Title:

Genetic mismatch explains sizable variation of COVID-19 vaccine efficacy in clinical trials

Authors:

Lirong Cao[1,2], Shi Zhao[1,2], Jingzhi Lou[1,2], Hong Zheng[1,2], Chris Ka Pun Mok[1], Renee WY Chan[3, 4], Marc Ka Chun Chong[1,2], Zigui Chen[5], Paul KS Chan[5], Benny Chung-Ying Zee[1,2], Maggie Haitian Wang[1,2*]

Affiliations:

[1] JC School of Public Health and Primary Care, Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China

[2] CUHK Shenzhen Research Institute, Shenzhen, China

[3] Department of Paediatrics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China

[4] CUHK-UMCU Joint Research Laboratory of Respiratory Virus & Immunobiology, Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China

[5] Department of Microbiology, Stanley Ho Centre for Emerging Infectious Diseases, Li Ka Shing Institute of Health Sciences, Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China

Correspondence: Maggie Haitian Wang, email: maggiew@cuhk.edu.hk, JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong SAR, China.

Abstract

Pulling vaccine efficacy and effectiveness (VE) outcomes from 17 reports of 9 different vaccine products and through sequence analysis, we found that genetic mismatch explained sizable variations in VE. The findings suggested the potential need of timely optimizing vaccine antigens as new dominant viral strains emerge.

Keywords: COVID-19; vaccine efficacy; genetic mismatch; Spike protein; receptor-binding domain.

Introduction

The novel coronavirus disease 2019 (COVID-19) has caused devastating consequences to global health. Vaccines against the pathogen, SARS-CoV-2, bear the hope of reducing the COVID-19 associated severe diseases, mortality and mitigating the pandemic scale. To date, over 90 vaccine candidates have been tested in phase 1/2/3 trials, and 13 products have been approved in at least one country [1]. Vaccine efficacy summarizes the relative proportion of protection against a disease outcome in the vaccinated group versus the placebo group in clinical trials, while vaccine effectiveness measures this quantity in a real-world scenario. The published clinical trial data for COVID-19 vaccines exhibited a large variation in efficacy ranging from 10.4% to 97.2% [2, 3]. While these vaccines incorporated different platforms (mRNA, viral vector, protein subunit and inactivated vaccines), the clinical trials were also conducted in different time periods, locations and populations. In our previous studies of the influenza virus, we demonstrated that genetic mismatch of circulating viruses against the vaccine strain significantly and negatively influenced vaccine effectiveness in population [4]. For COVID-19, experimental studies showed that the genetic variants, B.1.351, emerged in January 2021, escaped neutralization by South African donor plasma [5]; and a single mutation, E484K, may lead to a loss of neutralizing activity by vaccine-elicited antibodies [6]. We hypothesize that genetic mismatch against the vaccine virus can be used to account for the divergence of vaccine efficacy observed in clinical trials.

The spike (S) protein of SARS-CoV-2 is the immunodominant antigen during infection and immunization, and current vaccines incorporated the S protein as the main target derived from the initial Wuhan strain [7]. In this study, we calculated the average genetic distance of circulating strains against the vaccine strain during respective application periods and regions of vaccine trials, and analyzed their relationship with the reported vaccine efficacy or effectiveness (VE). We reported the extend of genetic drifts on affecting COVID-19 VE in population.

Methods

Vaccine Efficacy and Effectiveness Data We collected the VE published before April 22, 2021 in journal articles, Food and Drug Administration and other government reports. Seventeen reports were available, including thirteen phase 3 trials, two phase 2 trials, and three observational study, regarding mRNA vaccines, viral-vector vaccines, protein subunit vaccines, and inactivated vaccines. Twenty-two VEs for protecting the primary outcomes described in the studies were extracted for analysis, which is symptomatic COVID-19 infection or confirmed infection requiring medical care. The detailed information is available in **Supplementary Materials Table S1.1**.

Genetic Sequences The human SARS-CoV-2 strains were retrieved on April 22, 2021 from the global initiative on sharing all influenza data (GISAID) [8]. All available sequences matched to the period and location of the clinical trials or the observational study were downloaded. After removing duplicated strains, a

total number of 121,373 full-length genome sequences were prepared, sampled from Argentina, Bahrain, Brazil, Chile, Colombia, Egypt, Indonesia, Israel, Jordan, Mexico, Pakistan, Peru, Russia, South Africa, Turkey, the United Arab Emirates, the United Kingdom, and the United States of America. The collection date of the sequences ranged from January 1, 2020 to April 1, 2021. Multiple sequence alignment was performed using MAFFT (version 7). The 'Wuhan-Hu-1' genome (GenBank 'NC_045512.2', GISAID 'EPI_ISL_402125') was considered as the reference sequence. All sequence data used in the analysis were listed in the Supplementary Acknowledgement Table.

Statistical Methods The genetic mismatch (*G*) was calculated by the average Hamming distance between the reference sequence and all available circulating strains in a region during the trial periods. Three genomic segments were considered in the calculation, which include the complete S gene sequence covering 1,273 amino acids, receptor-binding domain (RBD) of the S protein containing 223 amino acids, and 16 selected key mutation sites in the S protein. The key mutation is defined as the prevalence of an amino acid substitution reaching dominance (>0.5) in population. Then, 3 sites in the RBD, 5 sites in the N-terminal domain and 8 from other codons in the S1 and S2 subunits were identified as key mutations (**Supplementary Materials Table S1.2**). The statistical relationship between genetic mismatch and VEs was evaluated following the general framework proposed in our previous study for influenza vaccine [4]. As the main target of the COVID-19 vaccine is S protein and restricted by sample size, we adopted a simpler version of the model with only one variable, genetic distance (G), as the predictor, and VE being the dependent variable, and the model reduced to a simple linear regression. The three genomic regions were evaluated separately. Statistical significance is declared if p-value < 0.05. All analyses were performed using **R** statistical software (version 4.0.3).

Results

We first described the general distributions of VEs and genetic mismatch across vaccine platforms. The mean VE is the highest for the mRNA vaccines of 93.6%, followed by the inactivated vaccines of 74.6%, whereas for protein subunit vaccines and viral-vectored vaccines, the VEs are 69.4% and 64.5%, respectively (ANOVA p-value = 0.048, Figure 1a). Interestingly, the degree of genetic mismatch shows a reversed order, which is lower for the mRNA and the inactivated vaccines and larger for the protein subunit and the viral-vectored vaccines (Figure 1a). These should be influenced by the amount of new genetic variants appeared in the trial locations and periods: the earlier the sample recruitment and the less new lineages circulation during the trial period would result in a smaller genetic mismatch. Next, we investigated the statistical relationship between the genetic mismatch and VE. The genetic mismatch (*G*) is significantly associated with vaccine protection: 64.8% of the variations in VEs of independent cohorts can be explained by the genetic mismatch calculated from the RBD (p-value < 0.0001, Figure 1b), while 51.7% of the VE deviations can be predicted by mismatch based on the whole S gene (p-value < 0.0001, Figure 1d), and 42.6% can be explained by mismatch of the key mutations (p-value < 0.001, Figure 1c).

The results show that the RBD region plays a dominant role in affecting the VEs. We estimated that the average VE reduction stemming from per amino acid substitution in the RBD is 22.6% [95% CI: 15.7 – 29.6%], see Figure 1b. The extremely low VE in Figure 1(b-d) is originated from a South African trial against the B.1.351 variant [3]. Remarkably, this point can be captured by the prediction interval of the regression line (Figure 1b). The RBD mutations with mutation prevalence >1% in each trial region are shown in **Supplementary Materials Figure S2**. Specifically, the mutations K417T/N, E484K, and N501Y were observed in Brazil and South Africa; in the United Kingdom, the set was N439K, S477N, N501Y; while in the United States, the relatively common mutations in the RBD were V382L, L452R, A475V, and A520S. By contrast, in the full-length S gene, per genetic mismatch would generate a lower decrease in effectiveness by 7.9% [95% CI: 4.7 – 11.1%] (Figure 1d). Although the key mutation group only consists of 16 sites, the explained variation can also reach 42.6% (Figure 1c).

Discussion

The VE estimates were evidently different across various vaccine platforms. Since clinical trials of the COVID-19 vaccine were conducted in different periods and places, it is difficult to directly compare vaccine performance or explain their difference. This study provides evidence of genetic factors affecting the efficacy of COVID-19 vaccines. We demonstrated that a larger genetic distance of the circulating SARS-CoV-2 viruses to the vaccine strain might cause weaker vaccine-induced protection. Similar findings have been reported for the influenza virus that influenza vaccine effectiveness can be affected by genetic mismatch in key mutations [4].

The decrease in VE associated with the RBD substitution is biologically reasonable according to the latest findings in literature. Evidence suggests that the neutralization potency of antibodies is reduced in the vaccine-elicited sera against SARS-CoV-2 variants encoding K417N/E484K/N501Y [6, 9]. Another mutation, N439K, was shown to modestly enhance the binding affinity to the host receptor hACE2 and confer resistance against several neutralizing monoclonal antibodies [10]. Moreover, the 475V and 477N variants were also reported to reduce antibody-mediated immunity [11, 12]. From the perspective of genetic epidemiology, our findings highlighted that RBD mutations may weaken vaccine protection. As for non-RBD substitutions, further investigation is needed to determine their roles in viral immune escape.

In summary, this study identified a negative relationship between VE and genetic mismatch of circulating SARS-CoV-2 strains against vaccine strains in three genomic regions. This study has several limitations. Although the current model reached good statistical significance, the complexity of the model is restricted by the sample size of the available studies reporting VE. Thus, the platform effect and population effect on vaccine performance cannot be addressed. Secondly, the existing genetic mismatch was evaluated by

measuring genetic distance, and other chemical and spatial structure changes associated with amino acid substitutions were not considered. Nevertheless, we demonstrated that genetic mismatch explains a sizable portion of VE deviations in clinical trials. These results imply that it is essential to monitor newly arising variants, and COVID-19 vaccines will need to be updated periodically to avoid potential loss of efficacy.

Conflicts of Interest

M.H.W and B.C.Y.Z are shareholders of Beth Bioinformatics Co., Ltd. B.C.Y.Z is a shareholder of Health View Bioanalytics Ltd. All other authors declare no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [31871340, 71974165], the Hong Kong Health and Medical Research Fund [INF-CUHK-1], and the Chinese University of Hong Kong Direct Grant [4054456, 4054524].

Acknowledgements

SARS-CoV-2 sequences were retrieved from the global initiative on sharing all influenza data at http://platform.gisaid.org/. The complete acknowledgement table could be found in online supplementary materials. We thank the contributions of the submitting and the originating laboratories.

Author Contributions

M.H.W conceived the study, L.C, M.H.W and S.Z wrote the manuscript. L.C and H.Z. collected and processed the data. L.C carried out the analysis. All authors critically read and revised the manuscript and gave final approval for publication.

Ethics approval and consent to participate

The ethical approval or individual consent was not applicable.

References

1. COVID-19 VACCINE TRACKER. Available at: https://covid19.trackvaccines.org/vaccines/. Accessed March 17, 2021.
2. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med **2020**; 383(27): 2603-15.
3. Madhi SA, Baillie V, Cutland CL, et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. N Engl J Med **2021**.
4. Cao L, Lou J, Zhao S, et al. In silico prediction of influenza vaccine effectiveness by sequence analysis. Vaccine **2021**.
5. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. Nat Med **2021**.
6. Collier DA, De Marco A, Ferreira IA, et al. SARS-CoV-2 B. 1.1. 7 escape from mRNA vaccine-elicited neutralizing antibodies. medRxiv **2021**.
7. Carvalho T, Krammer F, Iwasaki A. The first 12 months of COVID-19: a timeline of immunological insights. Nat Rev Immunol **2021**: 1-12.
8. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill **2017**; 22(13).
9. Wang Z, Schmidt F, Weisblum Y, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature **2021**.
10. Thomson EC, Rosen LE, Shepherd JG, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell **2021**; 184(5): 1171-87.e20.
11. Liu Z, VanBlargan LA, Bloyet LM, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe **2021**; 29(3): 477-88.e4.
12. Li Q, Wu J, Nie J, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell **2020**; 182(5): 1284-94.e9.

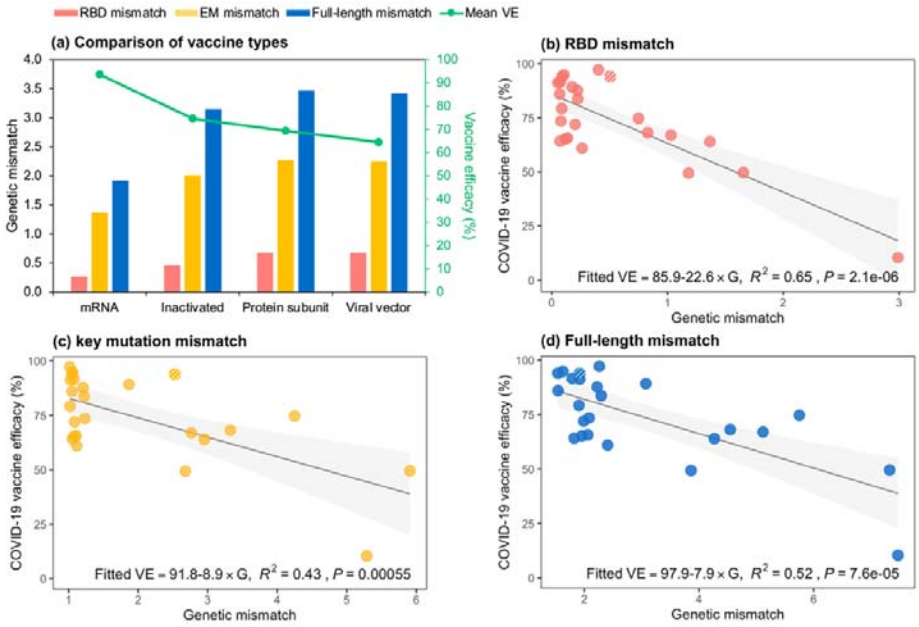

Figure 1. The comparison among four types of COVID-19 vaccines, and the relationship between vaccine efficacy (VE) and genetic mismatch of the circulating SARS-CoV-2 strains to the vaccine strain in S gene. Panel (a): comparison of genetic mismatch and VE for mRNA, inactivated, protein subunit and viral vector vaccines. Panels (b-d): negative linear relationships between VE and genetic mismatch for RBD (red dots), key mutations (yellow dots), and full-length sequence (blue dots), respectively. All points in panels (b-d) are vaccine efficacy except that the shaded point is vaccine effectiveness.