

SARS-CoV-2 *orf1b* Gene Sequence in the *NTNG1* Gene on Human Chromosome 1

STEVEN LEHRER and PETER H. RHEINSTEIN

¹Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, U.S.A.;
²Severn Health Solutions, Severna Park, MD, U.S.A.

Abstract. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense single-stranded RNA virus. It is contagious in humans and is the cause of the coronavirus disease 2019 (COVID-19) pandemic. In the current analysis, we searched for SARS-CoV-2 sequences within the human genome. To compare the SARS-CoV-2 genome to the human genome, we used the blast-like alignment tool (BLAT) of the University of California, Santa Cruz Genome Browser. BLAT can align a user sequence of 25 bases or more to the genome. BLAT search results revealed a 117-base pair SARS-CoV-2 sequence in the human genome with 94.6% identity. The sequence was in chromosome 1p within an intronic region of the netrin G1 (*NTNG1*) gene. The sequence matched a sequence in the SARS-CoV-2 *orf1b* (open reading frames) gene. The SARS-CoV-2 human sequence lies within non-structural proteins 14 and 15 (NSP14 and NSP15), and is quite close to the viral spike sequence, separated only by NSP16, a 904-base pair sequence. The mechanism for SARS-CoV-2 infection is the binding of the virus spike protein to the membrane-bound form of angiotensin-converting enzyme 2 and internalization of the complex by the host cell. It is probably no accident that a sequence from the SARS-CoV-2 *orf1b* gene is found in the human *NTNG1* gene, implicated in schizophrenia, and that haloperidol, used to treat schizophrenia, may also be a treatment for COVID-19. We suggest, therefore, that it is important to investigate other haloperidol analogs. Among them are benperidol, bromperidol, bromperidol decanoate, droperidol, seiperidol hydrochloride, and trifluperidol. These analogs might be valuable in the treatment of COVID-19 and other coronavirus infections.

This article is freely accessible online.

Correspondence to: Dr. Steven Lehrer, Box 1236 Radiation Oncology, Mount Sinai Medical Center, 1 Gustave L. Levy Place, New York 10029, NY, USA. E-mail: steven.lehrer@mssm.edu

Key Words: COVID-19, *orf1b* gene, *NTNG1* gene, haloperidol, the UCSC Genome Browser.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense single-stranded RNA virus (1). The rapid spread of the infection suggests that the virus could have adapted in the past to human hosts. If so, some of its gene sequences might be found in the human genome. To investigate this possibility, we utilized the UCSC Genome Browser, an on-line genome browser at the University of California, Santa Cruz (UCSC) (<https://genome.ucsc.edu>) (2).

To compare the SARS-CoV-2 genome to the human genome, we used the blast-like alignment tool (BLAT) of the UCSC Genome Browser (2). The SARS-CoV-2 sequence we analyzed with BLAT was FASTA Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome, NCBI Reference Sequence: NC_045512.2.

BLAT search results revealed a 117-base pair SARS-CoV-2 sequence in the human genome with 94.6% identity. The sequence was in chromosome 1p within an intronic region of the netrin G1 (*NTNG1*) gene (Figure 1). The sequence matched a sequence in the SARS-CoV-2 *Orf1b* (open reading frames) gene (Table I). Table II shows alignment of the SARS-CoV-2 sequence and Chr1p. Figure 2 shows side by side alignment of matching SARS-CoV-2 sequences and human Chr1p sequences. A BLAST search did not find the human sequence within the SARS-CoV-2 genome.

SARS-CoV strains have two *orf1* genes, *orf1a* and *orf1b*. The 16 *orf1ab* non-structural proteins (NSPs) are directly involved in viral replication. Five of the NSPs, NSP12 to NSP16, are on *orf1b* (3).

The SARS-CoV-2 human sequence lies within NSP14 and NSP15. As NSP12 duplicates the coronavirus genome, it sometimes adds a wrong 'letter' to the new copy. NSP14 cuts out these errors, so that the correct 'letter' can be added instead. NSP15 protein degrades residual virus RNA to evade the infected cell's antiviral defenses.

The SARS-CoV-2 human sequence is quite close to the viral spike sequence, separated only by NSP16, a 904-base pair sequence (Figure 3). The mechanism for SARS-CoV-2 infection is the binding of the virus spike protein to the membrane-bound form of angiotensin-converting enzyme 2 and internalization of the complex by the host cell (4).

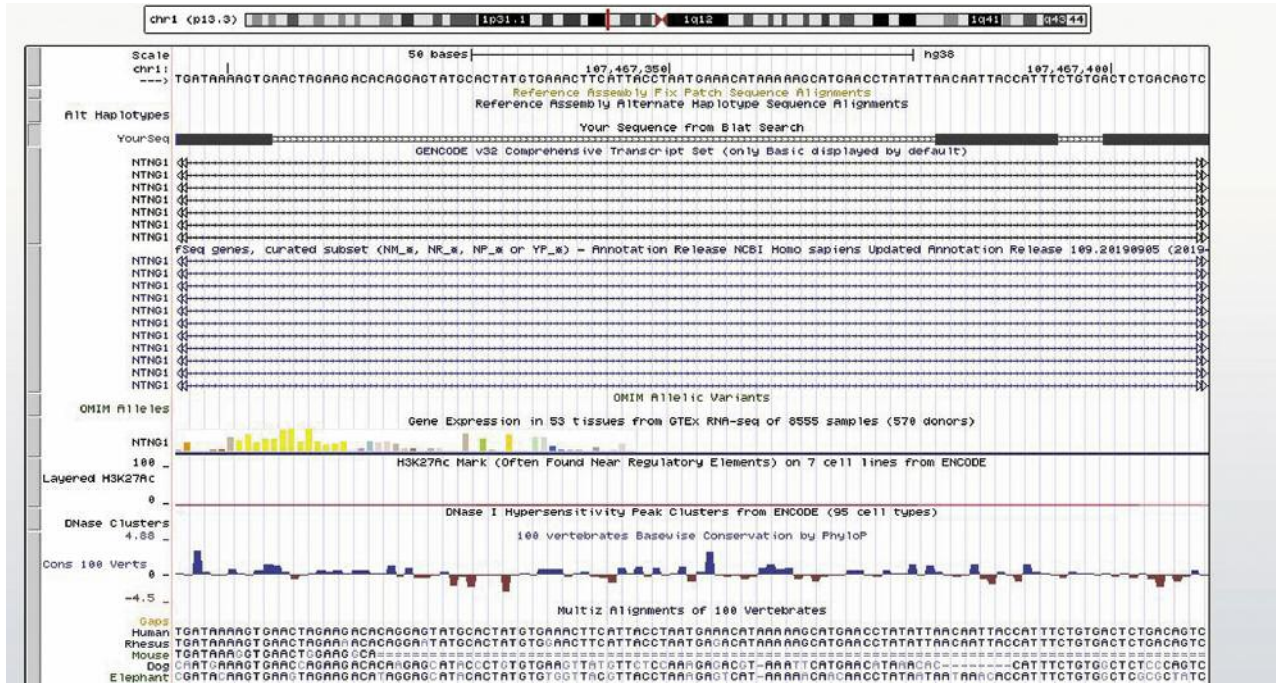


Figure 1. The 117-bp *orf1b* (your seq) gene sequence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in an intronic region of the human *netrin G1* (*NTNG1*) gene on chromosome 1p is shown in the University of California Santa Cruz genome browser.

SARS-CoV-2 sequences were present in chimpanzee (not shown in Figure 1). A 26-base sequence was on Chr3 with 100% identity; a 31-base sequence was present on Chr1 with 96.3% identity; and a 24-base sequence was present on Chr10 with 83.4% identity. No SARS-CoV-2 sequences were present in orangutan or bonobo.

Hu *et al.* analyzed the genomes of 15 SARS-CoV strains. Regardless of different host bat species, SARS-CoV and SARS-related-CoVs detected in bats from southwestern China (Yunnan, Guizhou and Guangxi province) formed one clade, in which SARS-related-CoV strains showing closer relationship to SARS-CoV were all from Yunnan. SARS-related-CoVs detected in southeastern, central and northern provinces, such as Hong Kong, Hubei and Shaanxi, formed the other clade, which was phylogenetically distant from human and civet SARS-CoVs (3).

Human *NTNG1* encodes a pre-pro-protein that is processed into a secreted protein containing eukaryotic growth factor-like domains. This protein acts to guide axon growth during neuronal development. Polymorphisms in this gene may be associated with schizophrenia (4).

Eight percent of DNA in the human genome comes from human endogenous retroviruses (HERV), and some human diseases have been attributed to this DNA. HERV sequences have occasionally been adapted by the human body to serve

a useful purpose, such as in the placenta, where they may safeguard fetal–maternal tolerance (5). However, SARS-CoV-2 is not a retrovirus.

Gordon *et al.* identified 332 high confidence SARS-CoV-2-human protein–protein interactions. Among these, they found 66 druggable human proteins or host factors targeted by 69 existing US Food and Drug Administration-approved drugs, drugs in clinical trials or preclinical compounds.

Of special interest are drugs interacting with the sigma receptors. Both sigma-1 and sigma-2 are human transmembrane receptors that act together with many non-polar, cationic drugs. Gordon *et al.* prioritized several of these drugs based on potency or potential disease relevance, including the antipsychotic Haloperidol, which binds to both sigma-1 and sigma-2 receptors (6). Sigma-1 has been implicated in schizophrenia, clinical depression, and the effects of cocaine abuse. Chloroquine, which is currently in clinical trials for SARS-CoV-2 acts with the sigma-1 receptor and sigma-2 receptor (7).

It is probably no accident that a sequence from the SARS-CoV-2 *orf1b* gene is found in the human *NTNG1* gene, implicated in schizophrenia, and that haloperidol, used to treat schizophrenia, may also be a treatment for COVID-19. We suggest, therefore, that it is important to investigate haloperidol analogs. Among them are benperidol,

Side by side alignment

```

000019332 tgataaaagt 000019342
>>>>>>> ||||| ||||| >>>>>>>
107467295 tgataaaagt 107467305

```

```

000019358 aacaattaccattt 000019371
>>>>>>> ||||| ||||| >>>>>>>
107467381 aacaattaccattt 107467394

```

```

000019379 actctgacagtc 000019390
>>>>>>> ||||| ||||| >>>>>>>
107467400 actctgacagtc 107467411

```

Figure 2. Side by side alignment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (upper sequences) and human chromosome 1p (lower sequences).

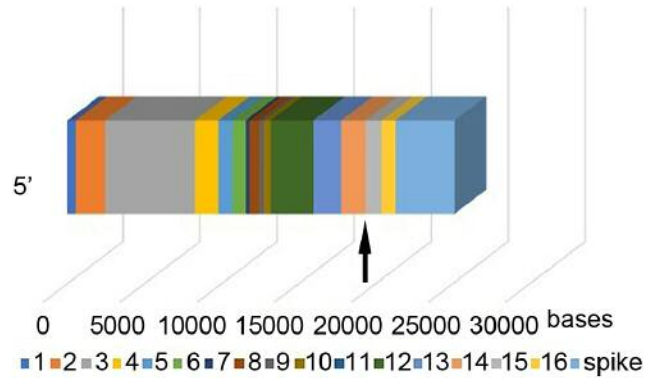


Figure 3. *Orf1ab* genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), showing the 16 non-structural proteins (NSPs) and the viral spike. The human sequence is within NSP 14 and NSP 15 (arrow). The human sequence is separated from the spike by NSP16, a small sequence of 904 bases. The mechanism for SARS-CoV-2 infection is the binding of the virus spike protein to the membrane-bound form of angiotensin-converting enzyme 2 and internalization of the complex by the host cell.

Table I. Alignment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome and human chromosome 1p. Bases 19300 to 19400 are within the SARS-CoV-2 *orf1b* gene. Matching bases in cDNA and genomic sequences are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence (often splice sites).

taaataaaca	tgcattccac	acaccagctt	tTGATAAAAG	TGcttttgtt	19350
aatttaaAAC	AATTACCATT	TttctattAC	TCTGACAGTC	catgtgagtc	19400

Table II. Alignment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome and human chromosome 1p. Bases 107467294 to 107467444 are within the human *netrin G1* (*NTNG1*) gene. Matching bases in cDNA and genomic sequences are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence (often splice sites).

TGATAAAAGT	Gaactagaag	acacaggagt	atgcactatg	tgaaactca	107467344
ttacctaag	aacataaaa	agcatgaacc	tatattAACA	ATTACCATTT	107467394
ctgtgACTCT	GACAGTCtat	cttttattg	aggacttga	gcagtgccgt	107467444

bromperidol, bromperidol decanoate, droperidol, seiperidol hydrochloride, and trifluoperidol. Benperidol was developed by Janssen and is marketed in Europe as Anquil by Kyowa Kirin. Testing immediate release bromperidol is likely to be more productive than testing bromperidol decanoate. Generally, the purpose of a decanoate salt is to facilitate formulation of a long-acting injection. Droperidol was marketed as Inapsine[®] and, in combination with fentanyl, as Innovar[®] by Janssen. It is the same as benperidol without the tetrahydropyridine ring. One or more of these analogs might be valuable in the treatment of COVID-19 and other coronavirus infections.

A weakness in our study is that the 117-base pair sequence of SARS-CoV-2 was discontinuously present in *NTNG1*. Hence, such a matching result is not proof of a genetic relationship between SARS-CoV-2 and the human genome, nor with the short sequences (24-31 bp) of SARS-CoV-2 found in chimpanzee. It would be worthwhile to examine the proteins to verify the similarity between SARS-CoV-2 *orf1b* and the *NTNG1* gene. Further studies are warranted.

Conflicts of Interest

There were no conflicts of interest.

Authors' Contributions

Dr. Lehrer and Dr. Rheinstein contributed equally to the conception, data analysis and writing.

References

- 1 Khan S, Siddique R, Shereen MA, Ali A, Liu J, Bai Q, Bashir N and Xue M: The emergence of a novel coronavirus (SARS-CoV-2), their biology and therapeutic options. *J Clin Microbiol*, 2020. PMID: 32161092. DOI: 10.1128/JCM.00187-20
- 2 Kuhn RM, Haussler D and Kent WJ: The UCSC genome browser and associated tools. *Brief Bioinform* 14(2): 144-161, 2013. PMID: 22908213. DOI: 10.1093/bib/bbs038
- 3 Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J and Shi ZL: Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog* 13(11): e1006698, 2017. PMID: 29190287. DOI: 10.1371/journal.ppat.1006698
- 4 Wilcox JA and Quadri S: Replication of ntng1 association in schizophrenia. *Psychiatr Genet* 24(6): 266-268, 2014. PMID: 25325217. DOI: 10.1097/YPG.0000000000000061
- 5 Kurth R and Bannert N: Beneficial and detrimental effects of human endogenous retroviruses. *Int J Cancer* 126(2): 306-314, 2010. PMID: 19795446. DOI: 10.1002/ijc.24902
- 6 Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, O'Meara MJ, Guo JZ, Swaney DL, Tummino TA, Huttenhain R, Kaake R, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco BJ, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Brewer M, Cakir M, McGregor MJ, Li Q, Naing ZZC, Zhou Y, Peng S, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Shen W, Shi Y, Zhang Z, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, Mathy CJP, Perica T, Pilla KB, Ganesan SJ, Saltzberg DJ, Ramachandran R, Liu X, Rosenthal SB, Calviello L, Venkataramanan S, Lin Y, Wankowicz SA, Bohn M, Trenker R, Young JM, Cavero D, Hiatt J, Roth T, Rathore U, Subramanian A, Noack J, Hubert M, Roesch F, Vallet T, Meyer B, White KM, Miorin L, Agard D, Emerman M, Ruggiero D, Garc, amp, iacute-Sastre A, Jura N, von Zastrow M, Taunton J, Schwartz O, Vignuzzi M, d'Enfert C, Mukherjee S, Jacobson M, Malik HS, Fujimori DG, Ideker T, Craik CS, Floor S, Fraser JS, Gross J, Sali A, Kortemme T, Beltrao P, Shokat K, Shoichet BK and Krogan NJ: A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. *bioRxiv*: 2020.2003.2022.002386, 2020. DOI: 10.1101/2020.03.22.002386
- 7 Maurice T and Su TP: The pharmacology of sigma-1 receptors. *Pharmacol Ther* 124(2): 195-206, 2009. PMID: 19619582. DOI: 10.1016/j.pharmthera.2009.07.001

Received April 14, 2020

Revised April 20, 2020

Accepted April 22, 2020