HLA predictions from the bronchoalveolar lavage fluid samples of five patients at the early stage of the Wuhan seafood market COVID-19 outbreak

René L Warren [1,*], and Inanç Birol [1]

[1]Genome Sciences Centre, BC Cancer, Vancouver, BC, V5Z 4S6, Canada

Abstract. We are in the midst of a global viral pandemic, one with no cure and a high mortality rate. The Human Leukocyte Antigen (HLA) gene complex plays a critical role in host immunity. Here we analyze data from five patient samples from the initial COVID-19 outbreak, and report on their HLA profiles. Our analysis results may help provide future insights on disease susceptibility.

We predicted HLA class I alleles from the transcriptome sequencing data prepared from the bronchoalveolar lavage fluid samples of five patients at the early stage of the COVID-19 outbreak. We identified the HLA-I allele A*24:02 in four out of five patients, which is higher than the expected frequency (17.2%) in the South Han Chinese population. The difference is statistically significant with a p-value less than 10^{-4}.

Availability: The tool used to perform the reported analysis results, HLAminer, is available from https://github.com/bcgsc/hlaminer. Predictions are available for download from https://www.bcgsc.ca/downloads/btl/SARS-CoV-2/BAL

*Contact: rwarren@bcgsc.ca

Introduction

SARS-CoV-2 infections have reached global pandemic proportions in early 2020, affecting over 1.5M people worldwide (as of this writing) and showing no signs of easing, except in a few jurisdictions where strict quarantine measures were implemented early on. The resulting coronavirus disease (COVID-19) has a relatively high ($\sim$3.4%) mortality rate [1] – a figure that varies widely between jurisdictions due to factors yet to be determined. Currently, no vaccines or effective treatments are available. Most current data analysis efforts are, understandably, focused on the virus itself for the purpose of vaccine development and tracking its evolution for diagnostics and infection monitoring purposes.

Curiously, it is estimated that as high as 18-30% or more of the population may be asymptomatic to SARS-CoV-2 infections [2] [3], while other affected individuals exhibit mild to severe to critical symptoms of infection. Thus, gaining insights on host susceptibility to the coronavirus is clearly another important aspect that needs to be worked on and understood. One would expect a link

between host immunity genes and susceptibility or resistance to infection. The Human Leukocyte Antigen (HLA) gene complex includes two classes of such genes, which encode the Major Histocompatibility Complex (MHC). Proteins of the MHC present (class I) internally- or (class II) externally-derived antigenic determinants (epitopes) to T cells, which upon recognition of the epitope-complex, will mount an immune response to defend against viral and bacterial infections. HLA genes are therefore cornerstone to acquired immunity in humans. HLA alleles have also been shown to be factors in susceptibility or resistance to certain diseases, and their frequency and composition in human populations vary widely (http://allelefrequencies.net). A previous study found HLA class I genes HLA-B*46:01 and HLA-B*54:01 to be associated with the 2003 severe acute respiratory syndrome (SARS) coronavirus infections in Taiwan [4] – a related disease to the current pandemic.

For over a decade, high-throughput transcriptome sequencing has proven a worthy instrument for measuring changes of gene expression in human diseases and beyond [5]. Transcriptome analysis has the potential to reveal key genes that are modulated in response to infections, but also has the potential to reveal the HLA composition of affected individuals. A few years ago, our group developed an approach for mining high-throughput next-generation shotgun sequencing data for the purpose of HLA determination [6], which has since been applied in a broader clinical context [7].

Here, we report our initial observations based on transcriptome sequencing (RNA-Seq) libraries prepared from the bronchoalveolar lavage fluid samples of five patients at the early stage of the Wuhan seafood market pneumonia coronavirus outbreak (see Methods). We identified the HLA class I group A allele A*24:02 in four out of the five individuals. Although HLA-A*24:02 is common and high-frequency in populations from Southeast Asia (reaching 86.3% in Taiwanese aborigines), it is coincidentally higher than the expected HLA-A*24:02 allele frequency in the Chinese population (est. 17.2%).

Methods

We downloaded Illumina HiSeq 3000 paired RNA-Seq reads from libraries prepared from the bronchoalveolar lavage fluid samples of five patients (https://www.ebi.ac.uk/ena/data/view/PRJNA605983 accessions SRX7730880-SRX7730884 denoted in the table below as Patient 1-5, respectively). We note that these are metagenomics samples, and, although not explicitly noted, we think that they were prepared for the primary purpose of identifying and characterizing the novel coronavirus at the outbreak epicentre. On each dataset, we ran HLAminer [6] in targeted assembly mode with default values (v1.4; contig length $\geq$200bp, seq. identity $\geq$99%, score $\geq$1000), and report 4-digit (HLA allele/protein) resolution when top-scoring predictions are unambiguous. Otherwise the 2-digit (allele group) resolution is reported.

Results and Discussion

We predicted and compiled the likely HLA class I alleles for five patients at the early stage of the COVID-19 outbreak in Wuhan, China (Table 1). Although the bronchoalveolar lavage (BAL) fluid samples were initially utilized to identify and characterize the novel coronavirus (similar justification in [8]), BAL metagenomics samples are expected to contain host cells / nucleic acids (DNA/RNA). Because HLA genes are expressed at the surface of all human nucleated cells, RNA-Seq data can be employed to determine HLA profiles from BAL samples.

Table 1. HLA-I predictions from the bronchoalveolar lavage fluid samples of five patients at the early stage of the Wuhan seafood market pneumonia coronavirus outbreak. Highest-scoring *HLAminer* predictions are shown for HLA class I genes A, B and C. Common HLA alleles between two or more patients are highlighted in bold face.

Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
A*01:01P	A*30:01P	**A*24:02P**	**A*24:02P**	A*29
A*24:02P	A*02:06P	A*26:01P	-	**A*24**
B*35/B*57	**B*51:01P**	B*15:01P	B*40:01P	B*54:01P
B*48	B*13:02P	**B*51:01P**	B*13:01P	B*07:05P
C*08:72P	C*14:02P	C*15:02P	C*04:03P	C*15
C*06:02P	**C*06:02P**	C*03:03P	C*03:04P	-

We observe the HLA-A*24:02 allele in four out of five (80%) patients. HLA-A*24 is a common group of alleles in Southeastern Asian populations, and the frequency of HLA-A*24:02 allele is high, especially in indigenous Taiwanese populations, reaching as high as 86.3%. However, our understanding is that the five patients are from the Wuhan market area, and the associated A*24:02 frequency in Chinese population is typically 17.2%, a value that is statistically significantly less than our observed frequency of 80% (p$< 10^{-4}$, 1-sided z-test).

We note that A*24 has not been previously reported as a risk factor for SARS infection [9], but there are reports of disease association with HLA-A*24:02, notably with diabetes [10] [11] [12] [13], which is a reported potential risk factor in COVID-19 patients [14]. HLA also informs vaccine development. This knowledge would help prioritize SARS-CoV-2 derived epitopes predicted to be stable HLA binders [15].

Future studies into host susceptibility and resistance to SARS-CoV-2 infections are sorely needed as they may help us better manage and mitigate risks of infections. We chose to communicate our early findings in this domain to facilitate rapid development of response strategies.

Funding

References

1. Rajgor, D.D. et al. (2020) The many estimates of the COVID-19 case fatality rate. The Lancet. doi:10.1016/S1473-3099(20)30244-9
2. Mizumoto, K. et al. (2020) Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill. 25, pii=2000180. doi:10.2807/1560-7917.ES.2020.25.10.2000180
3. Nishiura, H. et al. (2020) Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int. Journal of Infectious Diseases, doi:10.1016/j.ijid.2020.03.020
4. Lin, M. et al. (2003) Association of HLA class I with severe acute respiratory syndrome coronavirus infection. BMC Med Genet. 4, 9
5. Wang, Z. et al. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 10, 57-63
6. Warren, R.L. et al. (2012) Derivation of HLA types from shotgun sequence datasets. Genome Med. 4, 95
7. Brown, S.D. et al. (2014) Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. Genome Res. 24, 743-750
8. Wu F. et al. (2020) Complete genome characterisation of a novel coronavirus associated with severe human respiratory disease in Wuhan, China. bioRxiv. doi:10.1101/2020.01.24.919183
9. Sun, Y. and Xi, Y. (2014) Association Between HLA Gene Polymorphism and the Genetic Susceptibility of SARS Infection. Book: HLA and Associated Important Diseases. IntechOpen. 12, doi:10.5772/57561
10. Adamashvili, I. et al. (1997) Soluble HLA class I antigens in patients with type I diabetes and their family members. Human Immunol. 55, 176–83
11. Noble, J.A. et al. (2002) The HLA class I A locus affects susceptibility to type 1 diabetes. Human Immunol. 63, 657–664
12. Nakanishi, K. and Inoko, H. (2006) Combination of HLA-A24, -DQA1*03, and -DR9 contributes to acute-onset and early complete beta-cell destruction in type 1 diabetes: longitudinal study of residual beta-cell function. Diabetes. 55, 1862–1868
13. Kronenberg, D. et al. (2012) Circulating preproinsulin signal peptide-specific CD8 T cells restricted by the susceptibility molecule HLA-A24 are expand-ed at onset of type 1 diabetes and kill β-cells. Diabetes. 61, 1752–1759
14. Guan, W.J. et al. (2020) Comorbidity and its impact on 1590 patients with Covid-19 in China: A Nationwide Analysis. Eur Respir J. doi:10.1183/13993003.00547-2020
15. Prachar, M. et al. (2020) COVID-19 Vaccine Candidates: Prediction and Validation of 174 SARS-CoV-2 Epitopes. bioRxiv. doi:10.1101/2020.03.20.000794