Informatics in Medicine Unlocked 20 (2020) 100384

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Identification and classification of differentially expressed genes reveal potential molecular signature associated with SARS-CoV-2 infection in lung adenocarcinomal cells

Opeyemi S. Soremekun, Kehinde F. Omolabi, Mahmoud E.S. Soliman *

Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban, 4001, South Africa

ARTICLE INFO

Keywords:
Differentially expressed genes
SARS-CoV-2
COVID-19
Enrichment analysis
RNAseq

ABSTRACT

Genomic techniques such as next-generation sequencing and microarrays have facilitated the identification and classification of molecular signatures inherent in cells upon viral infection, for possible therapeutic targets. Therefore, in this study, we performed a differential gene expression analysis, pathway enrichment analysis, and gene ontology on RNAseq data obtained from SARS-CoV-2 infected A549 cells. Differential expression analysis revealed that 753 genes were up-regulated while 746 down-regulated. SNORA81, OAS2, SYCP2, LOC100506985, and SNORD35B are the top 5 upregulated genes upon SARS-Cov-2 infection. Expectedly, these genes have been implicated in the immune response to viral assaults. In the Ontology of protein classification, a high percentage of the genes are classified as Gene-specific transcriptional regulator, metabolite interconversion enzyme, and Protein modifying enzymes. Twenty pathways with P-value lower than 0.05 were enriched in the up-regulated genes while 18 pathways are enriched in the down-regulated DEGs. The toll-like receptor signalling pathway is one of the major pathways enriched. This pathway plays an important role in the innate immune system by identifying the pathogen-associated molecular signature emanating from various microorganisms. Taken together, our results present a novel understanding of genes and corresponding pathways upon SARS-Cov-2 infection, and could facilitate the identification of novel therapeutic targets and biomarkers in the treatment of COVID-19.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been implicated as the causative agent of the recent global pandemic disease named Coronavirus disease 2019 (COVID-19). SARS-CoV-2 belongs to a larger family of *Coronaviridae* (CoVs) which causes gastrointestinal, respiratory, and neurological diseases to a varying degree of severity [1,2]. CoVs are non-segmented positive-sense RNA viruses that are zoonotically introduced into the human population [3]. They are divided into four groups: the α, β,γ, and δ-CoVs [4]. SARS-CoV-2, SARS-CoV and MERS-CoV are β-coronaviruses that cause lethal respiratory infections [5]. It has been estimated that deaths from the ravage of SARS-CoV-2 in a short space of time, has been more than that which resulted from SARS-CoV, and MERS-CoV combined [6] (Fig. 1). The method of transmission of SARS-CoV-2 is mainly human to human especially by close contact with an infected host's aerosols and also the surfaces these have contaminated [7,8]. To date, there is no effective antiviral treatment for SARS-CoV-2 [9]. Instead, existing drugs are being repurposed for its treatment. Examples of these drugs among many others include serine protease inhibitors and a promising drug candidate, remdesivir, which has been found to have broad-spectrum activity against RNA viruses by hampering the mechanism of replication [10]. Other examples include favipiravir, hydroxychloroquine, co-administration of lopinavir and ritonavir, baricitinib, etc. [11–13]. These drugs and other supplemental therapies have functioned to resolve the infection to varying degrees of success [14–17]. Different approaches have been employed in the design of CoV vaccine, a majority of these strategies are developed to target the surface-exposed spike (S) glycoprotein, such as targeting the full S protein or S1-receptor-binding domain [43]. The similarity in the T-cell epitopes of MERS-CoV, SARS- and SARS-CoV-2 has been leveraged upon to examine the cross-reactivity of MERS-CoV and SAR-CoV vaccines [44]. A comparative study on the sequence length of S protein between SARS-CoV-2 and SARS-CoV showed that the variable amino acids were found on the S1

* Corresponding author.
E-mail address: soliman@ukzn.ac.za (M.E.S. Soliman).

https://doi.org/10.1016/j.imu.2020.100384
Received 26 April 2020; Received in revised form 20 June 2020; Accepted 21 June 2020
Available online 23 June 2020

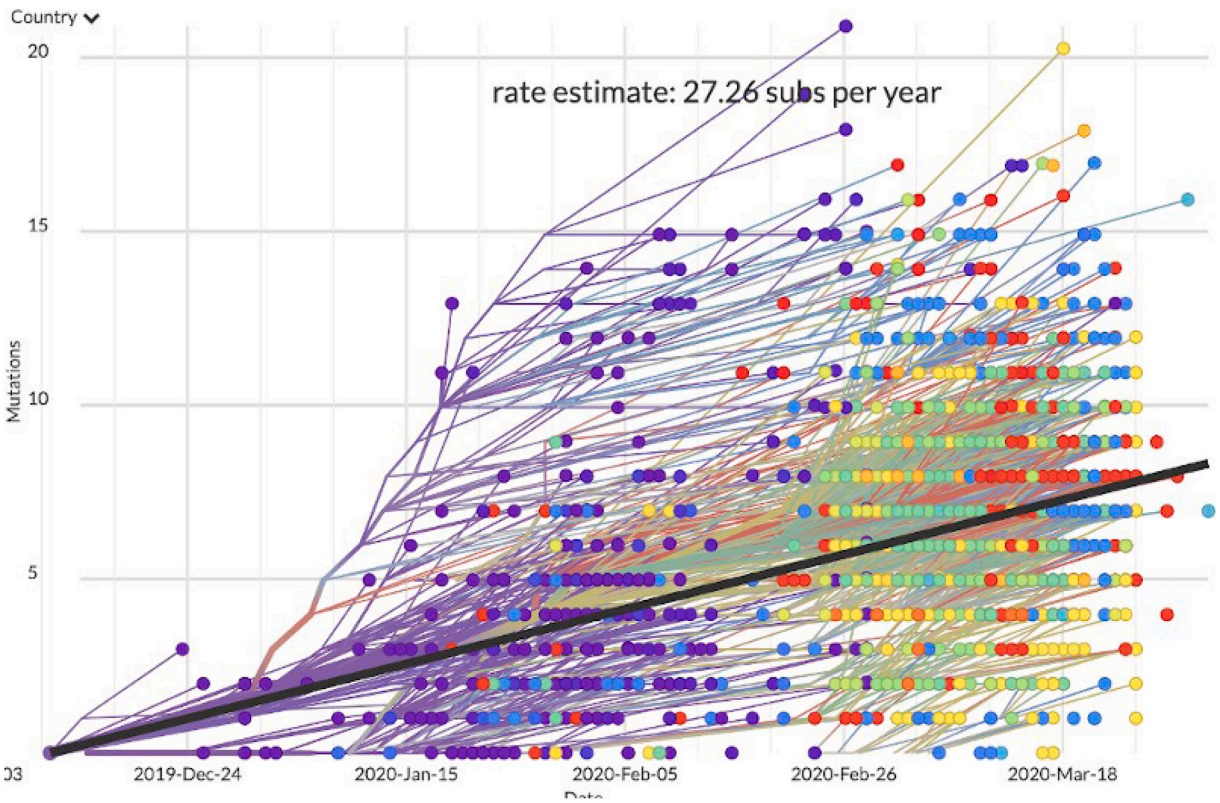

Fig. 1. COVID-19 infection rate and diversity based on geographical location.

subunit of S protein, the major CoV vaccine target [44]. Furthermore, immune-informatics approaches have also been employed in the discovery of candidate epitopes as vaccine targets [45]. Vaccine candidates in clinical trials include INO-4800, mRNA-1273 and ChAdOx1 nCoV-19 [18–20].

Over the years, genome sequencing (whole and targeted) has evolved as a viable and efficient route for drug discovery. It has been employed in understanding the molecular basis of the pathogenesis of diseases, the mechanism of drug actions/resistance, and also for the identification of chemotherapeutic targets in organisms [21]. The complete genome of the Wuhan-Hu-1 coronavirus (WHCV) a strain of SARS-CoV-2, has been sequenced. From this, a clearer insight into the molecular basis of its virulence, its establishment and main target in human hosts is being propounded.

High throughput sequencing technologies have facilitated the increase in the number of genetic biomarkers, many techniques such as protein-arrays, mass-spectrometry, and gene-expression have been employed in the discovery of genomic biomarkers. These techniques encourage the analysis of expressions and also enable the ability to determine the activity of these genes in different conditions. In this study, the RNAseq dataset of mock-treated A549 (MT_A549) cells and SARS-CoV-2 infected A549 (SI_A549) cells retrieved from the Gene expression Omnibus (GEO) database were analysed to identify differentially expressed genes between MT_A549 cells and SI_A549 cells. Furthermore, the objective of this study is to identify the dysregulated pathways between these two conditions.

2. Materials and methods

2.1. Data retrieval, processing, and differential expression analysis

We downloaded RNAseq datasets of MT_A549 cells (SRR11412249) and SI_A549 cells (SRR11412251) from the study with accession number GSE147507 [22]. In this study, transformed lung alveolar (A549) cells and human lung epithelium cells were mock treated with SARS-CoV-2. The Illumina NextSeq 500 platform was used in sequencing the RNA. Raw reads in fastq format of the MT_A549 cells and the SI_A549 cells were downloaded for our analysis. The statistical software R and encompassing packages from Bioconductor were used to analyse the MT_A549 cells and the SI_A549 cells. FastQC [23] was used to access the quality of the RNAseq data. Sequence trimming was then carried out with the aid of Trimmomatric [24]. We aligned the trimmed sequences to a reference human hg38 genome using Bowtie [25]. Furthermore, to have insight into the counts of individual genes and transcripts present in each condition, we used Htseq-count [26]. Each count was done both at the transcript level and gene level. However, genes possessing low expression were filtered out. Genes possessing negative fold change values are described as down-regulated while those possessing positive fold-change values are described as overexpressed. Identification of DEGs between the MT_A549 cells and the SI_A549 cells was carried out using DESeq2 [27].

2.2. Gene ontology, pathway enrichment analysis, and functional protein network construction

Gene Ontology is a method used in the categorization of gene expression attributes [28]. Protein ANalysis THrough Evolutionary Relationships [23] (PANTHER) is a straightforward visualization tool which is employed by researchers in the analysis of the biological functions of gene sets [29]. The 753 up-regulated genes and 746 down-regulated genes were used as input in the PANTHER server with a statistical difference set at $P < 0.05$. The PANTHER Classification System and analysis tool was therefore employed to classify the DEGs according to biological process, protein class, and molecular function, to ascertain their overrepresentation [29]. To analyse the pathways enriched in the MT_A549 cells and the SI_A549 cells, the 753 up-regulated genes and 746 down-regulated genes were used as input in the online server WebGestalt [30]. The Search Tool for the Retrieval of

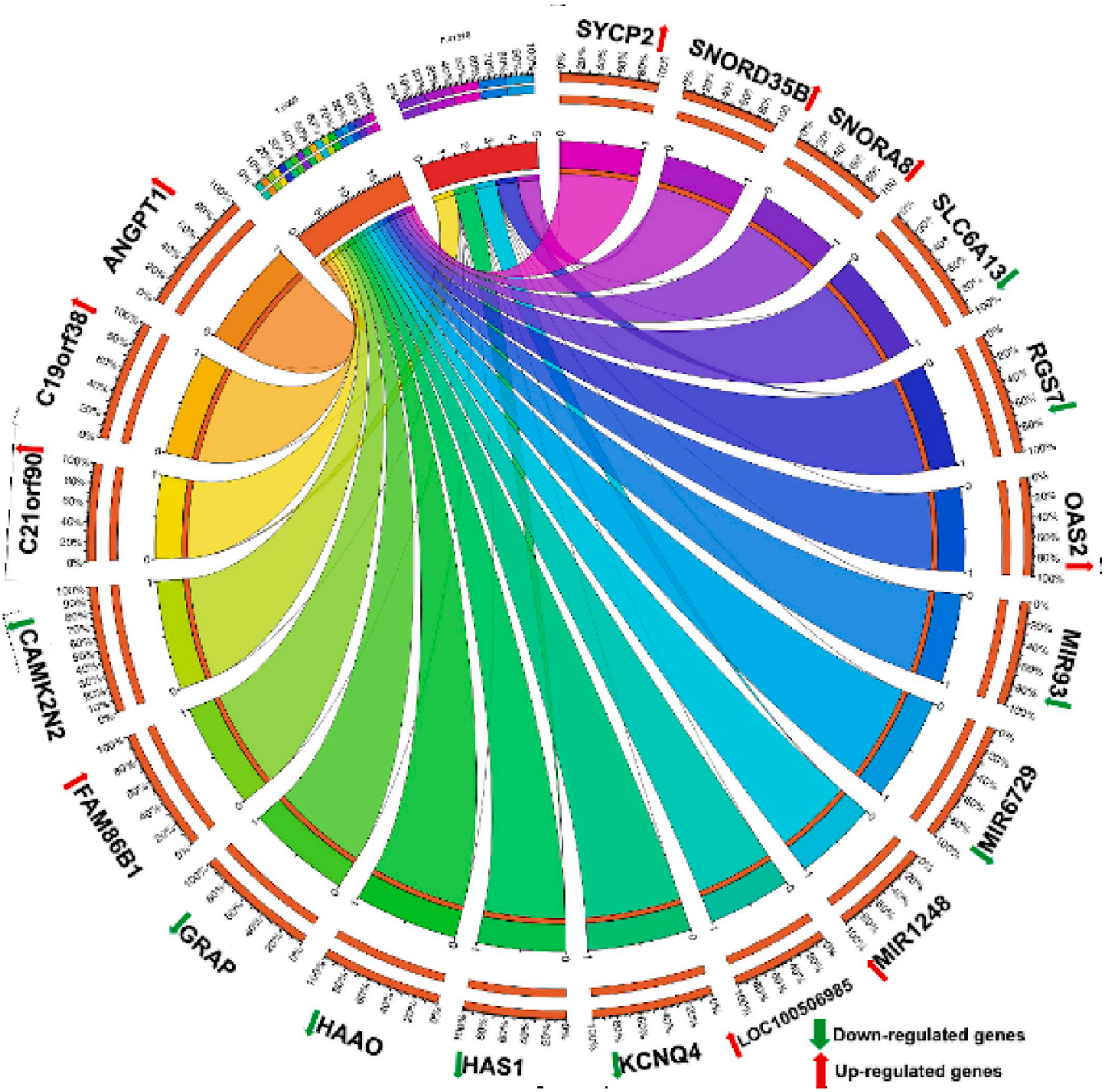

Fig. 2. Top 10 most up-regulated DEGs (**red arrow**) and 10 most down-regulated DEGs (**green arrow**) in the samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

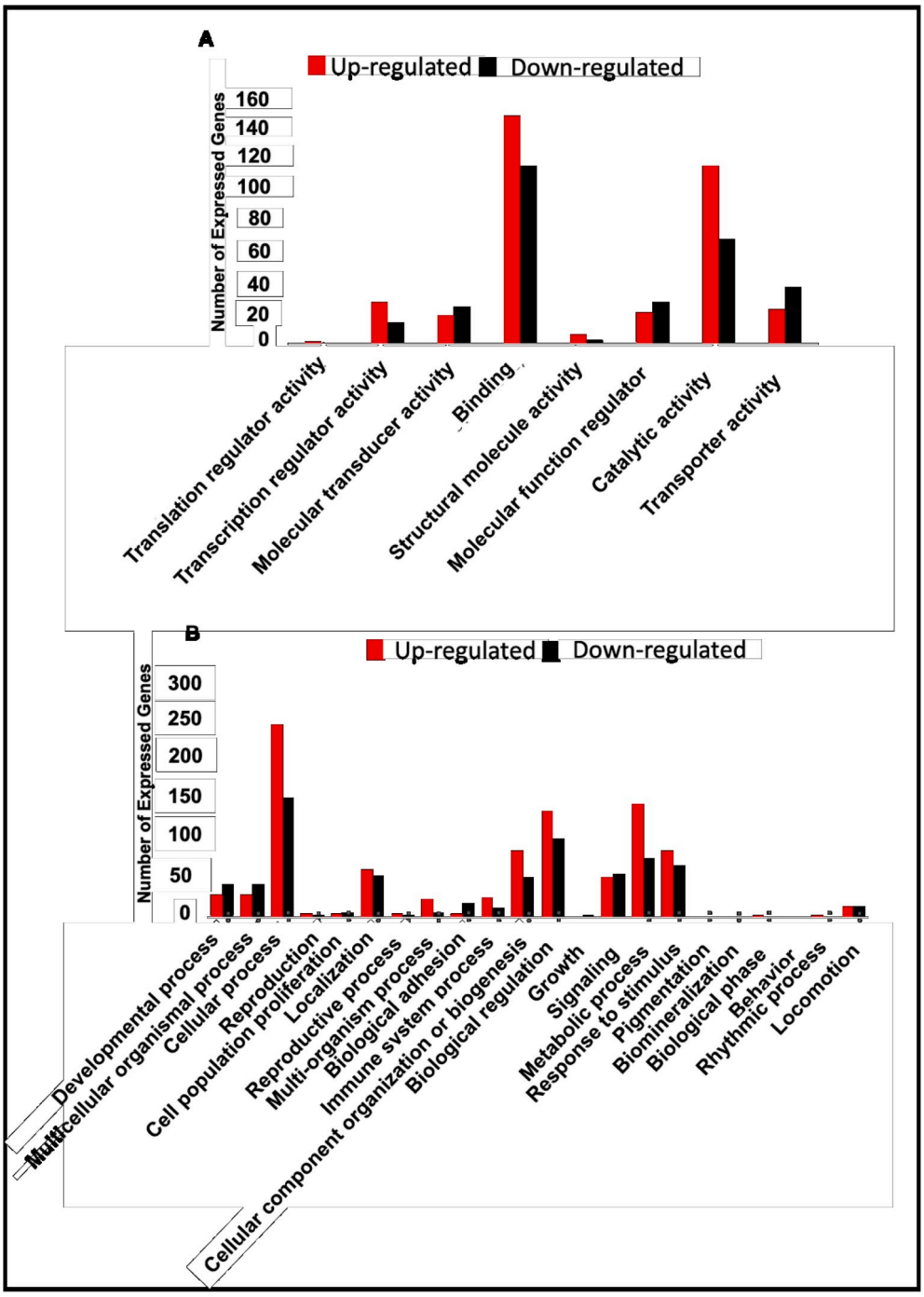

Fig. 3. Functional Classification of DEGs in MT_A549 cells and SI_A549 cells. (**A**) Molecular function ontology of overexpressed genes (red) and down regulated genes (black). (**B**) Biological process ontology of overexpressed genes (red) and down regulated genes (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Representation of DEGs according to protein class based on PANTHER classification.

Protein Class	Up-regulated genes	Down-regulated genes
Extracellular matrix protein	2	3
Cytoskeletal protein	16	12
Transporter	20	29
Scaffold/adaptor protein	10	13
Nucleic acid binding protein	30	
Intercellular signal molecule	12	16
Protein-binding activity modulator	18	19
Calcium-binding protein	5	1
Gene-specific transcriptional regulator	45	20
Defense/immunity protein	2	5
Translational protein	4	–
Metabolite interconversion enzyme	34	26
Protein modifying enzyme	40	23
Chromatin/chromatin-binding, or -regulatory protein	8	2
Transfer/carrier protein	1	–
Membrane traffic protein	11	7
Chaperone	1	
Transmembrane signal receptor	11	15
Cell adhesion	–	6
Structural protein	–	3
Cell junction	–	7

Interacting Genes (STRING) version 11.0, covering 24584628 proteins from 5090 species was employed to determine Protein-protein Interaction (PPIs). The interactions retrieved in STRING had a confidence score. During the prediction, we used experimentally validated interactions possessing a confidence score 0.4 and maximum additional interactor set at 0 to develop the protein-protein interaction network with the aid of Cytoscape software version 3.8.0 [46]. The String Enrichment plugin in Cytoscape was used to retrieve the functional enrichment. The statistically significant difference was set at $P < 0.05$.

2.3. 3D structure modeling and binding site characterization

Homology modeling is a technique used in computational biology for the determination of protein 3D structure using its amino acid sequence as building blocks [31]. It is adjudged the most accurate and effective method of computational structure prediction [31]. The protein sequence of OAS2, SYCP2L, GNB3, and ALOX12B with accession number P29728, Q5T4T6, P16520, and O75342 respectively, were retrieved from UniProt [32]. These amino acids sequences were used as input in the online template-based homology modeling server; SwissModel [33]. Validation of the built models was carried out with the aid of RAMPAGE [34], the ProSA web server [35], and Verify-3D [36]. Furthermore, after the structures were modeled, the binding sites needed to be identified; we therefore used metaPocket [37] to identify binding sites for potential drug binding sites. metaPocket is a webserver that uses different predictors (LIGSITE, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity, and POCASA) to determine a protein's binding site.

3. Results and discussion

3.1. Differential gene expression between MT_A549 cells and the SI_A549 cells

The DESeq2 package was employed to determine DEGs in the MT_A549 cells and the SI_A549 cells. At the Trimming stage, 7,461,512 (94.42%) reads of the (infected sample) passed the trim filter, 469, 643 (5.58%) were removed and 7,461,512 (99.11%) were aligned. For the mock-treated samples, 2440, 387 (89.20%) of the total reads passed the trim filter; out of this, 99.12% were aligned. At a P-value of 0.05 and 1% FDR ($Q < 0.01$), a total of 17179 transcripts and genes were differentially expressed between MT_A549 cells and the SI_A549 cells. Out of the 17179 expressed genes, 753 were up-regulated while 746 were down-regulated. However, the analysed result indicated that the level and range of expression of the down-regulated genes and transcripts were lower than that of the overexpressed genes. The top 5 overexpressed genes are SNORA81, OAS2, SYCP2, LOC100506985, and SNORD35B while the top 5 down-regulated genes are GNB3, TMED10P1, ALOX12B, MAFG-AS1 and SAYSD1 (Fig. 2). SNORA81, and SNORD35B belong to a highly expressed group of non-coding RNAs called Small nucleolar RNAs and C/D box snoRNAs, respectively. These proteins have been implicated in the regulation of pre-mRNA splicing, polyadenylation of genes,

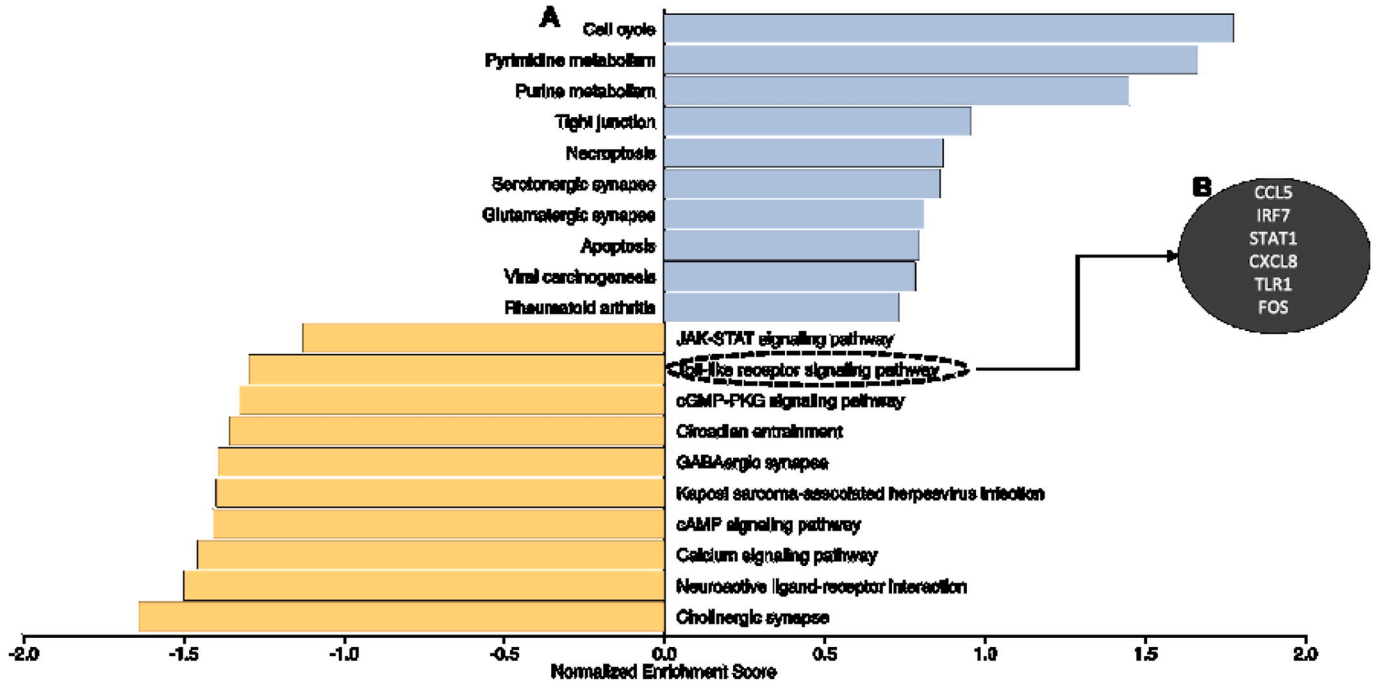

Fig. 4. Pathway enrichment analysis of the overexpressed DEGs using WebGestalt.

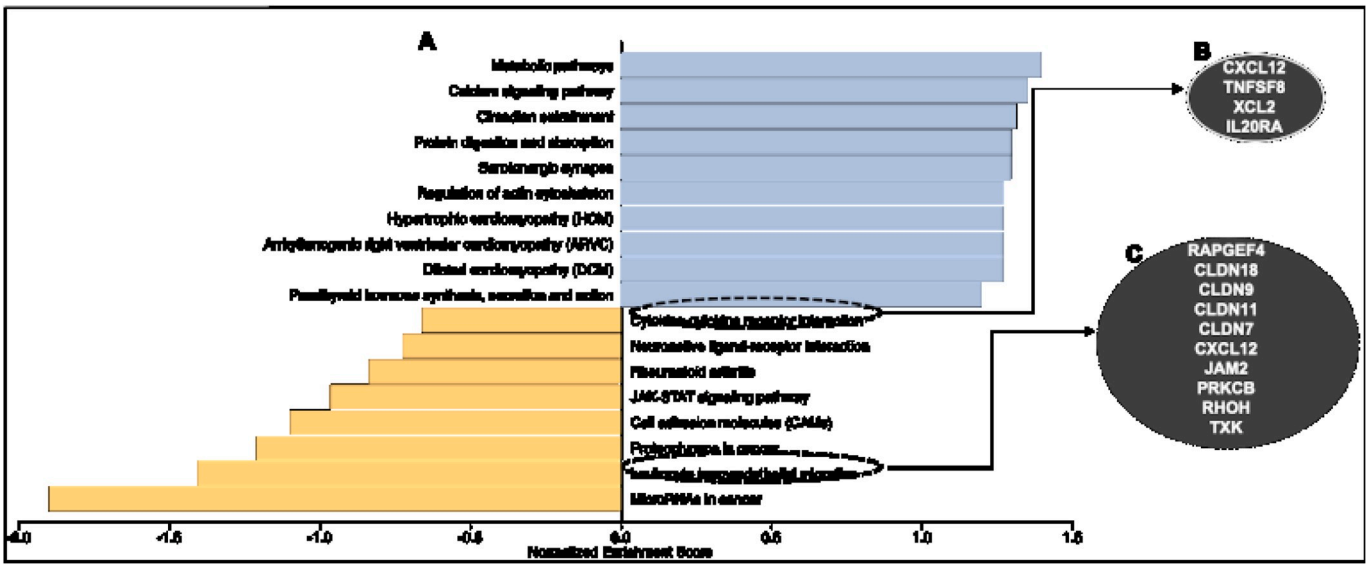

Fig. 5. Pathway enrichment analysis of the down regulated DEGs using WebGestalt.

and formation of protein complexes without the incorporation of fibrillarin [38]. Studies have shown how viruses use SNORD for host infectivity. According to research by Murray et al., 2014, using gene-trap insertional mutagenesis, 83 SNORDs and SNORAs were identified as major factors required for viral infectivity. This study suggested that the SNORDs, and not the host genes, were responsible for the replication of RNA viruses and DNA viruses [38] (see Fig. 3).

The second most upregulated gene OAS2 encodes the enzyme 2′-5′-oligoadenylate synthases 2; these enzymes are reputed for their ability to enhance intracellular antiviral mechanisms. After induction by type 1 interferons, 2′-5′-oligoadenylate synthases 2 facilitates the polymerization of ATP to 2′-5′-linked adenosine oligomers (2–5As). 2–5As subsequently activates the RNase L degradative pathway responsible for cleaving viral RNA and mitigates against infection [39].

3.2. Classification and overrepresentation analysis

To explore the DEGs at an in-depth functional level, the predicted DEGs were mapped onto the PANTHER database. Over-represented GO terms were identified in three ontologies. In the ontology of molecular function, the result indicated that the up-regulated genes were involved in majorly 6 molecular functions with the majority of the genes involved in binding (149 genes), catalytic activity (116 genes), transcription regulatory activity (27 genes), etc. Similarly, some down-regulated genes were also predicted to be implicated in catalytic activity, binding, and transcription regulatory activity. The abundance of genes in these processes is due to the fact that they contain genes that are necessary for basic life functionality. Granulocyte-macrophage colony stimulating factor is one of the genes involved in cellular process. This gene is a monomeric glycoprotein that serves as a cytokine that facilitates the development of the immune system and promotes defense against infections [40]. The cellular processes include cell communication, cell cycle, gene expression, protein targeting, metabolism, etc. Some genes classified under metabolic processes are involved in energy-producing pathways. These genes include PDK4, NLK, BATF2, ATP6AP1L, GRIK2, etc. Biological regulation describes genes that are involved in metabolism, apoptotic regulation, homeostasis, catalytic activity, and translation. The most abundant protein classification of the DEGs are Gene-specific transcriptional regulator, protein modifying enzyme, and metabolite interconversion enzyme (Table 1).

3.3. Signalling pathway analysis

To identify the dysregulated pathways on infection with SARS-CoV-2, pathway enrichment analysis was carried out with the aid of WebGestalt. 20 pathways with a P-value lower than 0.05 were enriched in the up-regulated genes while 18 pathways were enriched in the down-regulated DEGs. The Toll-like receptor (TLR) signalling pathway is one of the major pathways enriched among the DEGs (Fig. 4). TLRs play a very crucial function in the innate immune system by identifying the pathogen-associated molecular signature emanating from various microorganism [41]. Genes implicated in this pathway include C–C motif chemokine ligand 5, interferon regulatory factor 7, signal transducer and activator of transcription 1, C-X-C motif chemokine ligand 8, toll-like receptor 1 and Fos proto-oncogene (Fig. 4B). Furthermore, other pathways enriched are Leukocyte transendothelial migration (LEM), and Cytokine-cytokine receptor interaction. LEM movement from the blood vessel plays a crucial role in inflammation and immune surveillance. Leukocytes interact with endothelial cell adhesion molecules (CAM), and subsequently move across the vascular endothelium [42]. Some of the genes enriched in the cytokine-cytokine receptor interaction include C-X-C motif chemokine ligand 12, thrombopoietin, TNF superfamily member 8, X–C motif chemokine ligand 2, and interleukin 20 receptor subunit alpha (Fig. 5B).

3.4. Protein-protein interaction network analysis

PPI network analysis is crucial in the understanding of the biological responses of cells to infection. Based on the interactions result from STRING, we built a protein-protein interaction network using Cytoscape [46] (Fig. 6). A total of 810 nodes and 2300 edges were identified in the network. The STRING online database was used in the analysis of the predicted DEGs. The top 15 hub genes with the highest node degrees were ISG15, STAT1, HERC6, IRF7, HERC5, IFIH1, DDX58, GBP1, IRF9, OASL, MX1, OAS2, DDX60, OAS1, and IFIT1 (Table 2). The accuracy of the network was ascertained by the clustering coefficient (0.247), network density (0.247), and network centralization (0.063). These top 15 hub genes were investigated for betweenness centrality, MNC

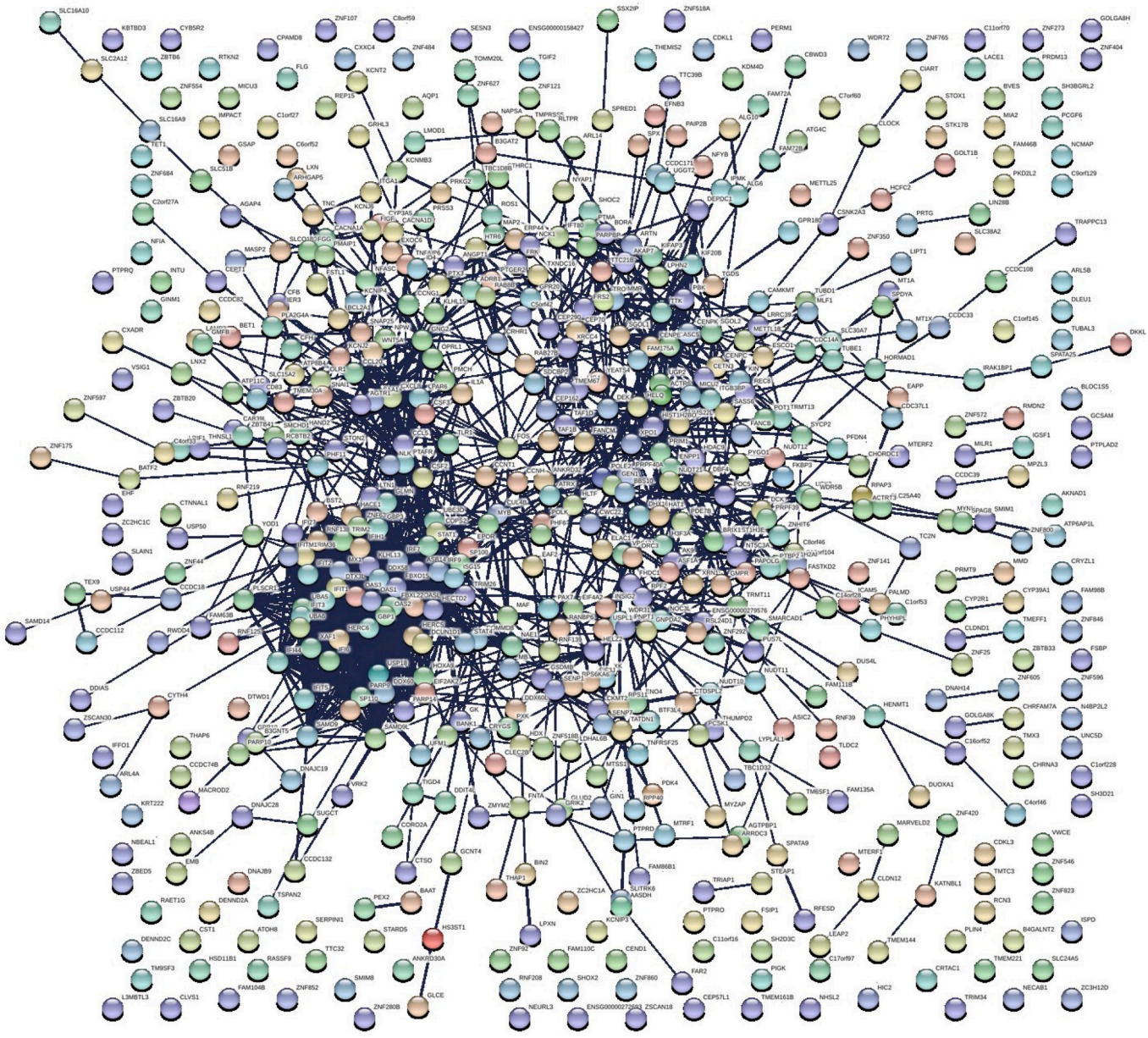

Fig. 6. Network analysis of the Differentially Expressed Genes based on betweenness centrality.

Table 2
Top 15 hub genes with highest degree on interaction.

Genes	Betweenness Centrality	Degree	Number Of Undirected Edges	Closeness Centrality
ISG15	0.0639	52	52	0.2075
STAT1	0.0530	50	50	0.2053
HERC6	0.0068	45	45	0.1841
IRF7	0.0114	45	45	0.1873
HERC5	0.0081	45	45	0.1843
IFIH1	0.0057	41	41	0.1862
DDX58	0.0135	41	41	0.1862
GBP1	0.0051	40	40	0.1853
IRF9	0.0110	40	40	0.1875
OASL	0.0073	40	40	0.1865
MX1	0.0020	40	40	0.1857
OAS2	0.0022	39	39	0.1852
DDX60	0.0041	37	37	0.1832
OAS1	0.0013	37	37	0.1843
IFIT1	9.0918E-4	37	37	0.1839

centrality, stress centrality, and degree centrality. ISG15 showed the highest node degree of 52 and betweenness centrality of 0.0639. OAS2 is strongly correlated with MX1; this connectivity could probably be due to the function of MX1 as protective genes against influenza virus infection [43] by interrupting viral ribonucleoprotein complex assembly [44].

3.5. Gene-products structure modeling

Homology modeling plays a crucial role in facilitating the drug discovery and development pipeline. We selected two representative genes from the up-regulated and down-regulated genes to be modeled. The protein sequence of OAS2, SYCP2L, GNB3, and ALOX12B with accession numbers P29728, Q5T4T6, P16520, and O75342 respectively, were retrieved from UniProt [32]. OAS2 has 719 amino acids; this sequence was used to build a model for OAS2, based on the in-built alignment algorithm in SwissModel [33], 2′-5′-oligoadenylate synthase 1 (4rwn) was used as template; there was 52.33% sequence identity between OAS2 and 2′-5′-oligoadenylate synthase 1. Similarly, this process was repeated for SYCP2L, GNB3, and ALOX12B. Synaptonemal complex

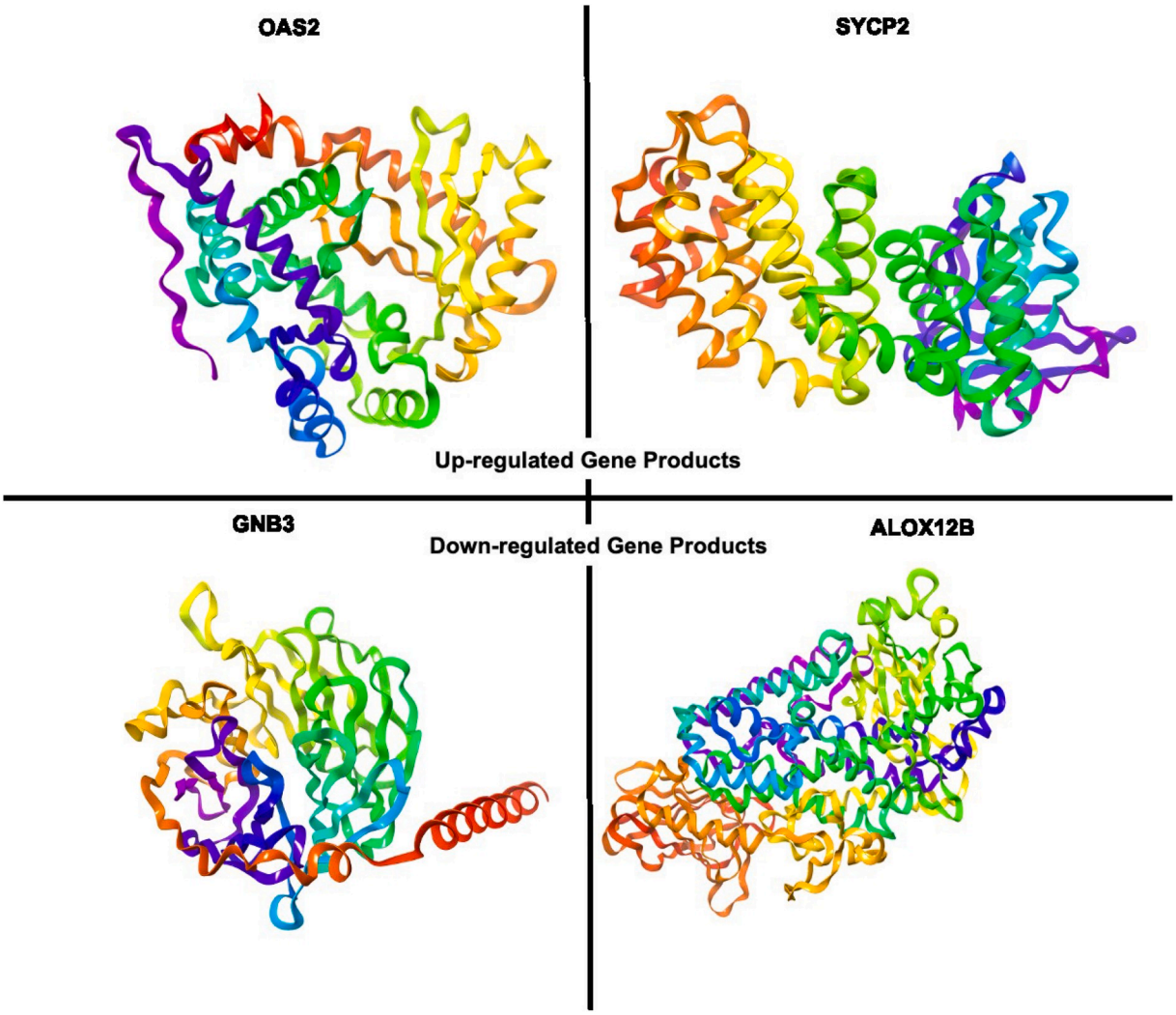

Fig. 7. Modeled structures of OAS2, SYCP2, GNB3, and ALOX12B using template-based 3D structure modeling.

Table 3
Assessment and validation of OAS2, SYC2L, GNB3, and ALOX12B using RAMPAGE and ProSA.

Genes	Favoured Region (%)	Allowed Region (%)	Outlier Region (%)	ProSA z-score
OAS2	97.1	1.7	1.2	−9.86
SYCP2L	97.1	2.1	0.8	−8.04
GNB3	94.7	3.6	1.8	−7.34
ALOX12B	94.1	4.6	1.3	−11.61

protein (5iwz) was used as the template to model SYCP2L; there was a 40.21% similarity between Synaptonemal complex protein and SYCP2L. An 83.24% sequence identity was found between Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 and GNB3. Lastly, in modeling ALOX12B, Arachidonate 15-lipoxygnase B was used as a template; these two proteins have 51.04% similarity (Fig. 7).

The quality of the selected model was validated by ProSA (z-score), Ramachandran plots, and Verify-3D. According to the RAMPAGE algorithm, 97.1% of OAS2 residues are found in the favoured region, 1.7% are in the allowed regions while 1.2% are located in the outlier region. 97.1%, 94.7%, and 94.1% amino acid residues of SYCP2L, GNB3, and ALOX12B respectively, were found in the favoured region (Table 3). Validation by ProSA revealed that the model structures are of good quality judging by the values of the z-score. (Table 3). Furthermore, the ProSA plots indicate that all of the modeled structures fall between the range of scores characteristic of proteins of similar size. Assessment of the model by Verify-3D showed that for OAS2, the profile score was 96.56%; this indicates that 96.56% of the residues have an average 3D-1D score of ≥0.2 (Fig. 8). The 3D-1D score ≥0.2 of SYCP2L, GNB3, and ALOX12B are 93.75, 88.20, and 93.71, respectively. Taken together, results obtained from these three servers suggest that the model OAS2, SYCP2L, GNB3 and ALOX12B structures are of high quality and could be used for further study.

3.5.1. Binding site characterization and potential for drug targeting

The function of a protein is dependent on its interaction with other biomolecules; these biomolecules could be called ligands [45]. The interaction between these ligands and proteins occur at a specific region of the protein called ligand binding sites; binding is facilitated by amino acid residues located in this ligand binding site [45]. Ligand binding sites have gained remarkable attention in drug-target interaction, molecular docking, drug design, etc. [46]. Hence, the identification of ligand binding sites aids in the understanding of the mechanism of intermolecular interaction between ligand and protein and hence its implication in the pathogenesis or modulation of disease condition [47]. We used metaPocket, which is a traditional machine learning-based ligand binding site server. metaPocket uses different predictors (LIGSITE, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity, and POCASA) to determine a protein's binding site. Binding site

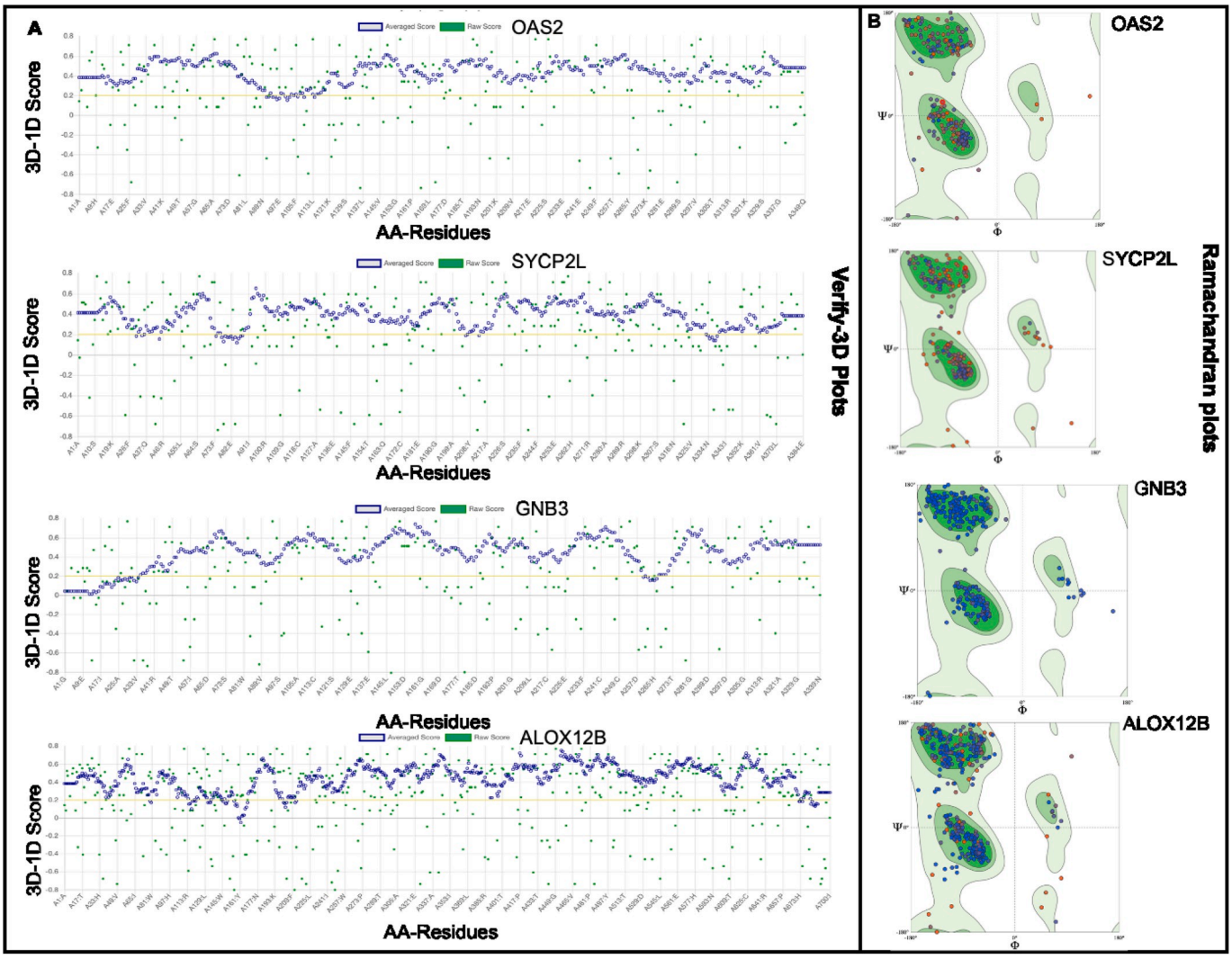

Fig. 8. Validation of OAS2, SYCP2L, GNB3 and ALOX12B model by: (A) Verify 3D and (B) Ramachandran plot.

prediction using metaPocket based on z-score and clustering highlighted two potential binding sites in the proteins. Fig. 9 highlights the two binding sites of OAS2, SYCP2, GNB3, and ALOX12B. Having highlighted these sites, they could be leveraged in the drug design and development pipeline. Most importantly, structure-based virtual screening (SBVS), SBVS is a method used in screening small molecule libraries or database for novel bioactive compounds against a certain protein [48]. This method makes use of the 3-dimensional structure, obtained from X-ray crystallography, Nuclear Magnetic Resonance, or homology modeling to dock a collection of small molecules into the binding pocket of proteins and subsequently select a subset of these compounds based on their binding pose [48]. Different online servers could be used in SBVS; for instance, when GNB3 was queried on the online Drug-Gene Interaction Database (DGIdb 3.0) [49], Lovastatin, Simvastatin, Olanzapine, Cerivastatin, and Hydrochlorothiazide were returned as potential drugs that could be used to target GNB3.

4. Conclusion

This research presents a differentially expressed gene analysis of RNAseq data of SARS-CoV-2 infected A549 cells. This expression profile was retrieved from Gene Expression Omnibus. We performed differential gene expression within the scope of gene sets and pathway analysis to explore genes and molecular-signature pathways that may be peculiar to Covid-19 infection. SNORA81, OAS2, SYCP2, LOC100506985, and SNORD35B are the top upregulated genes upon infection. To explore the DEGs at an in-depth functional level, the predicted DEGs were mapped onto the PANTHER database. Over-represented GO terms were identified in three ontologies. PPI network analysis is crucial in the understanding of the biological responses of cells to infection. The STRING online database was used in the analysis of the predicted DEGs. The top 20 overexpressed genes were investigated for betweenness centrality, MNC centrality, stress centrality, and degree centrality.

Ethical statement

This research did not involve animal model, hence ethical statement is not required.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

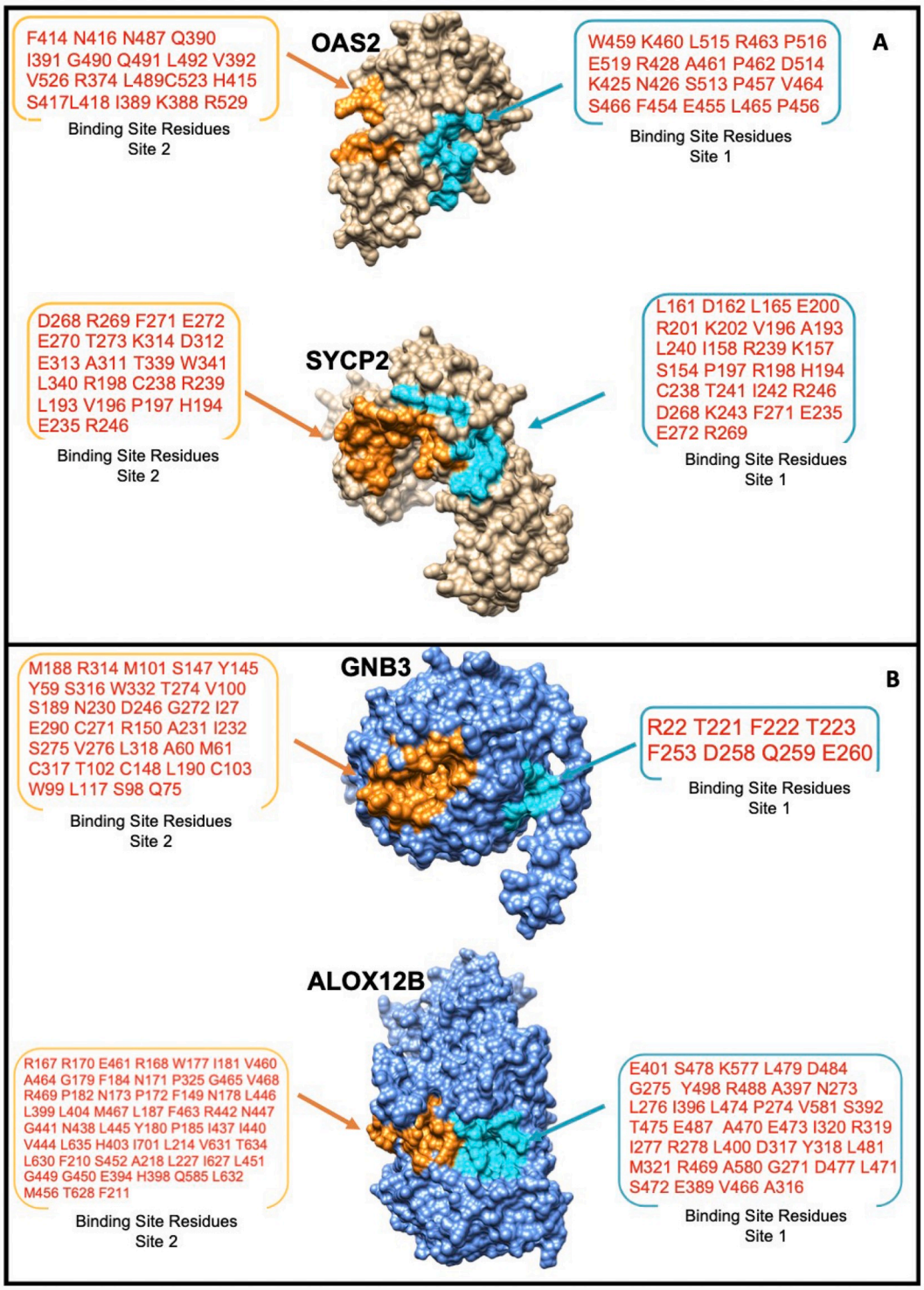

Fig. 9. Active site characterization of OAS2, SYCP2, GNB3, and ALOX12B using traditional machine learning-based ligand binding site prediction server.

the work reported in this paper.

Acknowledgement

The authors acknowledge the College of Health Sciences, UKZN, for their financial and infrastructural support, and at the same time thank the Centre for High Performance Computing (CHPC, www.chpc.ac.za), Cape Town, for computational resources.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2020.100384.

References

[1] Morfopoulou S, Brown JR, Davies EG, Anderson G, Virasami A, Qasim W, Breuer J. Human coronavirus OC43 associated with fatal encephalitis. N Engl J Med 2016,

August. https://doi.org/10.1056/NEJMc1509458. Massachussetts Medical Society.
[2] St-Jean JR, Jacomy H, Desforges M, Vabret A, Freymuth F, Talbot PJ. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. J Virol 2004;78 (16):8824–34. https://doi.org/10.1128/JVI.78.16.8824-8834.2004.
[3] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Tan W. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020;382(8):727–33. https://doi.org/10.1056/NEJMoa2001017.
[4] Woo PCY, Lau SKP, Lam CSF, Lau CCY, Tsang AKL, Lau JHN, Yuen K-Y. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavi. J Virol 2012;86(7):3995–4008. https://doi.org/10.1128/jvi.06540-11.
[5] Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology. Blackwell Publishing; 2018, February. https://doi.org/10.1111/resp.13196.
[6] Mahase E. Coronavirus : covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate, 641; 2020. p. 2020. https://doi.org/10.1136/bmj.m641. February.
[7] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Feng Z. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med 2020;382(13):1199–207. https://doi.org/10.1056/NEJMoa2001316.
[8] Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, Yuen KY. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020;395(10223):514–23. https://doi.org/10.1016/S0140-6736(20)30154-9.
[9] Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, Yan Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020;7(1):11. https://doi.org/10.1186/s40779-020-00240-0.
[10] Agostini ML, Andres EL, Sims AC, Graham RL, Sheahan TP, Lu X, Denison MR. Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. mBio 2018;9(2). https://doi.org/10.1128/mBio.00221-18.
[11] Gautret P, Lagier J-C, Parola P, Hoang VT, Meddeb L, Mailhe M, Raoult D. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents 2020:105949. https://doi.org/10.1016/j.ijantimicag.2020.105949.
[12] Furuta Y, Komeno T, Nakamura T. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. In: Proceedings of the Japan academy series B: physical and biological Sciences. Japan Academy; 2017. https://doi.org/10.2183/pjab.93.027.
[13] Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, Stebbing J. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. The Lancet. Lancet Publishing Group; 2020, February. https://doi.org/10.1016/S0140-6736(20)30304-4.
[14] Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Pillai SK. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020;382(10):929–36. https://doi.org/10.1056/NEJMoa2001191.
[15] Lim J, Jeon S, Shin HY, Kim MJ, Seong YM, Lee WJ, Park SJ. Case of the index patient who caused tertiary transmission of coronavirus disease 2019 in Korea: the application of lopinavir/ritonavir for the treatment of COVID-19 pneumonia monitored by quantitative RT-PCR. J Kor Med Sci 2020;35(6). https://doi.org/10.3346/jkms.2020.35.e79.
[16] Clinical management of severe acute respiratory infection when COVID-19 is suspected. ([n.d.]).
[17] Chen L, Xiong J, Bao L, Shi Y. Convalescent plasma as a potential therapy for COVID-19. The lancet infectious diseases. Lancet Publishing Group; 2020, April. https://doi.org/10.1016/S1473-3099(20)30141-9.
[18] Inovio collaborating with Beijing advaccine to advance INO-4800 vaccine against new coronavirus in China. ([n.d.]).
[19] NIH clinical trial of investigational vaccine for COVID-19 begins | National Institutes of Health (NIH). ([n.d.]).
[20] A study of a candidate COVID-19 vaccine (COV001) - full text view - ClinicalTrials.gov. ([n.d.]).
[21] Beeley LJ, Duckworth DM, Southan C. 1 the impact of genomics on drug discovery. Prog Med Chem 2000;37(C):1–43. https://doi.org/10.1016/S0079-6468(08)70056-0.
[22] Blanco-Melo D, Nilsson-Payant B, Liu W-C, Moeller R, Panis M, Sachs D, tenOever BR. SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. bioRxiv; 2020. https://doi.org/10.1101/2020.03.24.004655. 2020.03.24.004655.
[23] Andrews S. FastQC 1 . 1 what is FastQC 2 . Basic operations 2 . 1 opening a sequence file. Retrieved from, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; 2010.
[24] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.
[25] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
[26] Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31(2):166–9. https://doi.org/10.1093/bioinformatics/btu638.
[27] Michael A, Ahlmann-eltze C, Anders S, Huber W. Package ' DESeq2. 2020.
[28] Gene T, Consortium O. The gene ontology Resource : 20 years and still GOing strong, vol. 47; 2019. p. 330–8. https://doi.org/10.1093/nar/gky1055. November 2018.
[29] Mi H, Muruganujan A, Thomas PD. PANTHER in 2013 : modeling the evolution of gene function , and other gene attributes , in the context of phylogenetic trees, vol. 41; 2013. p. 377–86. https://doi.org/10.1093/nar/gks1118. November 2012.
[30] Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res 2019;47(W1):W199–205. https://doi.org/10.1093/nar/gkz401.
[31] Panda SK, Saxena S, Guruprasad L. Homology modeling, docking and structure-based virtual screening for new inhibitor identification of Klebsiella pneumoniae heptosyltransferase-III. J Biomol Struct Dyn 2020;38(7):1887–902. https://doi.org/10.1080/07391102.2019.1624296.
[32] Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Zhang J. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2017;45(D1):D158–69. https://doi.org/10.1093/nar/gkw1099.
[33] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003;31(13):3381–5. https://doi.org/10.1093/nar/gkg520.
[34] Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., … Richardson, D. C. (2003). Structure validation by C alpha geomF Altschul S, Gish W, Miller W, Myers E W, Lipman D J. Basic local alignment search tool. J Mol Biol 1990;50(August 2002):437–50. https://doi.org/10.1002/prot.10286. phi,psi and C beta deviation. Proteins-Structure Function and Genetics.
[35] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35 (SUPPL.2):407–10. https://doi.org/10.1093/nar/gkm290.
[36] Kresge CT, Leonowicz ME, Roth WJ, Vartuli JC, Beck JS. [2]»ÉøÍ,ÐÔµÃ © 19 9 2 nature publishing group. Nature 1992;359:710–3.
[37] Huang B. Metapocket: a meta approach to improve protein ligand binding site prediction. OMICS A J Integr Biol 2009;13(4):325–30. https://doi.org/10.1089/omi.2009.0045.
[38] Murray JL, Sheng J, Rubin DH. A role for H/ACA and C/D small nucleolar RNAs in viral replication. Mol Biotechnol 2014;56(5):429–37. https://doi.org/10.1007/s12033-013-9730-0.
[39] Leisching G, Cole V, Ali AT, Baker B. International Journal of Infectious Diseases OAS1 , OAS2 and OAS3 restrict intracellular M . tb replication and enhance cytokine secretion. Int J Infect Dis 2019;80:S77–84. https://doi.org/10.1016/j.ijid.2019.02.029.
[40] Shi Y, Liu CH, Roberts AI, Das J, Xu G, Ren G, Devadas S. Granulocyte-macrophage colony-stimulating factor (GM-CSF) and T-cell responses: what we do and don't know. Cell Res 2006;16(2):126–33. https://doi.org/10.1038/sj.cr.7310017.
[41] Kawasaki T, Kawai T. Toll-like receptor signaling pathways. Front Immunol 2014;5 (SEP):1–8. https://doi.org/10.3389/fimmu.2014.00461.
[42] Schimmel L, Heemskerk N, van Buul JD. Leukocyte transendothelial migration: a local affair. Small GTPases 2017;8(1):1–15. https://doi.org/10.1080/21541248.2016.1197872.
[43] Haller O, Staeheli P, Kochs G. Interferon-induced Mx proteins in antiviral host defense. Biochimie 2007;89(6–7):812–8. https://doi.org/10.1016/j.biochi.2007.04.015.
[44] Verhelst J, Parthoens E, Schepens B, Fiers W, Saelens X. Interferon-inducible protein Mx1 inhibits influenza virus by interfering with functional viral ribonucleoprotein complex assembly. J Virol 2012;86(24):13445–55. https://doi.org/10.1128/jvi.01682-12.
[45] Chen K, Mizianty MJ, Kurgan L. ATPsite : sequence-based prediction of ATP-binding residues 2011;9(Suppl 1):1–8.
[46] Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics 2018;34(17):i821–9. https://doi.org/10.1093/bioinformatics/bty593.
[47] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.
[48] Wu CH, Arighi C, Ross K. Protein bioinformatics: from protein modifications and networks to proteomics. Methods Mol Biol 2017;1558(1558):472. https://doi.org/10.1007/978-1-4939-6783-4.
[49] Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, Griffith M. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res 2018;46(D1):D1068–73. https://doi.org/10.1093/nar/gkx1143.