Short Communication

In silico exploration of enzymes involved in sialic acid biosynthesis and their possible role in SARS-CoV-2 infection

V.C. Divya [a], Balasubramanian Saravanakarthikeyan [b, *]

[a] Department of Oral Medicine and Radiology, SRM Kattankulathur Dental College and Hospitals, SRM Institute of Science and Technology, Chennai, 603 203, Tamil Nadu, India
[b] Department of Conservative Dentistry and Endodontics, SRM Dental College, Ramapuram, SRM Institute of Science and Technology, Chennai, 600 089, Tamil Nadu, India

A B S T R A C T

Salivary glands are considered important targets of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Recent evidence suggests that along with angiotensin converting enzyme 2, certain cell surface sialic acids (Sia) may function as receptors for binding SARS-CoV-2 spike protein. Over 50 forms of Sia have been identified in nature, with N-acetylneuraminic acid (Neu5Ac) being the most abundant. We explored the Human Protein Atlas repository to analyze important enzymes in Neu5Ac biosynthesis and propose a hypothesis that further highlights the significance of salivary glands in coronavirus disease 19 (COVID-19). This work may facilitate research into targeted drug therapies for COVID-19.

© 2021 Japanese Association for Oral Biology. Published by Elsevier B.V. All rights reserved.

Salivary glands have been suggested as important targets of infection by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), causing coronavirus disease 19 (COVID-19). Angiotensin converting enzyme 2 (ACE2), the primary receptor for SARS-CoV-2 entry into the host cell, along with TMPRSS2 (a type II transmembrane serine protease) and furin, which are responsible for priming and activating the spike protein of SARS-CoV-2, respectively, are highly expressed in the salivary glands [1–3]. Recently, researchers have provided *in silico* evidence of a dual strategy, wherein in addition to ACE2, certain sialic acids (Sia) present on the cell surface may also function as potential receptors for binding the spike protein of SARS-CoV-2 [4–7].

Sia are monosaccharides found on the outermost ends of sugar chains of glycoproteins or glycolipids (glycoconjugates), which are present on the cell surface of vertebrates, higher invertebrates, and few bacteria and play significant roles in several physiological and pathological processes [8,9]. More than 50 forms of Sia have been identified in nature, of which the most abundant is N-acetylneuraminic acid (Neu5Ac). Sia and other host sugar molecules are often used as receptors by a wide range of viruses, including coronaviruses [10,11]. Recently, it has also been demonstrated that Neu5Ac exhibits affinity for the SARS-CoV-2 spike protein [12].

We explored the Human Protein Atlas (HPA) repository [13] to investigate the expression of genes related to Sia metabolism in human tissues, with an emphasis on three main enzymes involved in the biosynthesis of Neu5Ac, namely, glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase (GNE) enzyme, Neu5Ac 9-phosphate synthase (NANS) and Neu5Ac-9-phosphate phosphatase (NANP) [9]. The HPA repository is a freely available interactive resource that maps the human tissue proteome to analyze tissue profiles of specific protein classes in order to achieve its spatial localization down to the single-cell level [13–15], as depicted in Fig. 1. All the information in HPA is provided without any restrictions to allow researchers to get a holistic map of the human body.
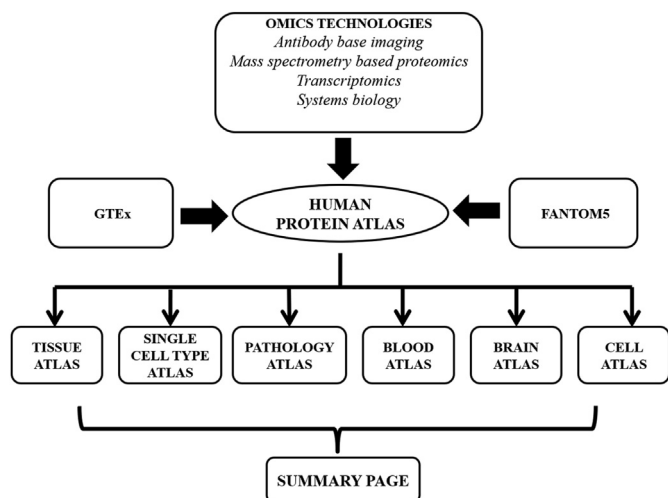
Fig. 1. Brief overview of Human Protein Atlas (HPA) repository. The available data in HPA repository comprises an integration of various omics technologies, such as antibody-based imaging, mass spectrometry-based proteomics, transcriptomics and systems biology. More than 13 million tissue-based immunohistochemistry images are available, each of which is annotated by pathologists. A comprehensive overview of protein expression patterns in human tissues is possible by stringently evaluating these immunohistochemical staining methods, RNA-seq data from internal and external sources, and protein/gene characterization, with an emphasis on RNA-seq. The HPA consists of six categories: tissue atlas, single cell type atlas, pathology atlas, blood atlas, brain atlas, and cell atlas, with each providing information on a different aspect of human proteins. It also includes a summary page for each protein, which can be navigated via the search engine and presents a summary gathered from the content of different aspects of the human atlas. GTEx, Genotype-Tissue Expression is an online portal to study tissue-specific gene expression as well as their regulation. RNA-seq data from 36 tissue types were mapped and included in the HPA for all corresponding genes. FANTOM5, the Functional Annotation of Mammalian Genomes 5 project presents comprehensive expression profiles and functional annotation of mammalian cell type-specific transcriptomes using cap analysis of gene expression (CAGE). The normalized tags per million for each gene were calculated and incorporated into HPA. The mRNA expression levels in human tissue, presented in the HPA repository, are based on RNA-seq data generated by the HPA, GTEx portal, and CAGE data generated by the FANTOM5 consortium. Consensus normalized expression (NX) levels for human tissues were obtained by the combination of data from the three datasets (HPA, GTEx, and FANTOM5). For all annotated cell types, the protein expression ranged from: not detected (n), low (l), medium (m), and high (h). Beneath this expression diagram, the protein expression in each of the annotated cell types associated with the specific tissue/organ are reported using similar units.

An overview of various enzymes involved in Sia synthesis, as well as their mRNA and protein expression in salivary glands and lungs, is presented in Table 1. Considering the consensus among the three data sets, HPA, GTEx, and FANTOM5, the normalized mRNA expression of ACE2 was only slightly higher in salivary glands (1.1 NX) than in the lungs (0.8 NX), as reported previously [1,13,15]. The mRNA expression in the HPA data set revealed 0.5 pTPM (1 million transcripts per kilobase million) in the salivary glands and 1.7 pTPM in the lungs. GTEx data set revealed 1.8 pTPM in the salivary glands and 1.1 pTPM in the lungs whereas FANTOM5 data set revealed 0.3 and 2.8 scaled tags per million in the salivary glands and lungs, respectively. ACE2 protein expression was neither present in the salivary glands nor in the lungs (Table 1).

Further exploration of HPA presented some noteworthy findings [13,15]. Interestingly, we observed that the GNE gene was also highly enhanced in the salivary glands compared to the lungs (Table 1) [16]. As mentioned previously, GNE plays a crucial role in the initiation and regulation of the synthesis of Neu5Ac, which is the precursor of Sia. Considering the consensus among three datasets such as GTEx, HPA, and FANTOM5, normalized GNE expression was higher in the salivary glands (45.3NX) than in the lungs (5.5NX) [13,15,16]. This is also in accordance with the findings

of the individual data sets. The mRNA expression in the HPA data set revealed 33.4 pTPM in the salivary glands and 13.7 pTPM in the lungs. Similarly, both GTEx and FANTOM5 data sets also revealed higher GNE expression in the salivary glands (17.8 pTPM and 47.3 scaled tags per million, respectively) than in the lungs (9.3 pTPM and 28.2 scaled tags per million). The salivary glands showed the highest GNE mRNA expression in the body following the liver, where the maximum expression was observed. Nevertheless, GNE protein expression was highest in the lungs next to the naso-pharynx, with high detection in macrophages and no detection in the alveolar cells (Table 1). These findings corroborate the observation that SARS-CoV-2 can engage in Sia found in human respiratory cells [13].

According to the HPA repository, the normalized mRNA expression levels of NANS were also reported to be higher in the salivary glands (80.0 NX) than in the lungs (17.1 NX), as revealed by HPA, GTEx, and FANTOM5 datasets [13,15,16]. This is also in accordance with the findings in the individual datasets. The mRNA expression in the HPA data set revealed an 80.4 pTPM in the salivary glands and a 57.3 pTPM in the lungs. GTEx data set revealed 101.7 pTPM in the salivary glands and 68.8 pTPM in the lungs. The FANTOM5 dataset revealed 56.2 and 62.4 scaled tags per million, respectively, in the salivary glands and lungs. Interestingly, according to the consensus data, the highest mRNA expression of NANS in the human body was reported in the salivary glands. In addition, NANS protein expression was high in the glandular cells of the salivary glands and lung macrophages, whereas it was reported to be absent in the alveolar cells (Table 1).

Based on the consensus among the HPA, GTEx, and FANTOM5 datasets, the normalized mRNA expression of NANP was higher in the salivary glands (7.8 NX) than in the lungs (6.7 NX) [13,15,16]. The mRNA expression in the HPA data set was 1.2 pTPM in the salivary glands and 4.4 pTPM in the lungs. The GTEx dataset revealed a 2.4 pTPM in the salivary glands and 2.7 pTPM in the lungs. The FANTOM5 dataset revealed 1.7 and 7.1 scaled tags per million in the salivary glands and lungs, respectively. However, no data are available regarding NANP protein expression (Table 1).

The central dogma of molecular biology closely connects DNA, RNA, and protein molecules [13,15,16]. The nucleotide sequence determines the sequence of its mRNA product, and the mRNA sequence determines the amino acid sequence of the resulting polypeptide. The relationship between the concentration of a transcript and that of a protein derived from a particular locus is not trivial. Systematic studies at the genomic level that quantify transcripts and proteins revealed the significance of several processes beyond transcript concentration that influence the protein expression level. These processes include translation rates, translation rate modulation, protein half-life, protein synthesis delay, and protein transport. Thus, a direct comparison between protein and mRNA abundances from the same location or from the same cell type may not be appropriate. This could explain the variations between the protein expression and mRNA levels of several enzymes observed in the data sets in the present study. For instance, GNE protein expression was minimally detected in the glandular cells of the salivary glands despite its high mRNA expression. In addition, the protein *in silico* findings in the HPA are concerned with immunohistochemistry, which may not be the best tool for quantitative evaluation as it simply provides information pertaining to localization [2]. This may also partially explain why the data repositories yielded contradictory information that could be reconciled by missing data in one being present in the other.

Hence, we hypothesize that the high expression of enzymes such as GNE, NANS, and NANP in the salivary glands, which play a pivotal role in the synthesis of Sia, needs to be considered. This correlation further implicates the significance of salivary glands as

Table 1

mRNA and protein expression of ACE2, GNE, NANS and NANP in salivary glands and lungs.

	Salivary Glands	Lungs
Angiotensin converting enzyme 2 (ACE2)		
RNA		
Consensus, NX	1.1	0.8
HPA, pTPM	0.5	1.7
GTEx, pTPM	1.8	1.1
FANTOM5, scaled tags per million	0.3	2.8
Protein	Glandular cells: not detected	Macrophages: not detected
		Alveolar cells: not detected
GNE (Glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase)		
RNA		
Consensus, NX	45.3	5.5
HPA, pTPM	33.9	13.7
GTEx, pTPM	17.8	9.3
FANTOM5, scaled tags per million	47.3	28.2
Protein	Glandular cells: low detection	Macrophages: high
		Alveolar cells: not detected
Neu5Ac 9-phosphate synthase (NANS)		
RNA		
Consensus, NX	80.0	17.1
HPA, pTPM	80.4	57.3
GTEx, pTPM	101.7	68.8
FANTOM5, scaled tags per million	56.2	62.4
Protein	Glandular cells: high detection	Macrophages: high detection
		Alveolar cells: not detected
Neu5Ac-9-phosphate phosphatase (NANP)		
RNA		
Consensus, NX	7.8	6.7
HPA, pTPM	1.2	4.4
GTEx, pTPM	2.4	2.7
FANTOM5, scaled tags per million	1.7	7.1
Protein	No data	No data

HPA, human protein atlas; GTEx, genotype tissue expression; FANTOM5, functional annotation of mammalian genomes 5; NX, normalized expression; pTPM, 1 million transcripts per kilobase million. Protein data are interpreted as follows (according to The Human Protein Atlas): "Protein expression score is manually annotated on immunohistochemical figures based on the intensity of staining (negative, weak, moderate or strong) as well as fraction of stained cells (<25%, 25%—75% or >75%): negative-no detection; weak <25%- no detection; weak combined with 25%—75% or >75%- low detection".

an important target in COVID-19 infection. Hypothetically, this observation could contribute to obtaining greater insight into the SARS-CoV-2 disease process, which will aid in developing future interventions and research.

Ethical approval

Ethical approval is not required.

CRediT authorship contribution statement

V.C. Divya: Literature search, writing, reviewing, and editing. **B. Saravanakarthikeyan:** Literature search, table, writing, and reviewing.

Conflicts of interest

The authors declare that there are no competing interests.

Acknowledgments

References

[1] Xu J, Li Y, Gan F, Du Y, Yao Y. Salivary glands: potential reservoirs for COVID-19 asymptomatic infection. J Dent Res 2020;99:989.

[2] Pascolo L, Zupin L, Melato M, Tricarico PM, Crovella S. TMPRSS2 and ACE2 Co-expression in SARS-CoV-2 salivary glands infection. J Dent Res 2020;99:1120—1.

[3] Zupin L, Pascolo L, Crovella S. Is *FURIN* gene expression in salivary glands related to SARS-CoV-2 infectivity through saliva? J Clin Pathol 2021;74:209—11.

[4] Morniroli D, Giannì ML, Consales A, Pietrasanta C, Mosca F. Human sialome and coronavirus disease-2019 (COVID-19) pandemic: an understated correlation? Front Immunol 2020;11:1480.

[5] Milanetti E, Miotto M, Di-Rienzo L, Monti M, Gosti G, Ruocco G. In-Silico evidence for two receptors based strategy of SARS-CoV-2. bioRxiv 2020. https://doi.org/10.1101/2020.03.24.006197.

[6] Kuchipudi SV, Nelli RK, Gontu A, Satyakumar R, Surendran NM, Subbiah M. Sialic acid receptors: the key to solving the enigma of zoonotic virus spillover. Viruses 2021;3:262.

[7] Sun XL. The role of cell surface sialic acids for SARS-CoV-2 infection. Glycobiology 2021 Apr 28:cwab032. https://doi.org/10.1093/glycob/cwab032. Epub ahead of print. PMID: 33909065.

[8] Varki A. Sialic acids in human health and disease. Trends Mol Med 2008;14:351—60.

[9] Li Y, Chen X. Sialic acid metabolism and sialyl transferases: natural functions and applications. Appl Microbiol Biotechnol 2012;94:887—905.

[10] Wasik BR, Barnard KN, Parrish CR. Effects of sialic acid modifications on virus binding and infection. Trends Microbiol 2016 Dec;24(12):991—1001. https://doi.org/10.1016/j.tim.2016.07.005. Epub 2016 Aug 1. PMID: 27491885; PMCID: PMC5123965.

[11] Distinct roles for sialoside and protein receptors in coronavirus infection. mBio 2020;11(1). https://doi.org/10.1128/mBio.02764-19. e02764-19.

[12] Baker AN, Richards SJ, Guy CS, Tr Congdon, Hasan M, Zwetsloot AJ, et al. The SARS-COV-2 spike protein binds sialic acids and enables rapid detection in a lateral flow point of care diagnostic device. ACS Cent Sci 2020;6:2046—52.

[13] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science 2015;347:1260419.

[14] Human protein atlas (HPA) available from: https://www.proteinatlas.org/about/assays+annotation [Last accessed online 21 July 2021].

[15] Genotype-tissue expression (GTEx) available from: https://www.gtexportal.org/home/documentationPage (Last accessed online 21 July 2021).

[16] Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. Cell 2016;165:535—50.