**WORKING PAPER 26 MAY 2020: Critical analysis of *Andersen et al. The proximal origin of SARS-CoV-2.***
**CDR Jean-Paul Chretien & Dr. Greg Cutlip (DIA/NCMI)**

**1. Background**. The origin of SARS-CoV-2 remains uncertain. Some of its features are unique among the most closely-related known coronaviruses, and a progenitor virus has not been identified. In February 2020, several experts in virology from the US, UK, and Australia co-authored an assessment, posted to a virology blog, of the notable features of SARS-CoV-2 and what they suggest about its origin (Andersen et al., 2020a). The post subsequently was published as a letter-to-the-editor in *Nature Medicine,* a prestigious scientific journal (Andersen et al., 2020b). Andersen et al. concluded that the virus probably arose naturally, not by any sort of laboratory manipulation. Prominent scientists have cited their paper as decisive support for a natural origin scenario (Calisher et al., 2020; Collins, 2020).

Here, we do not advance a particular SARS-CoV-2 origin scenario or take issue with Andersen et al.'s conclusions. We consider the evidence they present, and find that it does not prove that the virus arose naturally. In fact, the features of SARS-CoV-2 noted by Andersen et al. are consistent with another scenario: that SARS-CoV-2 was developed in a laboratory, by methods that leading coronavirus researchers commonly use to investigate how the viruses infect cells and cause disease, assess the potential for animal coronaviruses to jump to humans, and develop drugs and vaccines.

**2. Coronaviruses and SARS-CoV-2 overview**. SARS-CoV-2 is the seventh coronavirus known to infect humans. Four of the others typically cause mild illness, such as the common cold, while two—SARS CoV and MERS-CoV—cause disease with case-fatality rates of around 10% and 34%, respectively (compared to 1% or less in SARS-CoV-2 infection). Other coronaviruses infect only animals, and nearly all coronaviruses are thought to have originated in bats (Fiorni et al., 2017), which host hundreds or thousands of coronaviruses. Both SARS-CoV and MERS-CoV probably originated in bats, passing through intermediate hosts before infecting humans. As RNA viruses, coronaviruses mutate relatively quickly (though less so than other RNA viruses because of a proofreading mechanism). Coronaviruses also evolve through recombination, when distinct viruses infecting the same cell exchange genetic elements.

The closest known relative of SARS-CoV-2 is the bat coronavirus RaTG13. Scientists from the Wuhan Institute of Virology (WIV) first reported the virus in January 2020, but wrote that it was sampled in Yunnan province in 2013, as part of a coronavirus discovery effort (Zhou et al., 2020). Overall, RaTG13 is 96% identical genetically to SARS-CoV-2 (still too dissimilar to be the direct progenitor virus; in addition, no precursors or branches of evolution stemming from a less human-adapted SARS-CoV-2-like virus have been reported [Zhan et al., 2020]). In the regions that encode the spike protein, which mediates entry into host cells, though, RaTG13 shares only 1 of 5 key amino acids for binding to the SARS-CoV-2 receptor, ACE2. The closest match to SARS-CoV-2 in the spike protein was identified in March-August 2019 (reported in 2020) from Malayan pangolins rescued in China. While the pangolin coronavirus is not as close to SARS-CoV-2 as RaTG13 is overall, the spike protein receptor binding domain in the pangolin coronavirus is nearly identical to that of SARS-CoV-2, sharing all 5 key amino acids (Xiao et al., 2020).

Besides the novel combination of features similar to bat and pangolin coronaviruses, SARS-CoV also is unique among its closest known relatives (Betacoronavirus lineage b viruses, such as SARS-CoV, RaTG13, and other bat SARS-related coronaviruses) for the presence of a particular type of host enzyme target site within the spike protein. For a coronavirus to infect host cells, the spike protein must be activated by enzymes supplied by the host (Hoffman et al., 2018). These proteases vary in the degree of priming they effect, and in their distribution within the host and across host species. The specific proteases a coronavirus can hijack helps determine the potential host species and severity of infection.

The spike protein of SARS-CoV-2 includes an insertion of 4 amino acids that creates a cleavage site for furin, which is expressed ubiquitously in humans (Braun et al., 2018). Proteolytic cleavage at this site is required for SARS-CoV-2 infection of human lung cells (Hoffman et al., 2020), and is expected to be especially effective in priming the virus because the site includes 2 consecutive basic amino acids. Such "polybasic" cleavage sites are known to increase virulence in avian influenza viruses, and distinguish highly pathogenic from low-pathogenicity avian influenza (Nao et al., 2017; Xu et al. 2019). Consistent with this effect, disruption of the site in cell-cultured SARS-CoV-2 resulted in attenuated symptoms in hamsters (Lau et al., 2020), suggesting that the novel furin cleavage site may contribute to SARS-CoV-2 pathogenicity or transmissibility (Coutard et al., 2019; Walls et al., 2020; Wang et al., 2020).

**3. Discussion of Andersen et al.'s analysis**. Andersen et al. argue the notable features of SARS-CoV-2 mentioned above—the presence of key amino acids in the receptor binding domain not found in SARS-related CoVs, and the furin cleavage site—likely arose naturally, not through laboratory manipulation. We consider these arguments below, along with their third argument for natural origin based on the global dissimilarity of SARS-CoV-2 to coronaviruses that have been used in published laboratory experiments.

3.1. Receptor binding domain. Andersen et al.: *While the analyses above suggest that SARS-CoV-2 may bind human ACE2 with high affinity, computational analyses predict that the interaction is not ideal and that the RBD sequence is different from those shown in SARS-CoV to be optimal for receptor binding. Thus, the high-affinity binding of the SARS-CoV-2 spike protein to human ACE2 is most likely the result of natural selection on a human or human-like ACE2 that permits another optimal binding solution to arise. This is strong evidence that SARS-CoV-2 is not the product of purposeful manipulation.*

The argument is that if someone wanted to design a coronavirus that binds with high affinity to human ACE2, they would not have designed SARS-CoV-2, since the computational analysis they would have undertaken in the planning stage would have predicted a lower affinity binding to ACE2 than what is, in fact, observed.

This is not a scientific argument but rather an assumption of intent and methodology for a hypothesized scientist. Instead of aiming to design a virus that binds with high affinity to ACE2, a researcher may have chosen to investigate, empirically, the effect of one or more receptor binding domain variants on receptor binding or infectivity. In fact, leading coronavirus research laboratories have been doing this for years to study the potential for bat coronaviruses to infect humans, as described below.

A typical approach is to synthesize a chimeric coronavirus, usually with spike protein and remainder of the virus ("backbone") from different bat, human, or other sources, simulating recombination events that may occur naturally. Below are selected examples of this and other coronavirus manipulations to investigate infectivity and virulence:

- In 2007, scientists from the WIV investigated human ACE2 binding potential by manipulating the spike protein of SARS-CoV and bat SARS-like-CoVs, and constructing chimeras with different sequences of the SARS-CoV spike protein inserted into a bat SARS-like-CoV backbone (Ren et al., 2008). They identified a minimal sequence that converted the bat SARS-like-CoV spike protein from non-ACE2 binding to ACE2 binding, indicating that acquisition of this sequence could allow a species jump to humans.

- Also in 2007, investigators from the University of North Carolina, the National Institutes of Health, and the Centers for Disease Control and Prevention adapted SARS-CoV to mice through serial passage, resulting in a strain that was lethal to mice (Roberts et al., 2007). They identified 6 coding mutations associated with adaptation and virulence, and demonstrated their effect by introducing these into SARS-CoV and re-creating the phenotype.

- In 2015, a collaboration that included the University of North Carolina and the WIV generated a chimeric virus with the spike protein of a bat SARS-like-coronavirus in a (mouse-adapted) SARS-CoV backbone, and showed that the chimera could infect human airway cells (Menachery et al., 2015). Based on these results, they also synthetically re-derived the full-length bat SARS-like coronavirus and showed it was viable in human airway cells and mice. (University of North Carolina researchers were granted a patent for chimeric coronavirus spike proteins methods in 2018 [Baric et al., 2015]).

- In 2016, researchers from the WIV led a collaboration that constructed an array of chimeric coronaviruses with the SARS-like bat CoV backbone and variants of the spike protein from 8 other bat SARS-like CoVs (which were reported for the first time in this study) (Hu et al., 2017). All of these viruses we able to use the human ACE2 receptor to enter cells.

In the context of this research, SARS-CoV-2 could have been synthetized by combining a backbone from a coronavirus similar to RaTG13 with the receptor binding domain of a coronavirus similar to the one recently isolated from pangolins. Such research might have aimed to investigate pangolins as possible intermediate hosts for bat coronaviruses potentially pathogenic for humans, and would have been consistent with the longstanding line of investigations described above.

3.2. Furin cleavage site. Andersen et al.: *The acquisition of both the polybasic cleavage site and predicted O-linked glycans also argues against culture-based scenarios. New polybasic cleavage sites have been observed only after prolonged passage of low-pathogenicity avian influenza virus in vitro or in vivo. Furthermore, a hypothetical generation of SARS-CoV-2 by cell culture or animal passage would have required prior isolation of a progenitor virus with very high genetic similarity, which has not been described. Subsequent generation of a polybasic cleavage site would have then required repeated passage in cell culture or animals with ACE2 receptors similar to those of humans, but such work has also not previously been described. Finally, the generation of the predicted O-linked glycans is also unlikely to have occurred due to cell-culture passage, as such features suggest the involvement of an immune system.*

The central argument is that generation of the furin cleavage site would have required a progenitor virus, which has not been published. Like the first argument, this also is not based on scientific analysis but on an assumption that the prior work would have been published if it had been done. However, absence of a publication does not mean that the research was not done. Perhaps the experiments were aborted or not reported because of the SARS-CoV-2 outbreak? Perhaps the results were never intended for publication? In a recent example of delayed publication from the COVID-19 pandemic, WIV researchers first reported RaTG13 in January 2020, but stated that they had discovered the virus in 2013 (Zhou et al., 2020).

Andersen et al. acknowledge that polybasic cleavage sites have arisen during *in vitro* or *in vivo* passage, with avian influenza. Therefore, the possibility of the SARS-CoV-2 furin site arising during passage in the laboratory cannot be dismissed.

However, this is not the only possible method of generating a new furin cleavage site. Laboratories also have directly inserted furin cleavage sites into coronaviruses. For example:

- Scientists in the United States (Belouzard et al., 2009; Follis et al., 2006) and Japan (Watanabe et al., 2008) have inserted furin cleavage sites into the spike protein of SARS-CoV using pseudotype systems (a relatively safe way to study dangerous pathogens, which allows the evaluation of the coronavirus spike protein, or envelope proteins from other viruses, in the context of another virus without producing progeny of the chimeric virus).

- In 2015, a collaboration that included the University of North Carolina and the WIV investigated mutations that could enable HKU4, a MERS-CoV-related bat coronavirus, to infect humans. Noting a furin cleavage site present in MERS-CoV but absent in HKU4, they introduced a mutation into the spike protein of HKU4 that created one (in a pseudotype system), and showed this allowed it to use host protease to mediate human cell entry (Yang et al., 2015).

- In 2019, researchers at China Agricultural University constructed an infectious bronchitis virus (a bird coronavirus) with a polybasic cleavage site in the spike protein, and showed that this mutant virus led to higher mortality in chickens (Cheng et al., 2019).

Building on the possible origin scenario described at the end of section 3.1 above, SARS-CoV-2 could have been synthesized by generating a polybasic cleavage site in a chimeric virus with RaTG13-like backbone and pangolin coronavirus-like receptor binding domain (with additional mutations arising during cell-culture or animal adaptation, as occurred with increasing virulence after passage of SARS-CoV in a mouse model [Roberts et al., 2007]).

Andersen et al. note that the O-linked glycans predicted to flank the polybasic cleavage site argue against a cell culture-based laboratory scenario. O-linked glycans may help shield a broad set of viruses, including coronaviruses, from a host immune system, and typically arise under immune selection, which would not occur in cell culture. However, the glycans could have arisen in animal, rather than cell culture, experiments. There is ample precedent for use of animal models in coronavirus research, such as inoculation of rats with bat SARS-like CoVs, led by China's Third Military Medical University (Hu et al., 2018), and several of the efforts cited earlier. Pangolins, or other animals with similar ACE2 conformation, could have been used in such *in vivo* experiments. (We note also that some preliminary investigations have found evidence for the O-linked glycans predicted by Andersen et al. [Shajahan et al., 2020], while others have not [Walls et al., 2020]).

3.3. Reverse genetics system. Andersen et al.: *Furthermore, if genetic manipulation had been performed, one of the several reverse-genetic systems available for betacoronaviruses would probably have been used. However, the genetic data irrefutably show that SARS-CoV-2 is not derived from any previously used virus backbone.*

Reverse genetics approaches seek to understand the phenotypic effects of a genotype (in contrast to forward genetics, which begins with a phenotype and investigates its genetic basis). Scientists have developed reverse genetics systems for SARS-CoV, bat SARS-like-CoVs, and other human and animal coronaviruses (Almazan et al., 2014). These systems are used to synthesize coronaviruses so specific genomic sequences and regions can be manipulated, and enabled many of the studies noted earlier.

It is clear that SARS-CoV-2 does not derive from a published coronavirus reverse genetics system, since the virus is not a close match to a CoV backbone used in any published study. As noted in section 3.2, though, the lack of a published report does not mean that the study was not performed. In the 6 years before the COVID-19 outbreak, the number of potential bat SARS-like CoV backbones has increased with coronavirus discovery efforts, like those conducted the WIV (Ge et al., 2013; Ge et al., 2016; Hu et al., 2017). Some coronavirus reverse genetics systems yield synthetic viruses that are identical genetically to the original virus (Almazan et al., 2014), such as one developed in 2002 at the University of North Carolina (Yount et al., 2002).

Recent technological innovations make it easier than ever for scientists to develop new reverse genetics systems. For example, researchers from Huazhong Agricultural University used CRISPR (clustered regularly interspaced short palindromic repeat) gene editing technology in the first direct *in vitro* manipulation of full-length coronavirus cDNA (from which the virus is reverse-engineered) in April 2019 (Wang et al., 2019). The authors noted that their reverse genetics platform "will simplify the construction of mutant infectious clones and help accelerate progress in coronavirus research."

A further illustration of modern capabilities comes from the COVID-19 pandemic: several laboratories already have synthesized SARS-CoV-2, using reverse genetics systems they developed within weeks of publication of the first genome sequence (Thao et al., 2020; Xie et al., 2020).

**4. Conclusion**. The arguments that Andersen et al. use to support a natural-origin scenario for SARS-CoV-2 are based not on scientific analysis, but on unwarranted assumptions. A long line of research shows that leading coronavirus laboratories do not work as described in the laboratory-origin scenario Andersen et al. consider and dismiss. SARS-CoV-2—a bat coronavirus with pangolin coronavirus receptor binding domain—is consistent with the chimeric constructs these laboratories have developed and studied for more than a decade. While key components of a laboratory effort resulting in SARS-CoV-2, such as generation of the furin cleavage site and development of a new reverse genetics system, have not been reported, this does not prove they did not occur. Coronavirus researchers have conducted these studies for other coronaviruses; technically, they would not have been difficult. The recent RaTG13 report demonstrates coronavirus researchers do not publish all of their research at the time it is conducted.

This critique of Andersen et al.'s arguments does not aim to show that SARS-CoV-2 originated in a laboratory, much less to identify a specific laboratory source or to characterize the goals of research that could have generated the virus. We highlight the features of SARS-CoV-2, noted by Andersen et al., are consistent with long-standing and on-going laboratory experiments; the evidence Andersen et al. present does not lessen the plausibility of laboratory origin.

**5. References**

Almazan F, Gonzalez JM, Penzes Z. Coronavirus reverse genetic systems: Infectious clones and replicons. Virus Res. 2014; 189: 262–270. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727449/

Andersen KG, Rambaut A, Lipkin WI, Holmes, Garry RF. The proximal origin of SARS-CoV-2. Virological.org 2020a. http://virological.org/t/the-proximal-origin-of-sars-cov-2/398

Andersen KG, Rambaut A, Lipkin WI, Holmes, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020b; 26: 450–452. https://www.nature.com/articles/s41591-020-0820-9

Baric R, Agnihothram S, Yount B. Methods and compositions for chimeric coronavirus spike proteins. US Patent US9884895B2. 2018. https://patentimages.storage.googleapis.com/ba/78/39/b9581f7e428f61/US9884895.pdf

Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci USA 2009; 106:5871-5876. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2660061/

Braun E, Sauter D. Furin-mediated protein processing in infectious diseases and cancer. Clin Transl Immunology 2018; 8:e1073. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6682551/

Calisher C, Carroll D, Colwell R, Corley RB, Daszak P, Drosten C, et al. Statement in support of the scientists, public health professionals, and medical professionals of China combatting COVID-19. Lancet 2020; 295:E42-E43. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30418-9/fulltext

Cheng J, Zhao Y, Xu G. The S2 Subunit of QX-type infectious bronchitis coronavirus spike protein is an essential determinant of neurotropism. Viruses 2019; 11: 10. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6832359/

Collins F. Genomic study points to natural origin of COVID-19. NIH Director's Blog 2020. https://directorsblog.nih.gov/2020/03/26/genomic-research-points-to-natural-origin-of-covid-19/

Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019- nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res. 2020; 176: 104742. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7114094/

Fiorni D, Cagliani R, Clerici M, Sironi M. Molecular evolution of human coronavirus genomes. Trends Microbiol 2017; 25:35-48. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7111218/

Follis KE, York J, Nunberg JH. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cel-cell fusion but does not affect viral entry. Virology 2006; 350:358-369. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7111780/

Ge XY, Li JL, Yang XL, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 2013; 503:535-538. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5389864/

Ge XY, Wang N, Zhang W et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. Virol Sin. 2016; 31: 31–40.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7090819/

Hoffman M, Hoffman-Winkler H, Pohlmann S. Priming time: how cellular proteases arm coronavirus spike proteins. Activation of Viruses by Host Proteases 2018 Feb 16:71-98.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7122371/

Hoffman M, Kleine-Weber H, Pohlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection in human lung cells. Molecular Cell 2020.
https://www.sciencedirect.com/science/article/pii/S1097276520302641

Hu B, Zeng L-P, Yang X-L, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog. 2017; 13: e1006698.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5708621/

Hu D, Zhu C, Ai L, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg Microbes Infect. 2018; 7: 154.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6135831/

Lau SY, Wang P, Mok B W-Y, et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. Emerg Microbes Infect. 2020. https://www.tandfonline.com/doi/full/10.1080/22221751.2020.1756700

Menachery VD, Yount BL, Debbink K, et al., A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nature Medicine. 2015; 21: 12.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4797993/

Nao N, Yamagishi J, Miyamoto H, et al. Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutin. mBio 2017; 8:e02298-16.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5312086/

Ren W, Qu X, Li W, et al., Difference in Receptor Usage between Severe Acute Respiratory Syndrome (SARS) Coronavirus and SARS-Like Coronavirus of Bat Origin. J Virol, 2008 Feb; 82(4): 1899–1907.
https://jvi.asm.org/content/82/4/1899

Roberts A, Deming D, Paddock CD, et al. A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mise. PLoS Pathog 2007; 3:e5.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1769406/

Shajahan A, Supekar N, Gleinich A, Azadi P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. bioRxiv 2020.
https://www.biorxiv.org/content/10.1101/2020.04.01.020966v1

Thao TT, Labroussa F, Ebert N, et al. Rapid reconstruction of SARS-CoV-2 using a synthetics genomics platform. Nature 2020. https://www.nature.com/articles/s41586-020-2294-9

Walls A, Park Y-J, Tortorici M, et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7102599/

Wang G, Liang R, Liu Z, et al. The N-terminal domain of spike protein is not the enteric tropism determinant for transmissible gastroenteritis virus in piglets. Viruses 2019; 11:313. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6520731/

Wang Q, Qiu Y, Li J-Y, et al. A unique protease cleavage site predicted in the spike protein of the novel pneumonia Coronavirus (2019-nCoV) potentially related to viral transmissibility. Virol Sin. 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7091172/

Watanabe R, Matsuyama S, Shirato K, et al. Entry from the cell surface of Severe Acute Respiratory Syndrome Coronavirus with cleaved S protein as revealed by pseudotype virus bearing cleaved S protein. J Virol 2008; 82:11985-11991. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2583654/

Xiao K, Zhai J, Feng Y, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature 2020. https://www.nature.com/articles/s41586-020-2313-x

Xie X, Murato A, Lokugamage K, et al. An infectious cDNA clone of SARS-CoV-2. Cell Host Microbe 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7153529/

Xu S, Li X, Yang J, et al. Comparative Pathogenicity and Transmissibility of Pandemic H1N1, Avian H5N1, and Human H7N9 Influenza Viruses in Tree Shrews. Front. Microbiol 2019. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6933948/

Yang Y, Liu C, Du L, et al. Two mutations were critical for bat-to-human transmission of Middle East Respiratory Syndrome coronavirus. J Virol 2015; 89:9119-9123. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4524054/

Yount B, Denison MR, Weiss SR, Baric RS. Systematic assembly of full-length infectious cDNA of mouse hepatitis virus strain A59. J Virol 2002; 76:11065-11078. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC136593/

Zhan SH, Deverman BE, Chan YA. SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence? bioRxiv 2020. https://www.biorxiv.org/content/10.1101/2020.05.01.073262v1

Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020; 579:270-273. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7095418/