In silico analysis of gene expression and binding affinity to predict the susceptibility to Covid-19 of Pan-Cancer patients

Harim Chun, 20152629, School of Systems Biomedical Science, Soongsil University

Introduction

In recent times, all over the world is suffering from a tiny novel coronavirus (SARS-CoV-2) which causes severe acute respiratory syndrome (Covid-19). This virus is so harmful that taking the lives of the people. People who already have cardiovascular disease or underlying disease are more susceptible than healthy people [1]. Moreover, cancer patients are more vulnerable to SARS-CoV-2 infection [2].

SARS-CoV-2 has a spike protein that is needed to bind to the host cell. This protein targets human ACE2 as a receptor and invades our cell. Therefore, the viral infection is directly interconnected with the binding affinity of the ACE2 and SARS-CoV-2. Also, CXCL10 is a crucial cytokine that leads to cytokine storms. This cytokine is responsible for developing acute respiratory distress syndrome (ARDS). Considering the cancer patients' susceptibility to SARS-CoV-2, there would be a relation between ACE2 and CXCL10 expression or mutation and Covid-19. This study aimed for a comprehensive analysis of various cancer types' ACE2 and CXCL10 gene expression by using TCGA data. Homology modeling and binding affinity prediction were also conducted to investigate the difference in binding affinity. The overall study was progressed with various well-known databases.

Materials & Methods

1. Expression Analysis

1.1 Analyzing the gene expression in Various Cancer Type

ACE2 and CXCL10 mRNA expression pattern in various cancer type were acquired from Tumor Immune Estimation Resource (TIMER 2.0, http://timer.cistrome.org/)[1][2]. This web server-based resource provides a comprehensive analysis of immune infiltrates across diverse cancer types. This web-resource consists of 3 major menus. From those 3 menus, Cancer Exploration allows users to analyze the gene expression data and provides significant cancer types according to the differential expression of the specific gene.

1.2 Further Analysis of gene expression

Even TIMER 2.0 supports the gene expression data, there are insufficient data of the gene expression from normal and tumor patients. UALCAN web portal (http://ualcan.path.uab.edu/) offers a comprehensive resource for analyzing cancer omics data [3]. This web-based platform provides easy access to cancer omics data including pan-cancer gene expression from The Cancer Genome Atlas (TCGA) data. From UALCAN, users can get gene expression differences between the normal samples and tumor samples for 35 types of cancer [3].

2. Functional annotation of the targeted genes and Binding affinity prediction

2.1 Mutation of the gene according to the cancer types

Significant genetic change in ACE2 and CXCL10 was investigated using cBioPortal for Cancer Genomics (https://www.cbioportal.org/). This website provides visualization, analysis, and download of large-scale cancer genomics data sets. From an enormous analysis of the various cancer types, it provides

information about the gene amplification, variation, survival analysis, and gene fusion. Especially, the data about the missense, truncating, and in-frame mutations of a specific gene are offered [4]. Also, whether a single point protein mutation's Functional impact is provided by SIFT, PolyPhen-2, and Mutation Assessor.

2.2 Homology Modeling with Altered amino acid sequences

According to the genomic variation, protein's structure which translated from a mutated genome would be changed. There are various methods to predict the protein structure from a first-dimensional amino acid sequence such as Homology modeling, Template-Based Prediction, and the Ab initio prediction methods. However, the Ab initio prediction method is not adequate because of the computing resource. Therefore, with a web-based tool called SWISS-MODEL (https://swissmodel.expasy.org/)[5], protein structures of mutated genes were predicted. Also, fasta file containing a mutated amino acid sequence was generated by Python programming (Fig. 1, Supplementary Material. 1). UCSF Chimera was used to visualize 3D structures and retrieve the human ACE2 protein sequence [6].

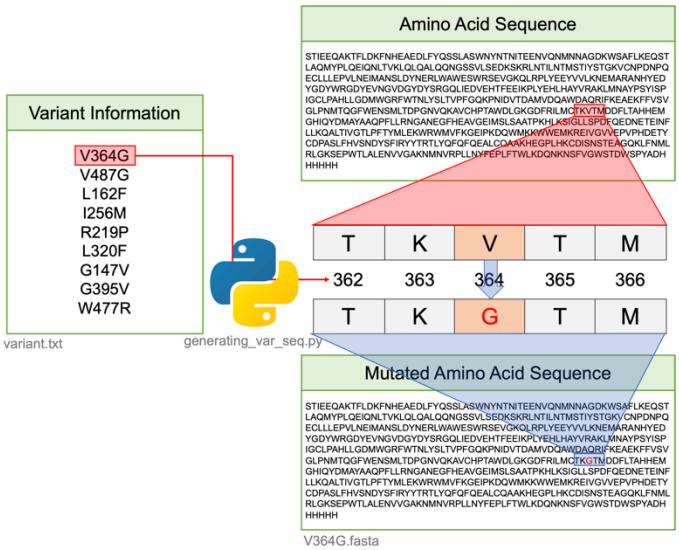

Figure 1. Python programming scheme Generating .fasta files containing the information about the mutated amino acid sequence from a wild type amino acid sequence. When variant information is given, python code automatically finds an exact mutation position, alter the amino acid sequence, and save the fasta file.

2.3 Predicting the binding affinity between SARS-CoV-2 spike protein and Mutated ace2

Protein's structure doesn't mean just a mere structure. Its structure is related to its function. Especially, structural alteration of ace2 protein which is a major receptor of SARS-CoV-2 spike protein might affect the binding affinity between them. Predicting the binding affinity is time and high-computing-demanding work. To solve this problem, this research was conducted by aiding ISLAND (In SiLlico protein AffiNity preDictor) [7]. ISLAND is a tool for predicting the binding affinity of the protein only by amino acid sequence based on machine learning.

3. Materials

SARS-CoV-2 spike glycoprotein sequence(P59594, Spike_CVHSA) was retrieved from Uniprot (https://www.uniprot.org/). Human ACE2 protein sequence was retrieved from a SARS-CoV-2 spike glycoprotein and ACE2 bound complex.

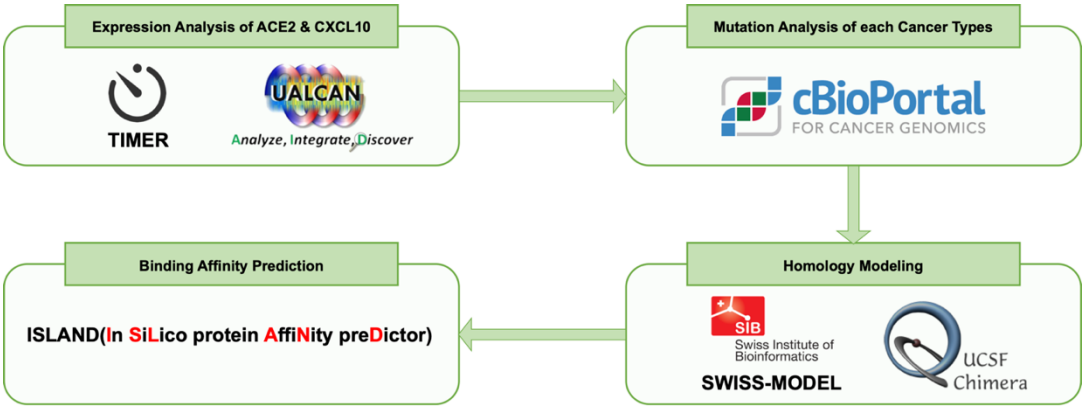

Figure 2. Overall Study Workflow Various web-based databases were used to conduct the study.

Results

1. Differentially Expressed ACE2 and CXCL10 in a certain Cancer Types

With TIMER 2.0, differential gene expression of ACE2 and CXCL10 between tumor and adjacent normal tissues were calculated. It also provides the statistical significance of the differential gene expression. According to the TIMER 2.0, ACE2 mRNA was differentially expressed in the BRCA(Breast Invasive Carcinoma), COAD(Colon Adenocarcinoma), KICH(Kidney Chromophobe), LIHC(Liver Hepatocellular carcinoma), PRAD(Prostate Adenocarcinoma), SKCM(Skin Cutaneous melanoma), STAD(Stomach adenocarcinoma), and THCA(Thyroid Cancer) where all the cases' p-value was <0.05. Differentially expressed CXCL10 was also captured in the BRCA, COAD, ESCA(Esophageal carcinoma), HNSC(Head and Neck squamous cell carcinoma), KICH, KIRC(Kidney renal clear cell carcinoma), LUAD(Lung adenocarcinoma), LUSC(Lung squamous cell carcinoma), PRAD(Prostate adenocarcinoma), READ(Rectum adenocarcinoma), SKCM, STAD, and UCEC(Uterine corpus endometrial carcinoma) with p-value < 0.05 (Fig. 3). After that, both significant cancer types were selected. Those cancer types were BRCA, COAD, KICH, PRAD, SKCM, STAD. Even LUSC and LUAD were not significant for the ACE2, they were also selected as significant cancer types because the COVID-19 is a respiratory disease.

In order to certain the difference of the gene expression, this study referenced the UALCAN. Statistical significance was calculated according to the differential expression of the ACE2 and CXCL10. Significant cancer types were filtered by the p-value < 0.05 and remains were BRCA, KICH, LUAD, and LUSC. Interestingly, ACE2 expression in BRCA tumor sample was lower than normal (Supplementary Fig. 1).

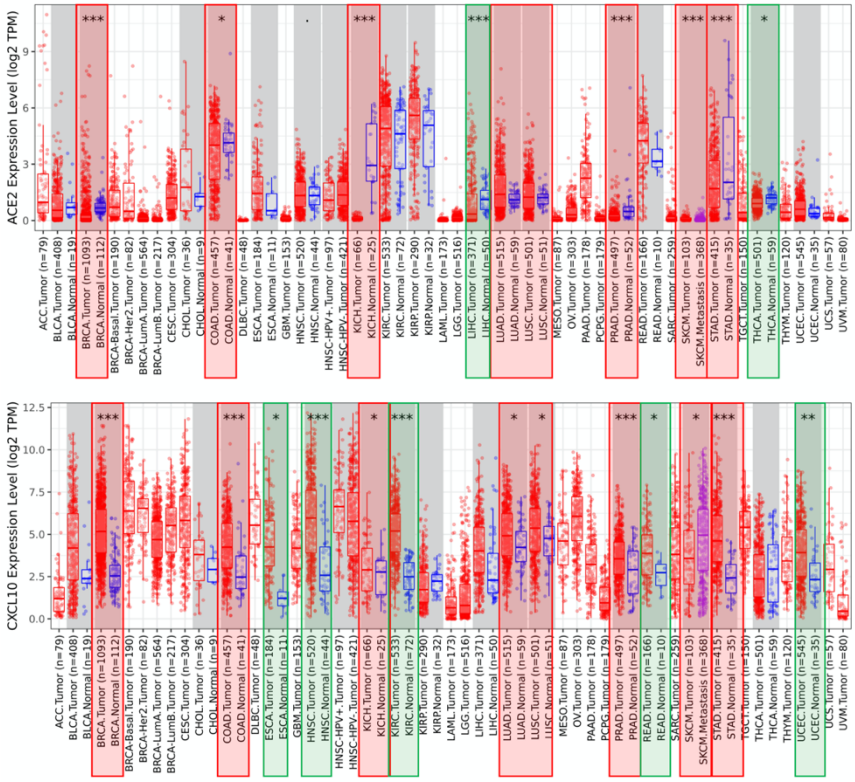

Figure 3. Differentially Expressed Gene in TIMER 2.0 Differentially expressed ACE2 and CXCL10 in various cancer types. Cancer types that are in the red box mean the significant cancer types both differentially expressed. Green boxed cancer types indicate that only a single gene is differentially expressed.

2. Functional annotation of the targeted genes and Binding affinity prediction

To evaluate the functional significance of ACE2 and CXCL10 in various cancer types, mutational information was retrieved from a cBioPortal for Cancer Genomics. 3,385 samples of BRCA patients, 1,382 samples of LUAD patients, 1,176 samples of LUSC patients, and 773 samples of KICH patients were collected and investigated ACE2 and CXCL10 mutation. There was no mutation in the CXCL10 in all filtered cancer types. However, 9 variations in ACE2 were detected in BRCA, LUAD, and LUSC each three. There was no mutation in ACE2 in KICH. 9 variations' functional impact deleterious refers to PolyPhen-2.

The altered amino acid sequence was generated automatically by a python programming. Those mutated amino acid sequences were used for the homology modeling by a SWISS-MODEL (Fig. 4).

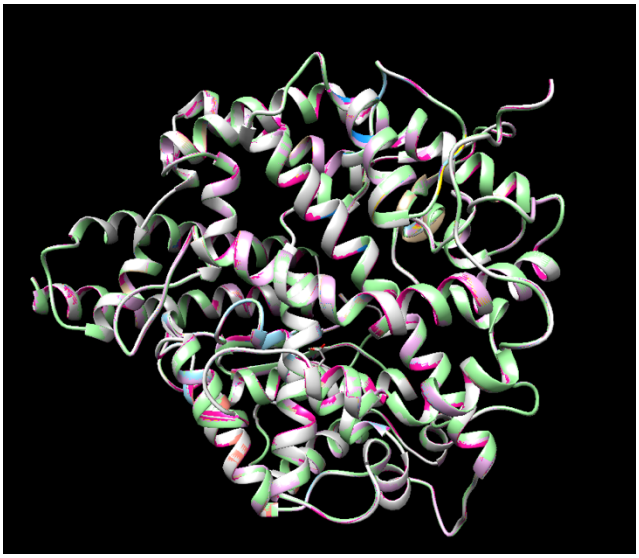

Figure 4. Mutated ACE2 protein structures 9 Mutated ACE2 protein structures. The protein structure was predicted by a SWISS-MODEL. Different colors indicate individual proteins. Most of the proteins are overlapped so that only a few colors are visible.

Wild type ACE2 was used as a template and 9 cases' 3-dimensional structure was constructed. Those models were gathered in a UCSF Chimera and no big structural difference between them. Also, there was no difference in binding affinity of ACE2 and SARS-CoV-2 spike protein predicted with ISLAND between normal and patients (Table. 1).

Variations	Biding Affinity(ΔΔG)	Dissociation Constant (Kd)	Cancer Types
Normal	-11.67	2.76E-09	-
V364G	-11.67	2.76E-09	BRCA
V487G	-11.67	2.76E-09	
L162F	-11.67	2.76E-09	
I256M	-11.678	2.72E-09	LUAD
R219P	-11.653	2.84E-09	
L320F	-11.67	2.76E-09	
G147F	-11.67	2.76E-09	LUSC
G395V	-11.67	2.76E-09	
W477R	-11.635	2.92E-09	

Table 1. Predicted binding affinity of each Variants

Discussion

Through a Differential Gene Expression analysis, most of the cancer types showed upregulation of ACE2 and CXCL10. ACE2 is well known for a receptor SARS-Co-V-2 spike protein utilize. Therefore, viral infection and pathogenicity are related to the binding affinity of the ACE2 receptor [8]. As ACE2 expression increased, the virus can enter into a cell by binding of spike proteins and ACE2. After SARS-CoV-2 enters the cell, they make various proteins they need, and our immune system is then activated. CXCL10 and other cytokines would be over-regulated and cause cytokine storm. This cytokine storm is responsible for the development of acute respiratory distress syndrome (ARDS). ACE2 and CXCL10's upregulation in LUSC and LUAD patients are strongly related to the viral infection. Co-expressed genes of ACE2 mainly have a function of biding and CXCL10 has immune response. Therefore, tumor patients would be more vulnerable to SARS-CoV-2 infection.

In a normal state, people can regulate blood pressure, anti-inflammation, and blood clotting by Nitrogenous oxide (NO) generated from endothelial cells. SARS-CoV-2 invades pulmonary alveolus' epithelial cells and causes ARDS. As a result, white blood cells gathered, and NO yields dropped which can cause uncontrollable blood clotting. It ends up with a destructive phase of inflammation (Fig. 5) [9]. However, the expression level of ACE2 in BRCA patients is decreased compared with normal samples. The function of ACE2 is mainly catalyzing the conversion of angiotensin II to angiotensin 1-7 and counterbalancing ACE activity. The enzymatic activity of ACE2 has a protective role in cardiovascular diseases. Therefore, BRCA patients' vascular environment would be adequate to be developed into the ARDS state at the lung tissue. Also, BRCA patients' mRNA expression is usually measured at the cancer type-specific sites. A decrease of ACE2 expression of patients would cause inflammation near the breast and affect other consequences.

The binding affinity of the ACE2 and SARS-CoV-2 spike protein was not significantly different. Because of the single amino acid sequence mutation, overall function or structure difference is not affected dynamically. However, when a gene fusion or copy number alterations of ACE2 occurs, its function and binding would be different.

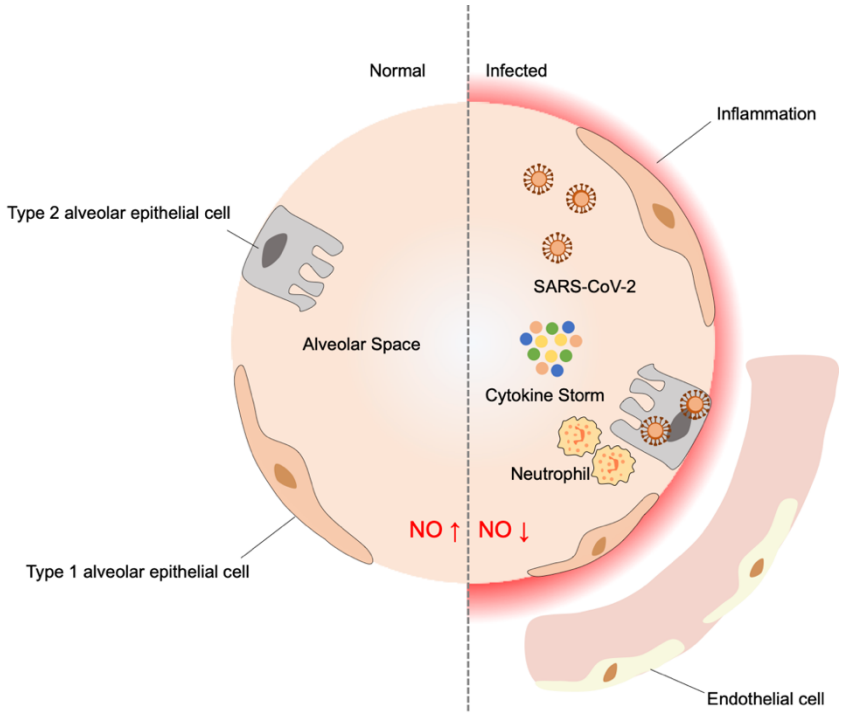

Figure 5. SARS-CoV-2 infected environment Environmental difference between normal and infected states.

Reference

[1] Yang, Jing, et al. "Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis." ***International Journal of Infectious Diseases*** 94 (2020): 91-95.

[2] Hanna, Timothy P et al. “Cancer, COVID-19 and the precautionary principle: prioritizing treatment during a global pandemic.” ***Nature reviews. Clinical oncology*** vol. 17,5 (2020): 268-270. doi:10.1038/s41571-020-0362-6

[3] Taiwen Li, Jingyu Fan, Binbin Wang, Nicole Traugh, Qianming Chen, Jun S. Liu, Bo Li, X. Shirley Liu. TIM ER: A web server for comprehensive analysis of tumor-infiltrating immune cells. ***Cancer Research.*** 2017;77(21):e108-e110.

[4] Taiwen Li, Jingyu Fan, Binbin Wang, Nicole Traugh, Qianming Chen, Jun S. Liu, Bo Li, X. Shirley Liu. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. ***Cancer Research.*** 2017;77(21):e108-e110.

[5] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Rodriguez IP, Chakravarthi BVSK and Varambally S. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. ***Neoplasia***. 2017 Aug;19(8):649-658. doi: 10.1016/j.neo.2017.05.002 [PMID:28732212]

[6] Cerami, Ethan et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. ***Cancer discovery vol. 2,5*** (2012): 401-4. doi:10.1158/2159-8290.CD-12-0095

[7] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. ***Nucleic Acids Res.*** 46(W1), W296-W303 (2018).

[8] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC *et al*. UCSF Chimera, a visualization system for exploratory research and analysis. ***Journal of computational chemistry*** 2004; *25*:1605-1612.

[9] Wajid Arshad Abbasi and Fayyaz-ul-Amir Afsar Minhas (2017), "ISLAND: *In silico* prediction of proteins binding affinity using sequence descriptors", DCIS, PIEAS

[10] Xu, H. *et al*. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int*. *J*. *Oral Sci*. 1–5 (2020) doi:10.1038/s41368-020-0074-x.

[11] Teuwen, L., Geldhof, V., Pasut, A. *et al.*COVID-19: the vasculature unleashed. ***Nat Rev Immunol*** (2020).