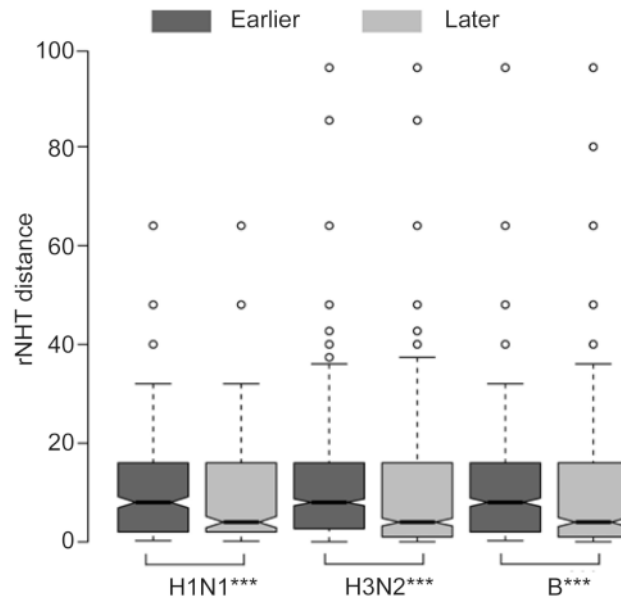
Figure S2. Comparison between the pairwise two-way rNHT distances with the virus generating antisera isolated chronologically for influenza A (H1N1), A (H3N2) and B viruses. "***", p-value < 0.0001.

Table S2. The correlation between the changes on HA1 protein and the difference of the rNHT distances for influenza A (H1N1), A (H3N2) and B virus. Epitopes A~E for influenza A (H1N1), A (H3N2) and B virus were adapted from Deem's (2009), Bush's (2000) and Ni's (2013) work, respectively. PCC, pearson correlation coefficient; SCC, spearman correlation coefficient; RBS, receptor-binding site; N-gly, N-glycosylation sites. All the residues on HA1 were numbered according to each (sub)type

Table S3. Summary of the least squares linear models in influenza H1N1, H3N2 and B virus which take the changes on HA1 protein (see Table S2) as variables to explain the variability of the differences of the pairwise two-way rNHT distances

	H1N1	H3N2	B
Number of variables	49	57	21
p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16
Multiple R-squared	0.54	0.50	0.45
Adjusted R-squared	0.46	0.45	0.42

Table S4. The performance of PREDAC-H3 in determining the antigenic relationship for those viral pairs with rNHT distance in the bin of 4~8 ($\geqslant$ 4 & < 8) for influenza A (H3N2) virus

Method	Accuracy	Sensitivity	Specificity
PREDAC-H3	0.65	0.51	0.81

REFERENCES

Bao YM, Bolotov P, Dernovoy D et al. 2008. J Virol, 82: 596–601.
Deem MW, Pan KY. 2009. Protein Eng Des Sel, 22: 543–546.
Fitch WM, Bush RM, Bender CA, 2000. J Hered, 91: 183–185.
Ni FY, Kondrashkina E, Wang QH. 2013. Virology, 446: 112–122.
Ndifon W, Dushoff J, Levin SA. 2009. Vaccine, 27: 2447–2452.
R Core Team. 2015; Available from: http://www.r-project.org/