TRANSFAC revealed IRF3 site in SNPs associated with high severity of Covid-19, missed by JASPAR

The COVID-19 pandemic has been estimated to cause 6.6 million deaths so far (1). The predominant cause of mortality is pneumonia and severe acute respiratory syndrome. Several genome-wide association studies (GWAS) were carried out for identifying candidate genes and pathways that play a role in COVID-19.

A genetic study by Zeberg & Pääbo (2), by comparing with Neanderthal DNA, has identified a gene cluster on chromosome 3 as a risk locus for respiratory failure post-infection with SARS-CoV-2. COVID-19 host genetics initiative (3) also proved that the risk locus on chromosome 3 is the major risk factor for severe symptoms of covid-19.

Using CRISPRi analysis (4) it was proposed that a region near rs11385942 at chromosome 3p21.31 (locus of highest significance with COVID-19 disease severity at intron 5 of LZTFL1) significantly affected the expression of LZTFL1 gene. LZTFL1 (leucine zipper transcription factor-like 1) gene is an airway cilia regulator and has been identified as a candidate effector gene at a COVID-19 risk locus (5).

Based on several previous studies, we analyzed the single nucleotide polymorphism (SNP) rs11385942 in the LZTFL1 gene using the TRANSFAC Professional database. rs11385942 has been identified in a recent genome-wide association study as one of the most important genetically predisposed risk factors of COVID-19 infection (6). For comparative analysis, we also did a similar analysis using the JASPAR database.

Our aim was to identify transcription factors governing the regulation of the LZTFL1 gene. It is important to note that when searching for transcription factors regulating the genes of your interest, it is of crucial importance which collection of position weight matrices (PWMs or DNA-binding models) is going to be used for identifying potential transcription factor binding sites.

As of 2022, TRANSFAC professional contains 10,297 position weight matrices (PWMs) covering vertebrates, plants, insects, and fungi. JASPAR contains 2430 position frequency matrices (PFMs) covering vertebrates, plants, insects, nematodes, and fungi.

In the TRANSFAC database, positional weight matrices (PWMs) are built by collecting and aligning known binding sites for a particular transcription factor and then counting the nucleotides in each position, and then with this PWM, a score is calculated using the program MATCH^{TM} (6) for predicting potential new binding sites in any regulatory region of the gene.

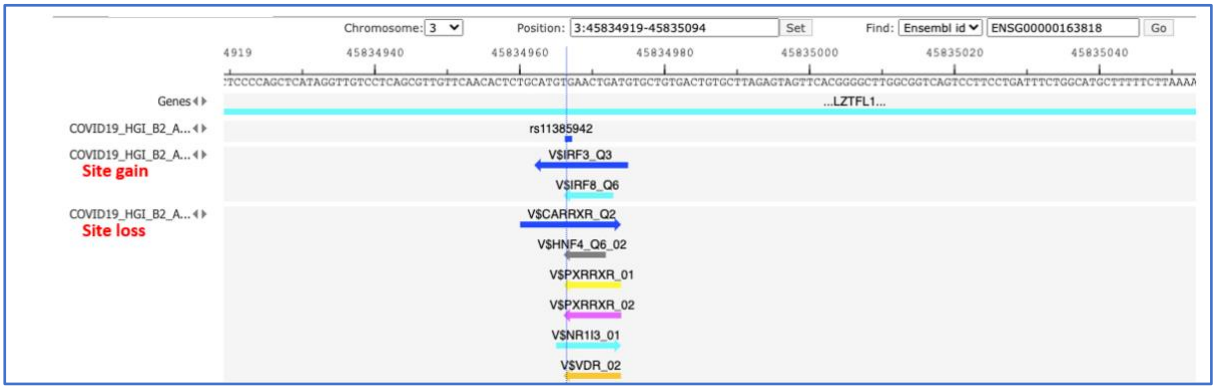
If there is an SNP in such a predicted binding site, its score either drops or increases leading to either a gain or loss of the binding site.

This information on SNPs was loaded from the coronavirus database (covid19hg.org) into the geneXplain platform to analyze them.

We used the full library of vertebrate matrices from the TRANSFAC Professional database (profile: vertebrate_human_p0.001) to calculate TFBS on the SNPs. Additionally, we have filtered the sites by the top core score to achieve highly significant results. When we observed regions around the SNP of our focus, rs11385942, we found that several TFs have gained or lost their sites due to nucleotide exchange. Two very important TFs VDR (Vitamin D receptor) and IRF-3 (interferon regulatory factor 3) are identified using

TRANSFAC Professional. Here it is very important to note that several studies have proved the role of IRF-3 in covid infection.

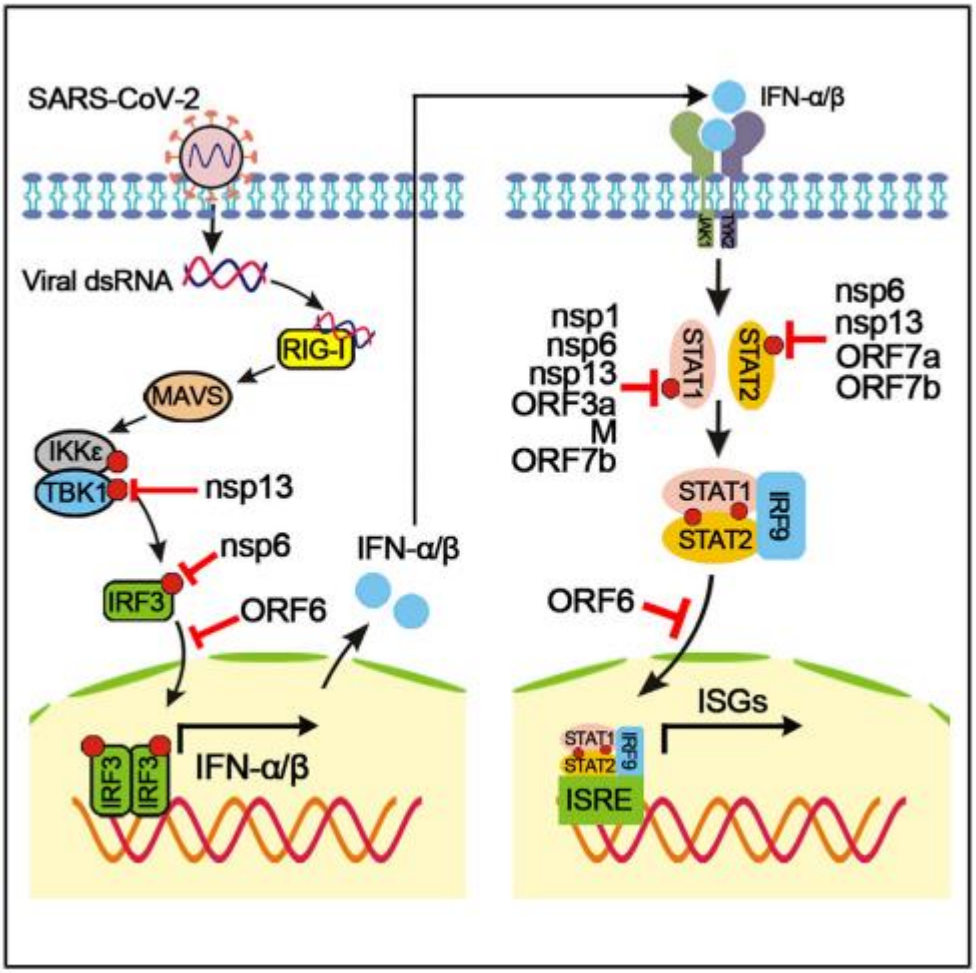
A study by Fretias et al. (7) has proved that the SARS-CoV-2 Spike protein potentiates proteasomal degradation of IRF-3. This may implicate a novel mechanism by which SARS-CoV-2 evades the host's innate immune response to facilitate COVID-19 pathogenesis.



IRF-3 correlates to many diseases, specifically coronavirus infections, as also curated and captured in the TRANSFAC database.

Xia et al. (8) also proved that SARS-CoV-2 proteins antagonize type I interferon (IFN-I) response by suppressing IRF-3 phosphorylation.

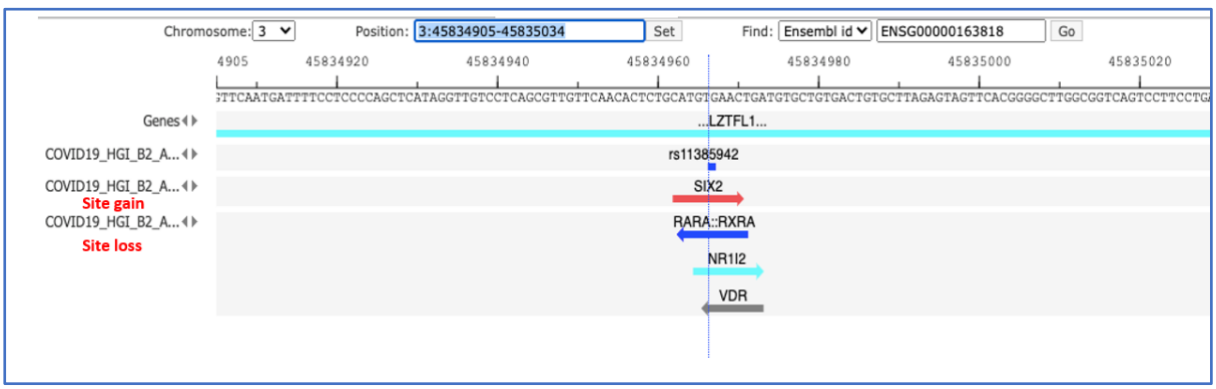
Based on these publications we can infer that IRF-3 identified in our analysis is critical for the infection with SARS-CoV-2.



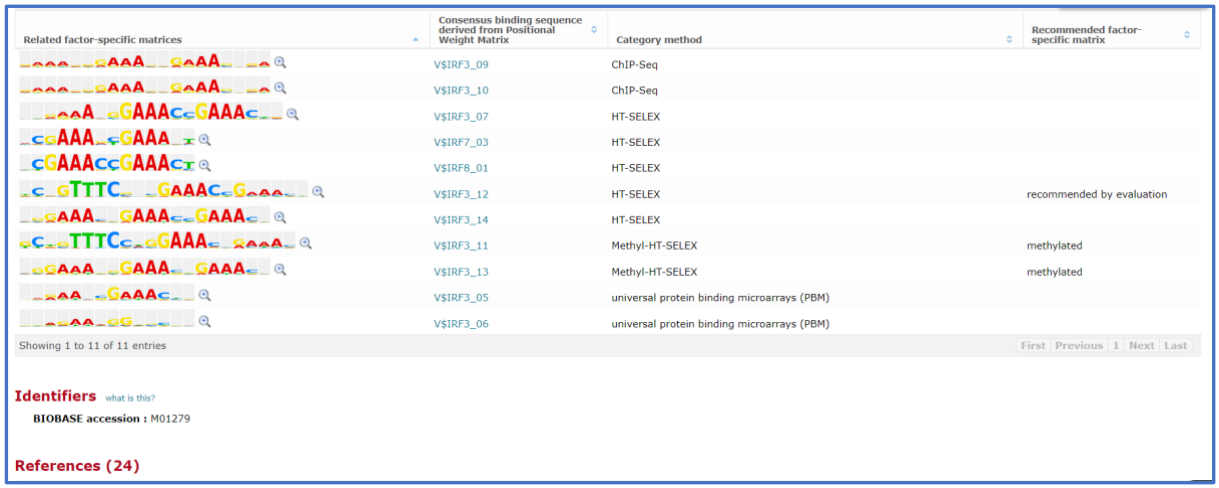
Pic reference: Xia-et-al (8)

Similarly, several publications that prove the role of vitamin D receptor in coronavirus infection severity (9), and our analysis indeed identified the VDR transcription factor in the studied SNP.

For comparison, we loaded the JASPAR matrix library into the MATCH tool and performed site search using the same parameters. The gained site of the VDR transcription factor was successfully identified, but the IRF-3 site was not found.



To summarize, via our analysis we further propose the role of IRF-3 and VDR in COVID-19 infections. With the help of TRANSFAC Professional but not with JASPAR we were able to identify a potentially crucial IRF-3 binding site. The reason is that in the JASPAR database, there was only one IRF-3 matrix, not well matching the site in the studied region, while TRANSFAC Professional had 11 IRF-3 matrices. One of these (M01279) was built by our team by using information from binding sites from SARS-CoV-2-infected cells.



Related factor-specific matrices	Consensus binding sequence derived from Positional Weight Matrix	Category method	Recommended factor-specific matrix
	V$IRF3_09	ChIP-Seq	
	V$IRF3_10	ChIP-Seq	
	V$IRF3_07	HT-SELEX	
	V$IRF7_03	HT-SELEX	
	V$IRF8_01	HT-SELEX	
	V$IRF3_12	HT-SELEX	recommended by evaluation
	V$IRF3_14	HT-SELEX	
	V$IRF3_11	Methyl-HT-SELEX	methylated
	V$IRF3_13	Methyl-HT-SELEX	methylated
	V$IRF3_05	universal protein binding microarrays (PBM)	
	V$IRF3_06	universal protein binding microarrays (PBM)	

Showing 1 to 11 of 11 entries First Previous 1 Next Last

Identifiers what is this?

BIOBASE accession : M01279

References (24)

Note: Subscription to TRANSFAC® Professional provides full access to regularly updated content that is unique in the breadth and depth of its content. It also provides customized pipelines that are designed to not only identify transcription factor binding sites but also do upstream analysis (integrated promoter and pathway analysis) with one click.

References:

1) https://covid19.who.int

2) Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. Nature 2020 Nov;587(7835):610-612. doi: 10.1038/s41586-020-2818-3. Epub 2020 Sep 30.

3) COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet Actions. 2020 Jun;28(6):715-718. doi: 10.1038/s41431-020-0636-6. Epub 2020 May 13.

4) Fink-Baldauf IM, Stuart WD, Brewington JJ, Guo M, Maeda Y. CRISPRi links COVID-19 GWAS loci to LZTFL1 and RAVER1. EBioMedicine. 2022 Jan;75:103806. doi: 10.1016/j.ebiom.2021.103806. Epub 2022 Jan 6. PMID: 34998241; PMCID: PMC8731227.

5) Downes DJ, Cross AR, Hua P, Roberts N, Schwessinger R, Cutler AJ, Munis AM, Brown J, Mielczarek O, de Andrea CE, Melero I; COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium, Gill DR, Hyde SC, Knight JC, Todd JA, Sansom SN, Issa F, Davies JOJ, Hughes JR. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. Nat Genet. 2021 Nov;53(11):1606-1615. doi: 10.1038/s41588-021-00955-3. Epub 2021 Nov 4. PMID: 34737427; PMCID: PMC7611960.

6) Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. 2003 Jul 1;31(13):3576-9. doi: 10.1093/nar/gkg585. PMID: 12824369; PMCID: PMC169193.

7) Freitas RS, Crum TF, Parvatiyar K. SARS-CoV-2 Spike Antagonizes Innate Antiviral Immunity by Targeting Interferon Regulatory Factor 3. Front Cell Infect Microbiol. 2022 Jan 10;11:789462. doi: 10.3389/fcimb.2021.789462. PMID: 35083167; PMCID: PMC8785962.

8) Xia H, Cao Z, Xie X, Zhang X, Chen JY, Wang H, Menachery VD, Rajsbaum R, Shi PY. Evasion of Type I Interferon by SARS-CoV-2. Cell Rep. 2020 Oct 6;33(1):108234. doi: 10.1016/j.celrep.2020.108234. Epub 2020 Sep 19. PMID: 32979938; PMCID: PMC7501843.

9) Wang Z, Joshi A, Leopold K, Jackson S, Christensen S, Nayfeh T, Mohammed K, Creo A, Tebben P, Kumar S. Association of vitamin D deficiency with COVID-19 infection severity: Systematic review and meta-analysis. Clin Endocrinol (Oxf). 2022 Mar;96(3):281-287. doi: 10.1111/cen.14540. Epub 2021 Jul 12. PMID: 34160843; PMCID: PMC8444883.