MULTI-PARAMETER ANALYSIS OF GENETIC STRUCTURE OF INTRA-HOST HCV POPULATIONS AT EARLY AND LATE STAGES OF INFECTION

Icer P., Artyomenko A., Ramachandran S., Zelikovsky A., Khudyakov Y., Skums P.

Georgia State University, Atlanta, GA, USA
Centers for Disease Control and Prevention, Atlanta, GA, USA
e-mail: pskums@bsu.by

Detection of incident hepatitis C virus (HCV) infections is crucial for identification of outbreaks and development of public health interventions. However, there are no diagnostic assays for distinguishing recent and chronic HCV infections. HCV is highly mutable. Each infected person hosts a heterogeneous population of genetically related HCV variants. Owing to complexity of structural development of intra-host populations affected by bouts of selective sweeps and negative selection during chronic infection [1, 2], simple metrics of genetic heterogeneity are not sufficiently accurate for staging HCV infections.

Here, using intra-host HCV populations sampled by next-generation sequencing of a highly heterogeneous genomic region (HVR1) from 108 recently and 257 chronically infected individuals, we developed a prediction model for differentiating recent and chronic infections. Analysis of 245,878 viral sequences was conducted using 12 parameters for evaluation of various characteristics of HCV populations, including diversity, topological structure, strength of selection and epistasis. In particular, diversity was measured using entropy of the k-mer distribution for each population, as well as by mean, standard deviation and coefficient of variance of pairwise distances among HCV variants. Metrics of selection and epistasis were derived from comparison of sampled and randomized populations generated to reduce effects of epistasis and selection while preserving allele structure. Correlation coefficient between frequency and eigenvector centrality of variants in the observed sequence space was calculated to estimate effect of selection on topological structure of population. A dynamic evolutionary model was applied to simulate variant frequencies for estimation of infection duration by comparing to the observed frequencies.

Support Vector Machine classifier built using the 12 parameters yielded a prediction accuracy of 96.16% before cross-validation and 87.40% in 10-fold cross-validation. The prediction method proposed in [3], which is based solely on the analysis of diversity, yielded an accuracy of 73.42 % when applied to our data.

References

1. Temporal variations in the hepatitis C virus intrahost population during chronic infection / S. Ramachandran et al. // J Virol. – 2011. – V. 85, N 13. – P. 6369–6380.

2. Exceptional Heterogeneity in Viral Evolutionary Dynamics Characterises Chronic Hepatitis C Virus Infection / J. Raghwani et al. // PLoS Pathog. – 2016. – V. 12, N 9. – e1005894.

3. Differentiation of acute from chronic hepatitis C virus infection by nonstructural 5B deep sequencing: A population level tool for incidence estimation /V. Montoya, V. et al. // Hepatology. – 2015. – V. 61, N 6. – P. 1842–1850.