

# The Likely Role of de-Mono-ADP-ribosylation of STAT1 by the SARS-CoV-2 nsp3 Protein in the Cytokine Storm Syndrome of COVID-19

Jean-Michel Claverie

Aix-Marseille University, CNRS

Structural & Genomic Information Laboratory (IGS, UMR 7256),

Mediterranean Institute of Microbiology (FR3479)

Marseille, France

**Keywords:** de-MARylation; interferon-stimulated gene; inflammation; PARP14; ACE2; coronavirus

## Abstract

As more cases of COVID-19 are studied and treated world-wide, it had become apparent that lethal and most severe cases of pneumonia are due to an out-of-control inflammatory response to the SARS-CoV2 virus. I explored the putative causes of this specific feature through a detailed genomic comparison with the closest SARS-CoV-2 relatives isolated from bats, as well as previous coronavirus strains responsible for the previous epidemics (SARS-CoV, and MERS-CoV). The high variability region of the nsp3 protein was confirmed to exhibit the most variations between closest strains. It was then studied in the context of physiological and molecular data available in the literature. A number of convergent findings point out de-mono-ADP-ribosylation (de-MARylation) of STAT1 by the SARS-CoV-2 nsp3 as a likely cause of the cytokine storm observed in the most severe cases of COVID-19. This may suggest new therapeutic approaches and an assay to predict the virulence of naturally circulating SARS-like animal coronaviruses.

## Introduction

Probably started in December 2019 from a zoonotic transmission from traded wild animals in a Wuhan market (Hubei province, China) (1, 2), the coronavirus disease 2019 (COVID-19) quickly reached a planetary dimension forcing billions of people into lockdown, grinding to a halt most of the world economies, and killing up 247,000 out of the 3.5 million of known infected people (as of May 4<sup>th</sup>). Initially catalogued as a respiratory disease (severe acute respiratory syndrome, SARS), the investigation of severe COVID-19 cases progressively revealed their association with a dysfunctional pro-inflammatory response leading to severe lung damages (3), but also cardiovascular complications (4) and multi-organ failures (5). The unexpected severity and dimension of the COVID-19 pandemics prompted national research organizations to ask their scientists, even without prior knowledge on coronaviruses, to use their expertise to contribute to the solution of this formidable challenge. This work is part of this approach, that of a novice researcher in the field, but competent in viral genomics. In this paper, I present a bioinformatic investigation of publicly available sequence data, in an attempt to identify which molecular process could be responsible of the specific clinical picture of COVID-19. Several coincidental findings examined in the light of the rich pre-existing bibliographical information strongly suggest that virus-induced de-MARylation of STAT1 is the key step leading to the inflammatory cytokine storm seen in the most severe COVID-19 cases.

## Results

As a simple, straightforward attempt to delineate genomic changes that might correlate with the crossing of the xenographic barrier to human, and with the clinical features of COVID-19, we compared the whole predicted proteome of SARS-CoV-2 with its closest animal-infecting relatives available in the NCBI public database. In a preliminary step, a comparison of the proteome sequences of 250 human-infecting SARS-CoV-2 isolates did not reveal any region of significant variability, as they were quasi-identical. For instance, out of 250 nsp3 protein sequences (1945 residues long), 214 were identical, and 36 exhibited a single amino-acid changes (99.95% identity), 11 of which conservative (data not shown). One sequence from the group of 214 identical sequence (Accession # YP\_009725299.1, Wu *et al.*, unpublished) was randomly chosen and used for subsequent comparative analyses

Variable regions became only detectable when comparing the polyprotein 1ab sequence of SARS-CoV-2 with that of its closest relatives in animal hosts. As previously noticed, the most similar isolate is bat coronavirus RaTG13, with 96.7 % similarity (Accession # QHR63299) (1), followed by two other bat isolates (# AVP78030, # AVP78041, 93.5% and 92.5% identical respectively) (6), and several similar pangolin isolates (e.g. # QIQ54046, 85.4% identical) (Cao *et al.*, unpublished).

Looking for amino-acid changes occurring in clusters, my attention was drawn to a segment located near the N-terminus of the nsp3 protein (Fig.1). While 82.3% of its 1945 positions exhibit strictly identical residues between the Human and bat homologues, this segment falls



down to 44.5% identical residues. In SARS-CoV-2 nsp3 this segment is strongly enriched in acidic amino-acids and coincides with a lack of predicted secondary structure elements (Fig.1). On the contrary, it is predicted as an intrinsically disordered protein region. Such domains form the basis of the dynamic “on-off” switch-type of interactions commonly found in signalling networks (reviewed in 7). This stretch of amino-acids roughly maps to the hyper variable region (HVR) previously recognized upon the comparison of nsp3 homologs from more distant coronaviruses (reviewed in 8), and to which no function has been assigned or predicted. I then focused further analyses on this well delineated island of changes within the otherwise highly similar nsp3 from human, bat, and pangolin SARS-CoV-2 closest relatives.

The nsp3 segment delineated as highly variable in the above comparison between the Human Sars-Cov-2 and its closest animal homologs, is also the most divergent upon alignment of the human SARS-CoV-2 with the previous SARS-CoV-1 (responsible of the 2002 epidemics) (9) (Fig2). In particular, a deletion of 25 consecutive residues in the SARS-CoV-1 makes it unlikely that these homologous regions could adopt an identical fold, or perform similar functions, if any. Finally, we also noticed that the 215 N-terminal residues of the SARS-CoV-2 nsp3, entirely encompassing the variable region delineated in Fig1, do not have homologues in the Middle East respiratory syndrome coronavirus (MERS-CoV) nsp3 protein (10) (data not shown).

The above analyses suggested that the N-terminal variable region (approximately from pos. 169 to pos. 402) of the nsp3 protein, being the one bearing the greatest variability between – even human- strains, might be linked to the specific clinical pictures of the associated diseases.

Nsp3 is a large multifunctional protein, several domains of which have been mapped and characterized (reviewed in 10). Starting from the N-terminus there are in succession: an ubiquitin-like domain (20-106) known to bind to single stranded RNA, an ADP-ribose-binding module (240-340), a single-stranded poly(A) binding domain (678-743), a papain like viral protease (746-1064), a nucleic acid-binding domain (1089-1201), and a seven-transmembrane G protein-coupled receptor C (1494-1563).

Interestingly, when the whole nsp3 sequence was compared to the entire human proteome using BlastP, the only reported significant matches ( $E\text{-value} < 10^{-6}$ ) were found to precisely encompass the ADP-ribose –binding module. Moreover, the top hits are human protein with a known function: all of them are mono-ADP-ribosyltransferases. This result strongly suggests that the ADP-ribose binding site predicted in the nsp3 protein is functional, as it has been experimentally verified in a number of other viruses, including SARS-CoV (reviewed in 11). This is the first brick on which I will build my hypothesis: the N terminal part of the SARS-CoV-2 nsp3 protein is constituted of a putative protein-binding domain immediately adjacent to the putative active site of an ADP-ribose (ADPR) processing enzyme.

Further inspections of the above BlastP output provided an exciting guess as to the putative target of the above ADPR enzyme. All the top matches of the nsp3 putative ADPR-binding site corresponded to that of various isoforms or variants of the PARP14 protein (Fig3). It turns out

HUMAN_COV2	APTK-VTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVAD	59
HUMAN_COV1	APIKGVTFGEDTVWEVQGYKNVRITFELDERVDKVLNEKCSVYTVESGTEVTEFACVVAE ** * ****:*** *****.*****.*****.*****.*****.*****:	
HUMAN_COV2	AVIKTLQPVSELLTPLGIDLDEWSMATYYLFDSEGEFKLASHMYCSFYPPDEDEE-EGDC	118
HUMAN_COV1	AVVKTLPVSDLLTNMGIDLDEWSVATFYLFDDAGEENFSSRMYSFYPPDEEEEDDAEC **:*****:*** :*****:***:***:*** :*:***:*****:*** :*:*	
HUMAN_COV2	EEEEFEPSTQYEGTEDDYQGKPLEFGATSAALQPEEQEEDWLDDDSQOTVGGQDGS	178
HUMAN_COV1	EEEEIDETCEHEYGTEDDYQGLPLEFGASAE TVRVEEEEEEDWLDDTTEQS----- ****: : : :***** *****: : : ***** :*:	
HUMAN_COV2	NQTTTIQTIVEVQPQLEMELTPVQVQTIENVNSFGYLKLTDNVYIKNADIVEEAKVKPTV	238
HUMAN_COV1	-----EIEP--EPEPTP---EFPVNFQFTGYLKLTDNVAIKCVDIVKEAQSANPMV *:* * * * **.*:***** ** .***:*.:.:* *	
HUMAN_COV2	VVNAANVYLKHGGGVAGALNKATNNAMQVESDDYIATNGPLKVGGSVLSGHNLAKHCLH	298
HUMAN_COV1	IVNAANIHLKHGGGVAGALNKATNGAMQKESDDYIKLNGPLTVGGSCLLSGHNLAKKCLH :*****:*****.*** ***** *****.*****:*****:***	
HUMAN_COV2	VVGPVNVKGEDIQLLKSAYENFNQHEVLLAPLLSAGIFGADPIHSLRVCVDTVTRTNVYLA	358
HUMAN_COV1	VVGPVNLNAGEDIQLLKAAYENFNSQDILLAPLLSAGIFGAKPQSLQVCVQTVRTQVYIA *****:* *****:*****. :*:*****. :*:***:***:***:***:	
HUMAN_COV2	VFDKNLYDKLVSSFLEMKSEKQVEQKIAEIPKEEVKPFITESK---PSVEQRKQDDK-KI	414
HUMAN_COV1	VNDKALYEQVVM DYLDNLKPR-----VEAPKQEEPPNTEDSKTEEKSVVQKPVDPKPKI * ** **:*. :* : . : . * ** * * :** ** * : * * *	
HUMAN_COV2	KACVEEVTTTLEETKFLTENLLLYIDINGNLHPDSATLVSDIDITFLKKDAPYIVGDVVQ	474
HUMAN_COV1	KACIDEVTTTLEETKFLTNKLLLFADINGKLYHDSQNMLRGEDMSFLEKDAPYVMGDVIT ***:*****:***: *****: * * . : . * :***:*****:****:	

**Fig.2.** Comparison of the N-term regions of the SARS-CoV-2 and SARS-CoV nsp3 polyprotein. The hypervariable region (HVR) is highlighted in green. The ADP-ribose binding site is highlighted in purple. A less variable region is also seen between position 365 and 459 (also highlighted in green). The rest of the nsp3 polyproteins are more than 80% identical.

```

>PARP14 [Homo sapiens] mono-ADP-ribosyltransferase
Sequence ID: XP_011511231.1 Length: 1209
Range 1: 818 to 937, Score:58.2 bits, Expect:7e-07,
Identities: 40/124 (32%), Positives: 61/124 (49%)

                238
NSP3      VVVNAANVYLLKHGGGVAGALNKATNNAMQVESDDYIATNGPLKVGGSVLSGHNLA-KHCLHVVG
          VVVNA+N  LKH GG+A AL+KA    +Q + D +   G L G + +   L   H +H VG
PARP14    VVVNASNEDLKHYGGLAAALSKAAGPELQADCDQIVKREGRLLPGNATISKAGKLPYHHVIHAVG
          818

NSP3      PNVNKGED---IQLLKSAYE-----NFNQHEVLLAPLLSAGIFGADPIHSLRVCVDTV 351
          P + E    + LL+ A +           ++ + P +S+G+FG      L CV+T+
PARP14    PRWSGYEAPRCVYLLRRAVQLSLCLAEKYKYRSIAIPAISGVEFG-----FPLGRCVETI 937

```

**Fig.3.** Best match from the Blast similarity search output of the entire SARS-CoV-2 nsp3 protein (1935 residues) against the human proteome subset of the NR database (19). This unique match precisely corresponds to the ADP-ribose binding site, central to the viral macrodomain defined by previous structural studies (18).

that this protein plays a significant role in the regulation of the inflammatory response that is out of control in severe cases of COVID-19.

PARP14 was shown to bind to a group of group of IFN-stimulated gene (ISG)-encoded proteins, most with an unknown function, and is required for their nuclear accumulation (12). But its best characterized function is to regulate macrophage activation via STAT1 mono-ADP-ribosylation (13) (MARylation).

STAT1 (for “signal transducer and activator of transcription 1”) is a member of the STAT protein family (i.e. STA1-6). In response to cytokines and growth factors, these proteins are phosphorylated by the receptor associated kinases, form homo- or heterodimers, and translocate to the cell nucleus.

Specifically, STAT1 plays a central role in macrophage activation. Upon IFN-gamma binding to cell surface receptors, Jak kinases (TYK2 and JAK1) are activated and lead to tyrosine and /or serine phosphorylation of STAT1. The phosphorylated STAT1 dimerizes and associates with ISGF3G/IRF-9 to form a transcription factor complex that enters the nucleus (reviewed in 10) where it activate the transcription of IFN-stimulated genes (ISG), which drive the cell in an antiviral state and a pro-inflammatory STAT1-dominated responses.

In the context of this complex interaction network, PARP14, counteracts the dimerization of STAT1 by hindering its phosphorylation through the MARylation at 'Glu-657' and 'Glu-705'. This precludes the translocation of STAT1 to the nucleus, hence down regulating pro-inflammatory cytokine production in macrophages responding to IFN stimulation (13).

This well-documented consequence of STAT1 MARylation on the inflammatory response together with the presence of a PARP14-like ADP-ribose binding site flanking the variable region of nsp3 is at the basis of my hypothesis on the origin of the cytokine storm associated to the severe cases of COVID-19, detailed below.

## Discussion

The ADPR-binding site that we identified in the SARS-CoV-2 nsp3 sequence corresponds to a conserved macrodomain (also called “X domain”) that follows the HVR in all coronaviruses (8). The macrodomain is a ubiquitous structural domain that removes mono-ADP-ribose from proteins, reversing the activity of ADP-ribosyltransferases. This domain has been functionally characterized (reviewed in 15). It was shown that the macrodomain from SARS-CoV (as well as that of other single-stranded positive-sense, single-stranded RNA viruses) can remove mono-ADP-ribose *in vitro* from a variety of target including (PARP1, PARP10, and PARP15) (16), although no specific protein targets for viral macrodomains have been identified *in vivo*. Here we propose that STAT1 is the *in vivo* target of the SARS-CoV-2 macrodomain, thus precisely counteracting its mono-ADP-ribosylation by human PARP14.



The fact that the ADP-ribose binding site of the SARS-CoV-2 nsp3 macrodomain is most resembling that of PARP14 (Fig. 3), suggests that they have a similar conformation, and are both compatible with the addition (for PARP14) or removal (for nsp3) of an ADP-ribose moiety at the same site in the context of the STAT1 3-D structure.

According to this molecular scenario, the expression of nsp3 in IFN-gamma-activated macrophages would indirectly promote a prolonged pro-inflammatory STAT1-dependent expression of interferon stimulated genes by inhibiting the downregulation of STAT1 through its MARylation by PARP14. This would result in the cytokine storm characteristic of severe COVID-19 cases.

Moreover, a recent unexpected finding makes the above hypothesis even more appealing. Following a monumental work, Ziegler et al. (13) just showed that the SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in cell types found in the in the upper airway (epithelial secretory goblet cells), in the lung (type II pneumocytes), and in the small intestine (absorptive enterocytes), all locations and organs linked to clinical symptoms and viral shedding (17). Accordingly, they found evidence for STAT1 binding sites in the promoter region of ACE2. In consequence, the higher level of STAT1 activity promoted by nsp3 in IFN- $\gamma$ -stimulated macrophages might lead to more SARS-CoV-2 receptors being expressed at the surface of bystander epithelial cells, enhancing the infection process. Once started, a positive feedback loop will proceed, with more infections generating more IFN, activating more STAT1, in turn generating more inflammation and more virus receptors, with nsp3 permanently counteracting the PARP14-dependent MARylation of STAT1, thus precluding its downregulation and the termination the pro-inflammatory phase.

In the context of this hypothesis, the variable disordered region flanking the conserved macrodomain, a predicted protein-protein interaction domain, would exhibit a differential affinity for STAT1 in different SARS-CoV strains, hence modulating the efficiency of its de-MARylation by nsp3, and the intensity of the inflammatory response. Alternatively, the nsp3 variable region of some strains could interact with different MARyled targets altogether, leading to different behaviour vis-à-vis the host innate immune response (11, 15, 18).

A relatively straightforward test of this hypothesis would involve the investigation of the ADP-ribosylation status of STAT1 during SARS-CoV-2 *in vitro* infection (using wild-type and macrodomain-mutated strains) and the experimental demonstration of a physical interaction between STA1 and the relevant N-terminal regions of nsp3. As the decrease of the Marylated STAT1 pool could indirectly arise from the inhibition of the MARylation function of PARP14, the eventual interaction of the later with nsp3 should also be investigated. Due to the lockdown constraints, such experiments could not be envisioned in a near future.

Finally, the discovery that ACE2 transcription is under the positive control a STAT1 (17), provides an elegant explanation for the conservation of a functional de-MARylation macrodomain, even though it was shown to be dispensable for the *in vitro* replication of the

virus (7). By following the natural incentive of all viruses to maximize their infectivity, here by increasing the density of its receptor, SARS-CoV-2 triggers a cytokine storm as an unfortunate side-effect of COVID-19. Beyond the fundamental interest of establishing a link between nsp3 and the inflammatory response associated to the disease, an *in vitro* measurement of the de-MARylation efficiency of STAT1 by synthetic constructs nsp3 designed from new SARS-like coronavirus genomic sequences could help predict the virulence of animal strains before they start circulating within human populations.

## Methods

Due to the lockdown constraints imposed during the COVID-19 epidemic, all the bioinformatics analyses presented in this article were performed on publicly accessible bioinformatic platforms. Databases searches were performed at NCBI using the online BlastP on various subsets of NR (15), structural analysis were performed using the PSIPRED Workbench at University College London (<http://bioinf.cs.ucl.ac.uk/psipred/>), and sequence alignments at the EMBL-European Bioinformatic Institute ([www.ebi.ac.uk/Tools/msa/](http://www.ebi.ac.uk/Tools/msa/)) with the Clustal Omega and Water services.

## Acknowledgements

I thank Dr. Chantal Abergel for her support, for correcting earlier versions of the manuscript, and for helpful discussions.

## References

- 1) Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020, 579(7798):270-273. doi: 10.1038/s41586-020-2012-7.
- 2) Lam TT, Shum MH, Zhu HC, Tong YG, Ni XB, Liao YS, Wei W, Cheung WY, Li WJ, Li LF, Leung GM, Holmes EC, Hu YL, Guan Y. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*. 2020 doi:10.1038/s41586-020-2169-0.
- 3) Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*. 2020 doi:10.1038/s41577-020-0311-8.
- 4) Atri D, Siddiqi HK, Lang J, Nauffal V, Morrow DA, Bohula EA. COVID-19 for the Cardiologist: A Current Review of the Virology, Clinical Epidemiology, Cardiac and Other Clinical Manifestations and Potential Therapeutic Strategies. *JACC Basic Transl Sci*. 2020 doi: 10.1016/j.jacbts.2020.04.002.

- 5) Liu PP, Blet A, Smyth D, Li H. The Science Underlying COVID-19: Implications for the Cardiovascular System. *Circulation*. 2020 doi:10.1161/CIRCULATIONAHA.120.047549.
- 6) Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, Yang L, Ding C, Zhu X, Lv R, Zhu J, Hassan B, Feng Y, Tan W, Wang C. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg Microbes Infect*. 2018 7(1):154. doi: 10.1038/s41426-018-0155-5.
- 7) Uversky VN. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys*. 2019 7:10. doi:10.3389/fphy.2019.00010.
- 8) Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res*. 2018 149:58-74. doi:10.1016/j.antiviral.2017.11.001.
- 9) Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Peñaranda S, Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Günther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*. 2003 300(5624):1394-9. doi:10.1126/science.1085952
- 10) Moreno A, Lelli D, de Sabato L, Zaccaria G, Boni A, Sozzi E, Prosperi A, Lavazza A, Cella E, Castrucci MR, Ciccozzi M, Vaccari G. Detection and full genome characterization of two beta CoV viruses related to Middle East respiratory syndrome from bats in Italy. *Virol J*. 2017 14(1):239. doi:10.1186/s12985-017-0907-1.
- 11) Alhammad YMO, Fehr AR. The Viral Macrodomain Counters Host Antiviral ADP-Ribosylation. *Viruses*. 2020 12(4). pii: E384. doi: 10.3390/v12040384.
- 12) Caprara G, Prosperini E, Piccolo V, Sigismondo G, Melacarne A, Cuomo A, Boothby M, Rescigno M, Bonaldi T, Natoli G. PARP14 Controls the Nuclear Accumulation of a Subset of Type I IFN-Inducible Proteins. *J Immunol*. 2018 200(7):2439-2454. doi: 10.4049/jimmunol.1701117.
- 13) Iwata H, Goettsch C, Sharma A, Ricchiuto P, Goh WW, Halu A, Yamada I, Yoshida H, Hara T, Wei M, Inoue N, Fukuda D, Mojcher A, Mattson PC, Barabási AL, Boothby M, Aikawa E, Singh SA, Aikawa M. PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. *Nat Commun*. 2016 7:12849. doi:10.1038/ncomms12849.
- 14) Michalska A, Blaszczyk K, Wesoly J, Bluysen HAR. A Positive Feedback Amplifier Circuit That Regulates Interferon (IFN)-Stimulated Gene Expression and Controls Type I and Type II IFN Responses. *Front Immunol*. 2018 9:1135. doi: 10.3389/fimmu.2018.01135.
- 15) Fehr AR, Jankevicius G, Ahel I, Perlman S. Viral Macrodomains: Unique Mediators of Viral Replication and Pathogenesis. *Trends Microbiol*. 2018 26(7):598-610. doi: 10.1016/j.tim.2017.11.011.
- 16) Fehr AR, Channappanavar R, Jankevicius G, Fett C, Zhao J, Athmer J, Meyerholz DK, Ahel I, Perlman S. The Conserved Coronavirus Macrodomain Promotes Virulence and

- Suppresses the Innate Immune Response during Severe Acute Respiratory Syndrome Coronavirus Infection. *mBio*. 2016 7(6). pii: e01721-16. doi:10.1128/mBio.01721-16.
- 17) Ziegler CGK *et al.* SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* 2020 pii:S0092-8674(20)30500-6. doi:10.1016/j.cell.2020.04.035.
- 18) Li C, Debing Y, Jankevicius G, Neyts J, Ahel I, Coutard B, Canard B. Viral Macro Domains Reverse Protein ADP-Ribosylation. *J Virol*. 2016 90(19):8478-86. doi: 10.1128/JVI.00705-16.
- 19) Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, Funk K, Ketter A, Kim S, Kimchi A, Kitts PA, Kuznetsov A, Lathrop S, Lu Z, McGarvey K, Madden TL, Murphy TD, O'Leary N, Phan L, Schneider VA, Thibaud-Nissen F, Trawick BW, Pruitt KD, Ostell J. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2020 48(D1):D9-D16. doi:10.1093/nar/gkz899.