

SUSD1: A CANDIDATE GENE FOR COVID-19 SEVERITY

Arturo Tozzi

Center for Nonlinear Science, University of North Texas
1155 Union Circle, #311427, Denton, TX 76203-5017 USA
tozziarturo@libero.it
Arturo.Tozzi@unt.edu

SARS-CoV and SARS-CoV-2 cause a storm of proinflammatory cytokines, leading to a severe acute respiratory syndrome with case-fatality rate of 2-3% (Chen and Li, 2020; Li et al., 2020). One of the most promising candidate to trigger the abnormal immunological host response is the SARS-CoV open reading frame 3a (ORF3a). The ORF3a gene encodes an accessory protein that, expressed in infected cells and localized at the endoplasmic reticulum/Golgi intermediate compartment, forms multimeric complexes acting as ion channels and regulating virus release (Ding et al., 2017). It has been demonstrated that ORF3a activates the NLRP3 inflammasome, by promoting TNF receptor-associated factor 3-mediated ubiquitination of apoptosis-associated speck-like protein containing a caspase recruitment domain (Siu et al., 2019). Further, the ORF3a protein is a strong activator of pro-IL-1 β gene transcription and protein maturation, through NF- κ B. Apart from proinflammatory effects which cause life-threatening forms of COVID-19, the ORF3a protein acts as a calcium-binding viroporin that regulates virus production and induction of host cell apoptosis (Yount et al., 2005; Minakshi et al., 2014). These results, summarized by Zhang et al. (2014), suggest the importance of ORF3a for the life cycle of Coronaviruses. It is noteworthy that the ORF3a gene displays rapid accumulation of non-synonymous mutations, with a possible impact on B-cell like epitope formation (Issa et al., 2020; Wang et al., 2020). Mutations involving SARS-CoV-2's ORF3a appear to be spreading worldwide, which deserves close attention.

Using Uniprot (<https://www.uniprot.org/blast/uniprot/B20200405A94466D2655679D1FD8953E075198DA804AEFC0>), we performed a Blast alignment of the 274 amino acids of the Uncharacterized protein OS=Human SARS coronavirus OX=694009 GN=ORF3a. We found a partial alignment (E-value: 9.8e0, Score: 66, Ident.: 25.3%) with the SUSD1_HUMAN - Sushi domain-containing protein 1 Homo sapiens (Human). In particular, we found alignment with 123-182 SUSD1, corresponding to a Calcium binding EGF-like domain, and with 192-212 SUSD1, corresponding to a SCR repeat zone.

SUSD1 protein coding gene is a calcium ion binding integral component of membrane, classified among protein containing domains EGF_3, EGF_CA, and CCP (Marchler-Bauer et al., 2017). The SUSD1 gene (location 9q31.3-q32) has ubiquitous expression in placenta (RPKM 5.9), urinary bladder (RPKM 5.8) and many other tissues (see: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SUSD1>). In particular, the SUSD1 protein contains a number of Complement control protein modules, or short consensus repeats, also known as Sushi domains. These domains are known to exist in a wide variety of complement and adhesion proteins, suggesting that SUSD1 may play a role in complement system of the immune system (<https://www.prosci-inc.com/susd1-antibody-8609.html>). It is noteworthy that the mRNA differential expression in normal tissues according to GTEx suggests that SUSD1 Gene is overexpressed in Whole Blood (x5.6), while the protein differential expression in normal tissues from HIPED suggests that SUSD1 Gene is overexpressed in Peripheral blood mononuclear cells (48.1) and Platelet (20.9) (<http://amp.pharm.mssm.edu/Harmonizome/gene/SUSD1>).

Data from virus perturbations changing expression of SUSD1 gene from the GEO Signatures of Differentially Expressed Genes for Viral Infections dataset (<http://amp.pharm.mssm.edu/Harmonizome/gene/SUSD1>) suggest that SUSD1 displays eight increased expression associations, including several SARS viruses:

HIV_Infected-mDC_None_GSE42058 [2.42027]; SARS-BatSRBD_Day1_None_GSE50000 [2.1444]; A-CA-04-2009(H1N1)_24Hour_23935999_GSE47962 [1.70559]; A-Vietnam-1203_CIP048_RG4-2004(H5N1)NS1trunc124_1day-MOI-10³_None_GSE44445 [1.62128]; A-CA-04-2009(H1N1)_48Hour_None_GSE37571 [1.494]; icSARS CoV_72Hour_None_GSE37827 [1.45884]; SARS-CoV MA15_Day1-PFU-10⁵_None_GSE50000 [1.40423]; SARS-dORF6_24Hour_23935999_GSE47962 [1.20463]

Furthermore, SUSD1 has been associated with several human diseases, including pathologies somewhat correlated with COVID-19: venous thromboembolism, childhood obesity, arthritis severity (Brenner et al., 2013; Tang et al., 2013; see also: <https://www.ncbi.nlm.nih.gov/protein/544063434>).

In sum, we hypothesize that the increased expression of the SUSD1 host gene caused by Coronavirus infection may contribute to the lower respiratory tract cytokine storm that makes severe acute respiratory syndrome a life-threatening

disease. Also, we speculate that low-symptomatic or asymptomatic COVID-19-positive adults and the relatively spared pediatric population might display a lower expression of SUSD1.

REFERENCES

1. Brenner M, Laragione T, Gulko PS. 2013. Arthritis severity locus *Cia4* is an early regulator of IL-6, IL-1 β , and NF- κ B activators' expression in pristane-induced arthritis. *Physiol Genomics*. 2013 Jul 2;45(13):552-64. doi: 10.1152/physiolgenomics.00029.2013..
2. Ding Q, Heller B, Capuccino JM, Song B, Nimgaonkar I, et al. 2017. Hepatitis E virus ORF3 is a functional ion channel required for release of infectious particles. *Proc Natl Acad Sci U S A*. 2017 Jan 31;114(5):1147-1152. doi: 10.1073/pnas.1614955114.
3. Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. 2020. SARS-CoV-2 and ORF3a: Non-Synonymous Mutations and Polyproline Regions. doi: <https://doi.org/10.1101/2020.03.27.012013>.
4. Li X, Geng M, Peng Y, Meng L, Lu S. 2020. Molecular immune pathogenesis and diagnosis of COVID-19. *Journal of Pharmaceutical Analysis*. <https://doi.org/10.1016/j.jpha.2020.03.001>.
5. Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", *Nucleic Acids Res*.45(D)200-3Minakshi R, Padhan K, Rehman S, Hassan MI, Ahmad F. 2014. The SARS Coronavirus 3a protein binds calcium in its cytoplasmic domain. *Virus Res*. 2014 Oct 13;191:180-3. doi: 10.1016/j.virusres.2014.08.001.
6. Siu KL, Yuen KS, Castaño-Rodríguez C, Ye ZW, Yeung ML, et al. 2019. Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *FASEB J*. 2019 Aug;33(8):8865-8877. doi: 10.1096/fj.201802418R.
7. Tang W, Teichert M, Chasman DI, Heit JA, Morange PE, et al. 2013. A genome-wide association study for venous thromboembolism: the extended cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Genet Epidemiol*. 2013 Jul;37(5):512-521. doi: 10.1002/gepi.21731.
8. Yount B, Roberts RS, Sims AC, Deming D, Frieman MB, et al. 2005. Severe Acute Respiratory Syndrome Coronavirus Group-Specific Open Reading Frames Encode Nonessential Functions for Replication in Cell Cultures and Mice. *J Virol*. 2005 Dec; 79(23): 14909–14922. doi: 10.1128/JVI.79.23.14909-14922.2005
9. Wang M, Li M, Ren R, Brave A, van der Werf S, et al. 2020. International expansion of a novel SARS-CoV-2 mutant. doi: <https://doi.org/10.1101/2020.03.15.20035204>.
10. Zhang R, Wang K, Lv W, Yu W, Xie S, et al. The ORF4a protein of human coronavirus 229E functions as a viroporin that regulates viral production. *Biochim Biophys Acta*. 2014 Apr;1838(4):1088-95. doi: 10.1016/j.bbame.2013.07.025.

SUPPLEMENTARY INFORMATION

Examining the Uncharacterized protein OS=Human SARS coronavirus OX=694009 GN=ORF3a PE=4 SV=1, with the protein sequence:

```
MDLFMRFFTLXSITAQPVKIDNASXASTVHATATIPLQASLPGWLIVIGVAFLAVFQSAT
KIIALNKRWQLALYKGFQFICNLLLLFVTIYSHLLLVAAGMEAQFLYLYALIYFLQCINA
CRIIMRCWLCWKCKSKNPLLYDANYFVCWHTHNYDYCIPYNSVTDITIVVTEGDGISTPKL
KEDYQIGGYSEDRHSGVKDYVVVHGYFTEVYYQLESTQITTDGTIENATFFIFNKLVDKP
PNVQIHTIDGSSGVANPAMDPIYDEPTTTTTSVPL,
```

Selected alignment(s) from match Q6UWL2-3

Q6UWL2-3 SUSD1_HUMAN - Isoform 3 of Sushi domain-containing protein 1 Homo sapiens (Human)

E-value: 9.8e0
 Score: 66
 Ident: 25.3%
 Positives : 41.8%
 Query Length: 274
 Match Length: 747

J9TEM7	J9TEM7_CVHSA	121	CRIIMRCWLCWKCKSKNPLLYDANYFVCWHTHNY---DYCIPYNSVTDITIVVTEGDGIST	177
			C I C + C+ + F C+ Y + P++ TD TE D +	
Q6UWL2-3	SUSD1_HUMAN	123	CTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTP	182
J9TEM7	J9TEM7_CVHSA	178	PKLKEDYQIGGYSEDRHSGVKDYVVVHGYFT	208
			P++ + Y IG Y+ S V+ Y G+F+	
Q6UWL2-3	SUSD1_HUMAN	183	PEVDPGYIIGNYNTSSILGSQVR-YACREGFFS	212

Architecture

aa



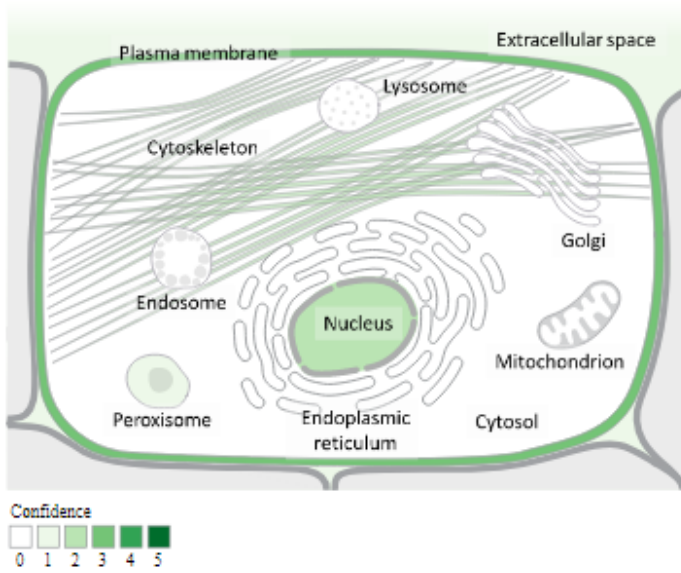
747



NP_071931.2 747 aa ✕

	42 - 69 aa	pfam12947	EGF_3: EGF domain
	73 - 111 aa	smart00179	EGF_CA: Calcium-binding EGF-like domain
	125 - 156 aa	smart00179	EGF_CA: Calcium-binding EGF-like domain
	179 - 234 aa	pfam00084	Sushi: Sushi repeat (SCR repeat)
	239 - 294 aa	cd00033	CCP: Complement control protein (CCP) modules (aka short consensus repeats SCRs or SUSHI repeats) have been identified in several proteins of the complement system

Subcellular locations from COMPARTMENTS ?



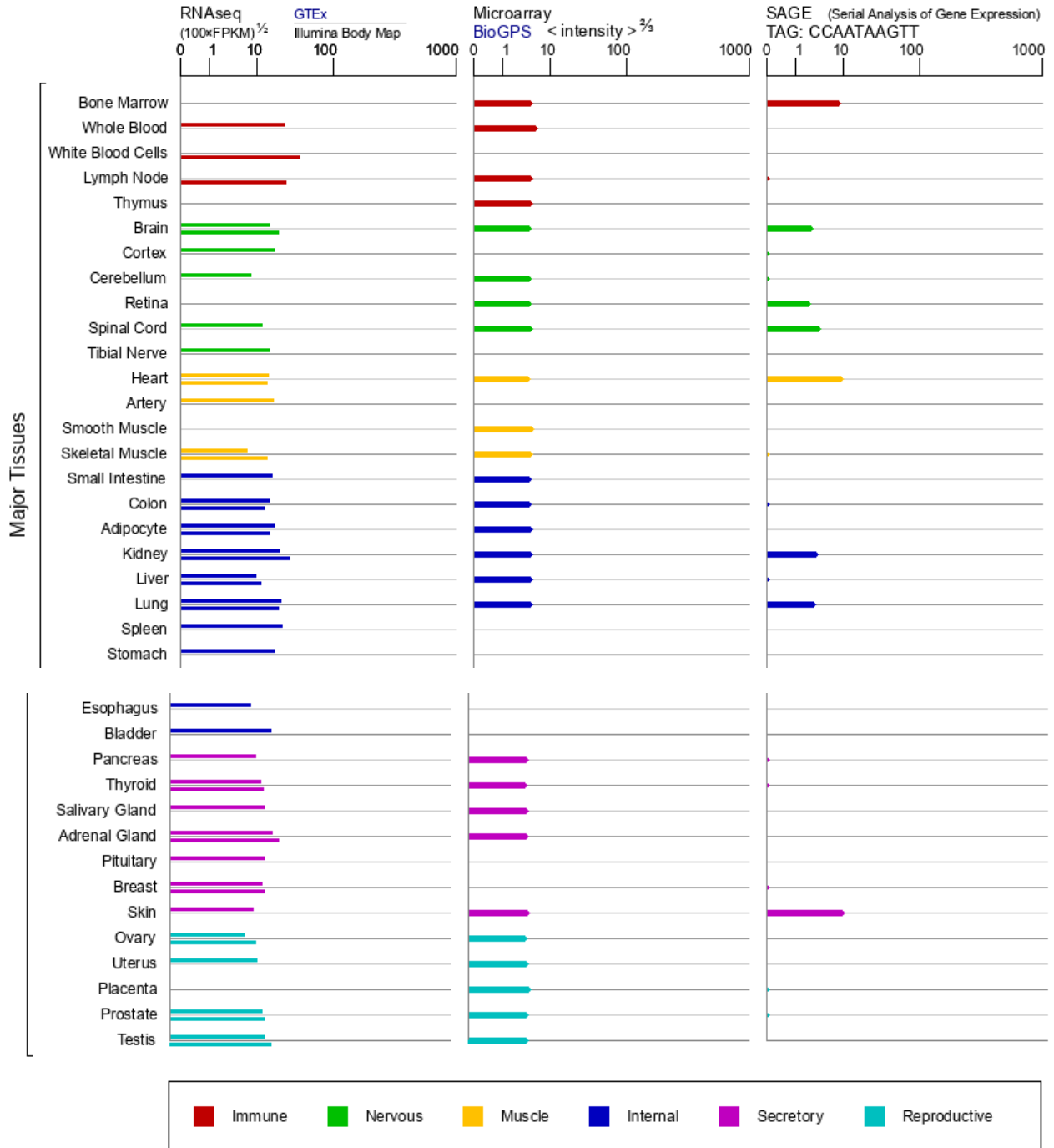
Compartment	Confidence
plasma membrane	3
cytoskeleton	2
nucleus	2
extracellular	1
peroxisome	1
mitochondrion	0
endoplasmic reticulum	0
golgi apparatus	0

(<https://www.genecards.org/cgi-bin/carddisp.pl?gene=SUSD1>)

Expression for SUSD1 Gene

Products: Primer / Gene Expression Assay

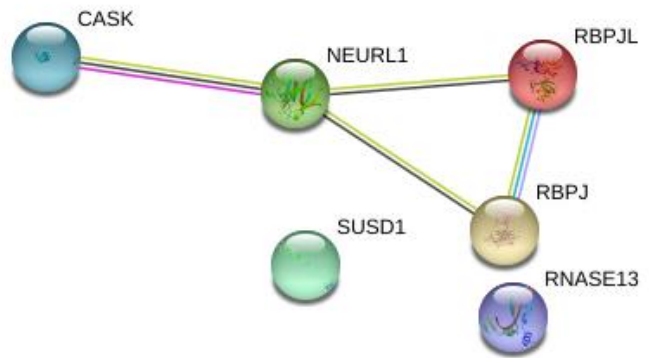
mRNA expression in normal human tissues from [GTEx](#), [Illumina](#), [BioGPS](#), and [SAGE](#) for SUSD1 Gene ?



(<https://www.genecards.org/cgi-bin/carddisp.pl?gene=SUSD1>)

Interacting Proteins for SUSD1 Gene

STRING Interaction Network Preview (showing top 5 STRING interactants - click image to see top 25)



(<https://www.genecards.org/cgi-bin/carddisp.pl?gene=SUSD1>)