Peer-reviewed version available at *Int. J. Mol. Sci.* **2019**, *20*, 6226; doi:10.3390/ijms20246226

An integrated pan-cancer analysis and structure-based virtual screening of GPR15

Yanjing Wang[#1], *Xiangeng Wang*[#1], *Yi Xiong*[1], *Cheng-Dong Li*[1], *Qin Xu*[1], *Lu Shen*[2], *Aman Chandra Kaushik**[1], *Dong-Qing Wei**[1]

[1]State Key Laboratory of Microbial Metabolism and School of life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, PR China

[2]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200030, PR China

*Corresponding author:

Professor: Dong-Qing Wei

Email: dqwei@sjtu.edu.cn

[#]contributed equally

Abstract

G protein-coupled receptor 15 (GPR15, also known as BOB) is an extensively studied orphan GPCR involving HIV infection, colonic inflammation and smoking-related diseases. Recently, GPR15 was deorphanized and its corresponding natural ligand demonstrated an ability of inhibiting cancer cell growth. However, no study reported the potential role of GPR15 in a pan-cancer manner. Using large-scale publicly available data from the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases, we found that *GPR15* expression is significantly lower in colon adenocarcinoma (COAD) than in normal tissues. And among 33 cancer types, *GPR15* expression is significantly correlated with the prognoses of COAD, neck squamous carcinoma (HNSC), lung adenocarcinoma (LUAD) positively and stomach adenocarcinoma (STAD) negatively. This study also revealed that commonly upregulated gene set in the high GPR15 expression group (stratified via median) of COAD, HNSC, LUAD and STAD are enriched in immune systems, indicating that GPR15 might be considered as a potential target for cancer immunotherapy. Furthermore, we modelled the 3D structure of GPR15 and conducted

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 6226; doi:10.3390/ijms20246226

the structure-based virtual screening. The top 8 hits compounds were screened and then subjected to molecular dynamics (MD) simulation for stability analysis. Our study provided novel insights into the role of GPR15 in a pan-cancer manner and discovered a potential hit compound for GPR15 antagonists.

Keywords: *GPR15, pan-cancer, TCGA, cancer immunity, differential gene expression, prognosis, virtual screening*

1. INTRODUCTION

G protein-coupled receptors (GPCRs), also known as seven-transmembrane domain receptors, constitute the largest family of cell signaling receptors[1]. GPCRs respond to a wide range of extracellular signals and regulate various cellular and physiological processes, including hormone regulation, vision, immune responses, neuronal communication, and behavior[2]. Overwhelming evidences have demonstrated that GPCRs and their downstream signaling targets play critical roles in cancer initiation and progression by regulating signal transduction and cellular processes (including cell proliferation, apoptosis, stress signals, immune escape, invasion, angiogenesis and metastasis, ion and nutrient transport and migration)[3, 4]. It has also been demonstrated that diverse GPCRs were overexpressed in a variety of tumors[5]. Both orphan and well-characterized GPCRs have been reported to be involved in cancer development[6-10], which provide opportunities for developing new strategies of cancer prevention and treatment. At present GPCR-targeted drugs for cancer treatment are still few. The limited concrete knowledge about the role of GPCRs in cancers might be the cause of this lack of GPCR-targeted drugs as a treatment for cancer.

G protein-coupled receptor 15 (also known as BOB) is an extensively studied orphan GPCR[11]. It is a chemokine co-receptor of human immunodeficiency virus type 1 and 2[12] and a meditator of homing control in the large intestine and skin[8]. Dozens of studies have demonstrated the significant association between GPR15 and immune system. For example, GPR15 was found to be expressed in memory B cells, plasmablasts and regulatory T cell subsets[13, 14]. It directs T cell homing to the developing epidermis as well as to the colon and regulates colitis[13, 15-17]. When GPR15 controls the homing of $FOXP3^{+}$ regulatory T cells (Tregs) to the large intestine lamina propria, it alleviates colonic inflammation[8]. Mounting evidences have suggested that inflammation may help the tumor cells to evade the defense from the immune system[18]. The altered expression level and epigenetic regulation of GPR15 could also have a significant influence in the health status of smokers[19-21]. Moreover, recent studies reported that GPR15 was

deorphanized and its ligand can also bind to SUSD2. The co-expression pattern of GPR15L and SUSD2 can suppress proliferation of several tumoral cell lines via G1 arrest [22-24]. This finding indicated that GPR15 may be actively involved in cancer progression. Therefore, it is necessary to characterize the role of GPR15 in carcinogenesis.

In this study, we firstly perform pan-cancer analysis to elucidate the potential role of GPR15 in cancers. Earlier studies using similar methods have been published to provide new insights of specific genes in carcinogenesis[25-28]. The expression levels of *GPR15* were evaluated in 33 different cancers using the data from the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases. The function of *GPR15* were predicted by integrated network analysis. Our study identified a number of common genes that are in the GPR15 regulatory network in four cancer. We provide evidences that GPR15 acts as an immunomodulator and can be considered as a novel target for immunotherapy for four cancers. We also predicted the 3D structure of human GPR15 and apply structure-based virtual screening (SBVS) approaches[29-33]to discover potential antagonists that bind to the predicted active site. These results could help us understand the role of GPR15 in carcinogenesis and its future prospective for STAD drug development.

2. RESULTS

2.1 Pan-cancer mutational and expression landscape of *GPR15*

Among all of the 33 cancer types, *GPR15* is most frequently mutated in rectal adenocarcinoma (READ), uterine carcinosarcoma (UCS), lung squamous carcinoma (LUSC), colon adenocarcinoma (COAD), and uterine corpus endometrial carcinoma. Cancers with lower *GPR15* expression (on the basis of difference of median expression between cancer samples and paired normal samples) include COAD, bladder urothelial carcinoma (BLCA), esophageal carcinoma (ESCA), READ, and head and neck squamous carcinoma (HNSC). Among all of the cancer types, COAD shows the significant lower expression in tumor tissue compared to healthy tissues from the Genotype-Tissue Expression (GTEx) project[34]. It is worth noting that GPR15 in COAD is both hypermutated and significantly downregulated compared to that in normal tissues. This pattern implies that alterations in *GPR15*-meditor T-cell homing[8] may have undiscovered effects on the pathophysiology of COAD. The mutational and expression landscape of *GPR15* in TCGA cohorts are shown in **Figure 1**.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at *Int. J. Mol. Sci.* **2019**, *20*, 6226; doi:10.3390/ijms20246226

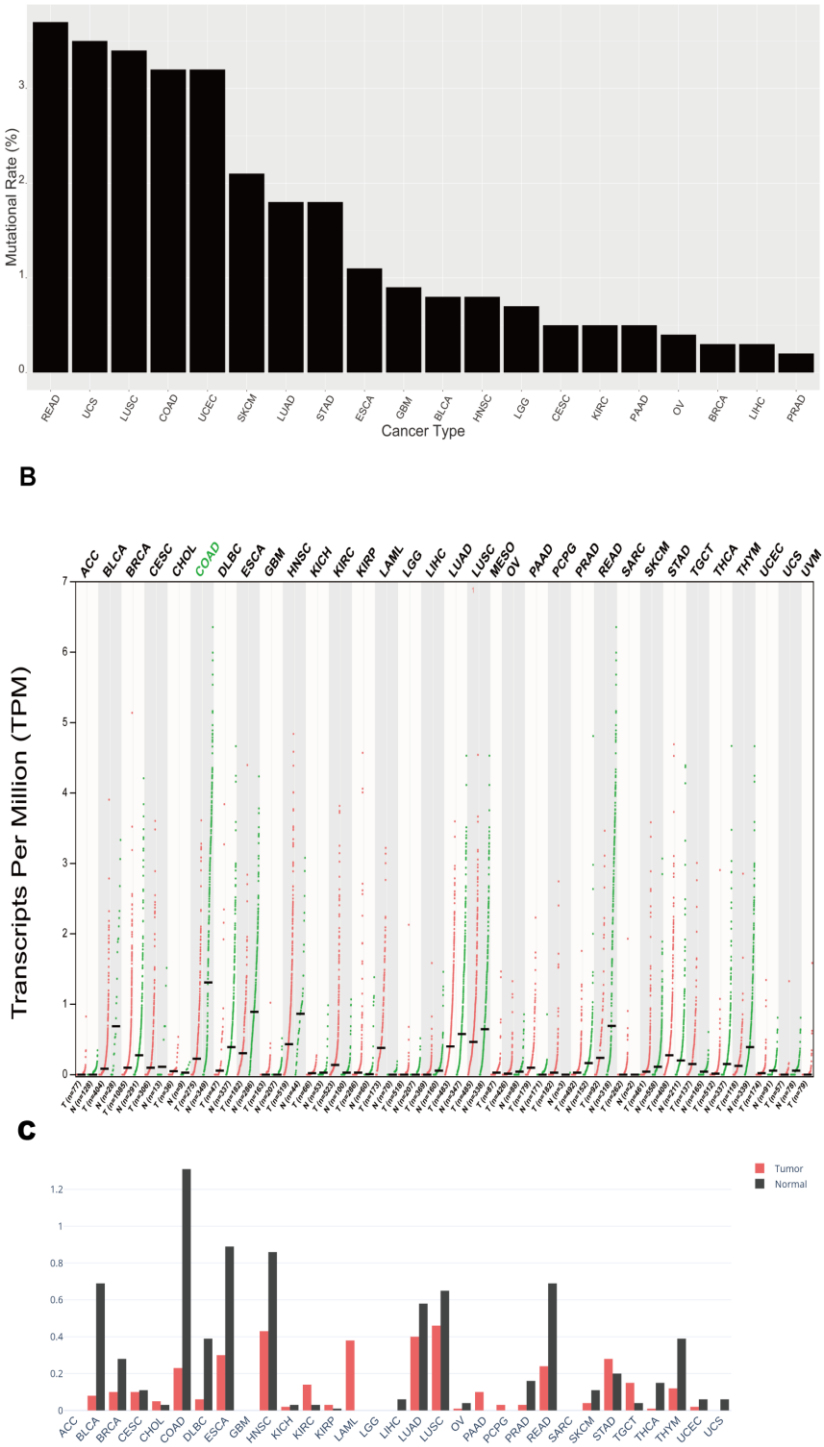

Figure 1. Expression and mutational landscape of GPR15 in TCGA cohorts. A. Barplot of the mutational rate of GPR15 in 20 cancer types and cancer types whose GPR15 mutational rate is 0 are excluded. **B.** Pan-cancer expression landscape of GPR15. "T" stands for tumor tissue and "N" stands for paired normal tissue. The expression abundance is measured by log-normalized transcripts per million (TPM). The green color of cancer type means that GPR15 is differentially expressed between tumor tissue

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 6226; doi:10.3390/ijms20246226

of bar represents the median expression (log-normalized TPM) of certain tumor type or normal tissue.

2.2 Integrated network analysis of *GPR15*

To obtain more functional insights for *GPR15*, we performed integrative network analysis on GPR15[35] as shown in **Figure 2**. The network was built upon co-expression, physical interaction, genetic interaction, shared protein domains, and pathway data, where we found that the most related protein is *YWHAB*, and the linkage is supported by direct physical interactions. *YWHAB* encodes the protein 14-3-3 protein beta/alpha, which plays a role in mitogenic signaling and cell cycle machinery[36]. Integrated network analysis revealed that, apart from immunity control, GPR15 may have effects on cell growth, thereby affecting carcinogenesis. Top 5 GPR15-related genes with highest scores are shown in **Table 1**.

Pathway analysis was conducted on the top 50 genes in the network to illustrate their biological function by Reactome platform. We found the Butyrate Response Factor 1 (BRF1) binding and Tristetraprolin (TTP, ZFP36) binding are the most two significant pathways. Both pathways involve *YWHAB*, which further implies the close interaction between GPR15 and *YWHAB*. Top 5 most related pathways are shown in **Table 2 and Table S1**.

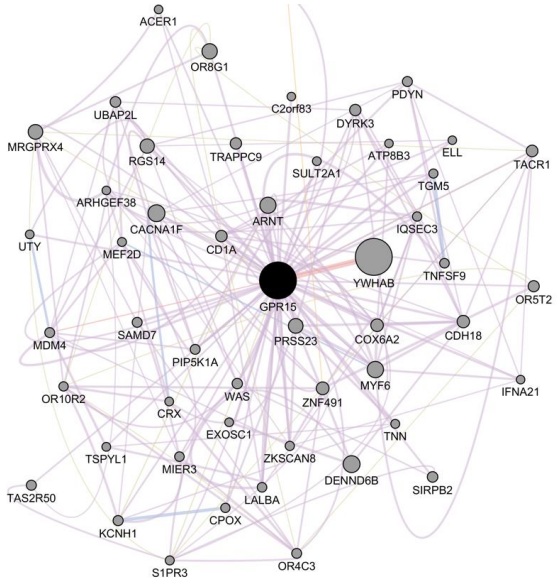

Figure 2. The integrated network of *GPR15*. The edge color represents supporting data. Pink means the line is based on physical interactions. Purple means the line is generated from co-expression profiles. Blue means co-localization, and yellow stands for predicted interaction. Node size stands for its weight in the network.

Table 1. Top 5 GPR15-related genes with highest scores

Gene	Score	Network group	Network resource
YWHAB	0.22097643	Physical Interactions	BioGRID-small-scale-studies[37]
YWHAB	0.14404532	Physical Interactions	IREF-INTACT[38]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 6226; doi:10.3390/ijms20246226

TAS2R9	0.034424774	Co-expression	Hannenhalli-Cappola-2006[40]
SPDYE4	0.031733938	Co-expression	Coelho-Hearing-2015[41]
GPR182	0.029303862	Co-expression	Scholtysik-Kuppers-2015[42]

Table 2. Top 5 pathways of genes in integrated network.

Pathway id	Pathway name	*p*-value	entities found
R-HSA-450385	Butyrate Response Factor 1 (BRF1) binds and destabilizes mRNA	0.00231	**YWHAB**;EXOSC1
R-HSA-450513	Tristetraprolin (TTP, ZFP36) binds and destabilizes mRNA	0.00231	**YWHAB**;EXOSC1
R-HSA-525793	Myogenesis	0.00636	MYF6;MEF2D
R-HSA-375170	CDO in myogenesis	0.00636	MYF6;MEF2D
R-HSA-388396	GPCR downstream signalling	0.00967	GPR15

2.3 Pan-cancer analysis of *GPR15* expression and prognostic association

To evaluate *GPR15* expression and prognosis in pan-cancer manner, we used the pre-train multiple variate Cox regression model, which combined specific gene expression value and basic clinical data provided by OncoLnc[43] to identify the TCGA cohorts of which the prognosis is significant with the GPR15 expression value. We found that the prognoses of the four cancer types, COAD, HNSC, LUAD, and stomach adenocarcinoma (STAD), are possibly ($p \leq 0.15$) associated with GPR15 expression (**Table 3, Table S2**). In addition, based on Cox coefficients, the hazards of COAD, HNSC, and LUAD are negatively associated with GPR15 expression, whereas the expression of GPR15 has positive correlation with the hazard of STAD.

Then, we stratified the patients in each cohort based on expression median into high and low expression groups. Afterward, we built KM plots for the GPR15 group in COAD, HNSC, LUAD, and STAD separately. We found that the prognoses of COAD (p = 0.014), HNSC (p = 0.0058), LUAD (p=0.0033), and STAD (p=0.0092) are significantly correlated with GPR15 expression groups (**Figure 3 B-E**).

GPR15 can reduce the inflammation level in the large intestine by controlling T-cell homing[8]. We thus hypothesized that the high expression of GPR15 in COAD can contribute to the homing and infiltration of $FOXP3^+$ regulatory T cells (T_{regs}), which in turn boost immunity response of tumor and results in better prognosis. This pattern was also observed in HNSC and LUAD, which suggests that GPR15 may perform similar immunity control functions in head, neck, and lung tissues. However, the observed upregulation of GPR15 in STAD implies poorer prognosis, which may suggest the opposite effects of GPR15 on stomach tissue.

abundance rank among all genes.

Cancer	Cox Coefficient	*p*-value	Rank
STAD	0.27	0.002	269
HNSC	-0.205	0.006	707
LUAD	-0.161	0.039	3711
COAD	-0.159	0.150	4299
READ	-0.328	0.160	2696
LUSC	0.07	0.330	6956
KIRC	-0.059	0.480	13174
LAML	0.078	0.510	9516
ESCA	0.021	0.880	14833

2.4 Commonly upregulated gene set in high GPR15 groups of COAD, HNSC, LUAD, and STAD

To dissect the effects of the expression of *GPR15* in a genome-wide manner, we performed differential gene expression (DEG) analysis[44, 45] on the *GPR15* low expression group compared to the GPR15 high expression group in the four cancer types. We found that 357, 487, 346, and 333 genes were differentially expressed of COAD, HNSC, LUAD, and STAD, respectively (**Supplementary List**). The profiles of top 200 differential expressed genes in COAD, HNSC, LUAD, and STAD are shown in **Figure S1**. Interestingly, we found that there were 146 genes that are commonly upregulated (**Figure 3A**). These genes are defined as commonly upregulated gene set (CUPGS). This considerable number of CUPGS implies shared regulatory mechanism of *GPR15* in COAD, HNSC, LUAD, and STAD.

GO enrichment analysis of the CUPGS was conducted and the results are shown in **Figures 4A-C**. Intuitively, we found that these genes were significantly enriched in the functional category of antigen binding ($p = 2.41 \times 10^{-159}$), cellular component of immunoglobulin complex ($p = 1.32 \times 10^{-50}$), and biological process of various immunological response processes. The results of KEGG pathway analysis are shown in **Figures 4 D-F**, which also showed that the CUPGS enriched the categories of B cell receptor immunology pathway, intestinal immune network for IgA production, and primary immunology. Go enrichment results are listed in **Table 4**. The gene-concept network for CUPGS is depicted in **Figure 5**. Surprisingly, besides the established function of T-cell homing of *GPR15*, these CUPGS were also significantly associated with B-cell meditated immunity, and these genes are upregulated in the high *GPR15* expression groups, which suggests that *GPR15* exerts a broader immunological impact.

Together, the results of enrichment analysis revealed that the regulatory role of GPR15 in four cancers is strongly correlated to immunity function. But the prognosis of STAD is opposite from other three types of cancer, which implies different underlying mechanism.

Table 4. Top 10 enriched GO terms for each GO category. BP stands for biological process, CC stands for cellular component, MF stands for molecular function.

ID	Description	*p*-adjust	Category
GO:0006958	complement activation, classical pathway	2.78E-120	BP
GO:0002455	**humoral immune response mediated by circulating immunoglobulin**	3.03E-117	BP
GO:0006956	complement activation	2.21E-112	BP
GO:0072376	protein activation cascade	1.16E-107	BP
GO:0016064	immunoglobulin mediated immune response	1.36E-104	BP
GO:0019724	B cell mediated immunity	1.66E-104	BP
GO:0002429	immune response-activating cell surface receptor signaling pathway	3.22E-96	BP
GO:0006959	humoral immune response	4.60E-94	BP
GO:0002768	immune response-regulating cell surface receptor signaling pathway	8.99E-94	BP
GO:0002460	adaptive immune response based on somatic recombination of immune receptors	1.21E-91	BP
GO:0019814	**immunoglobulin complex**	1.32E-80	CC
GO:0042571	immunoglobulin complex, circulating	3.13E-77	CC
GO:0009897	external side of plasma membrane	1.50E-44	CC
GO:0072562	blood microparticle	2.80E-18	CC
GO:0098802	plasma membrane receptor complex	0.513620478	CC
GO:0042101	T cell receptor complex	0.513620478	CC
GO:0008180	COP9 signalosome	0.721923256	CC
GO:0043235	receptor complex	0.721923256	CC
GO:0000788	nuclear nucleosome	0.721923256	CC
GO:0005771	multivesicular body	0.754761177	MF
GO:0003823	**antigen binding**	2.41E-159	MF
GO:0034987	immunoglobulin receptor binding	2.28E-71	MF
GO:0004252	serine-type endopeptidase activity	2.71E-46	MF
GO:0008236	serine-type peptidase activity	9.31E-45	MF
GO:0017171	serine hydrolase activity	1.40E-44	MF
GO:0005068	transmembrane receptor protein tyrosine kinase adaptor activity	0.03233891	MF
GO:0042834	peptidoglycan binding	0.055957242	MF
GO:0031210	phosphatidylcholine binding	0.100244352	MF
GO:0050997	quaternary ammonium group binding	0.100244352	MF
GO:0035591	signaling adaptor activity	0.102967022	MF

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 6226; doi:10.3390/ijms20246226

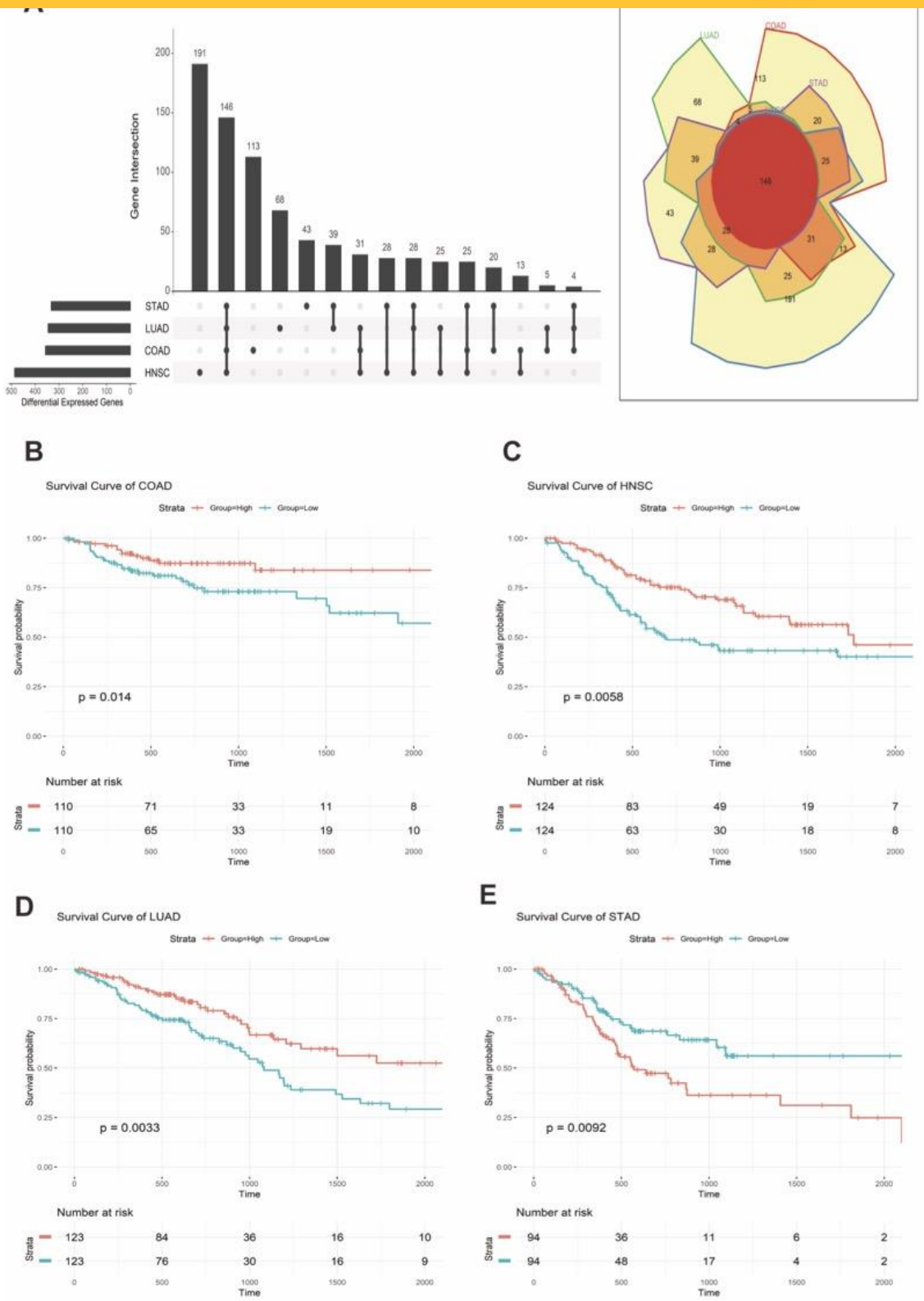

Figure 3. DEG analysis results and KM plots. **A.** Upset and Venn diagram of DEGs in COAD, HNSC, LUAD, and STAD. **B-E.** KM plots of the GPR15 expression groups in COAD, HNSC, LUAD, and STAD.

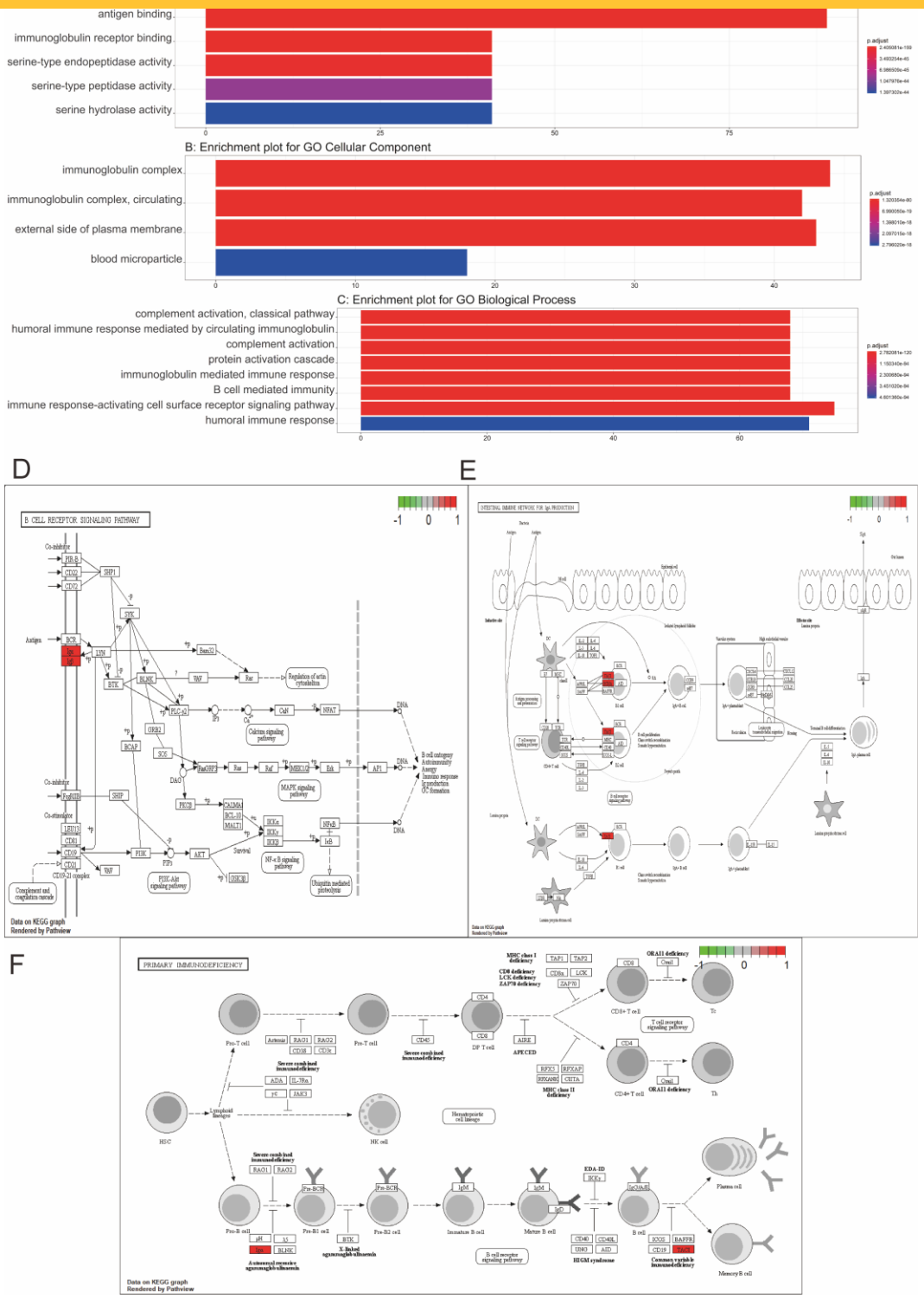

Figure 4. Enrichment results of CUPGS. A–C. GO enrichment results of molecular function, cellular component, and biological process. Color scale stands for the significance of each specific GO term. Bar length stands for the number of common altered genes within a specific GO term. **D–F.** KEGG enrichment results. Gene name in red means that the gene in a specific pathway is in the CUPGS.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at *Int. J. Mol. Sci.* **2019**, *20*, 6226; doi:10.3390/ijms20246226

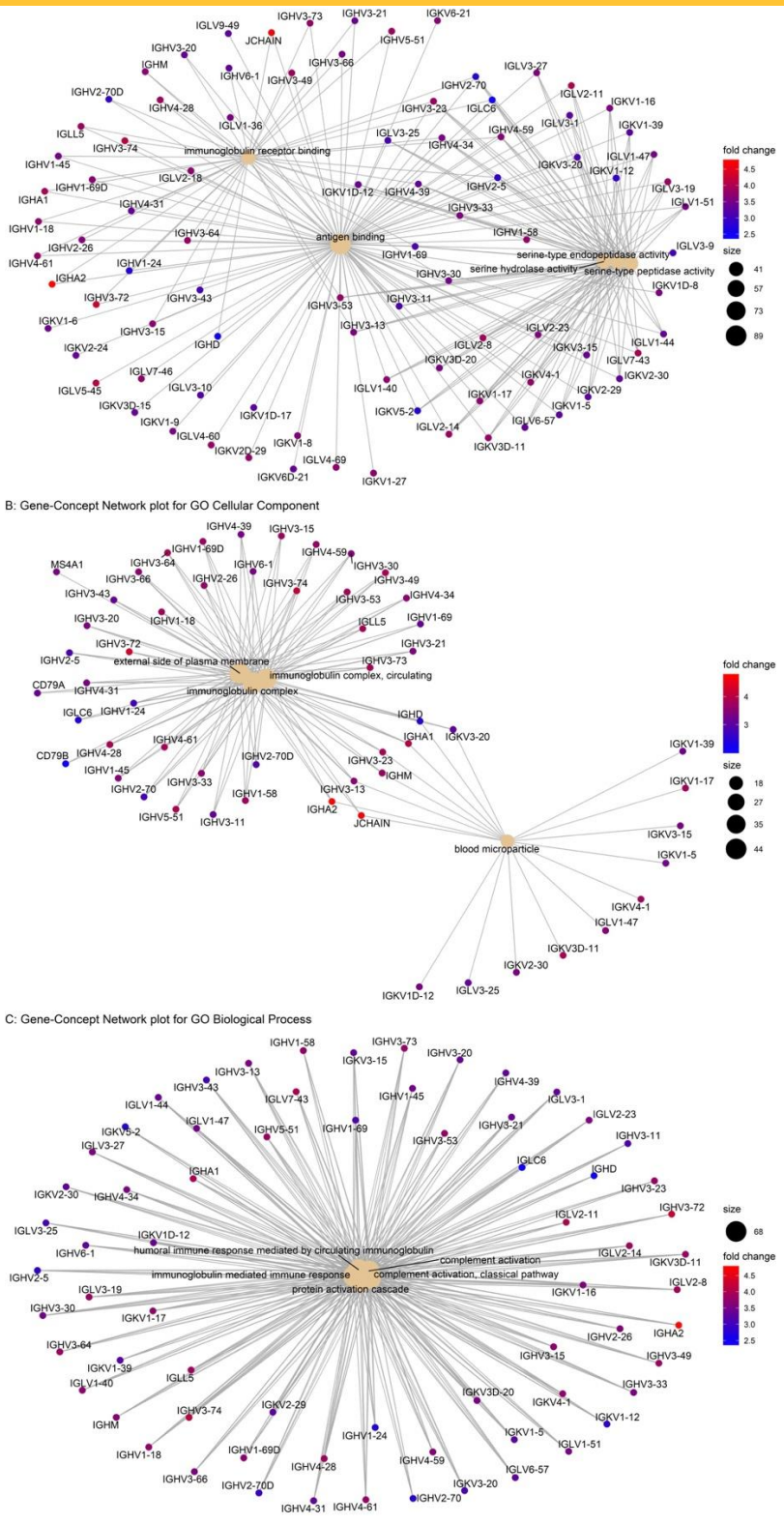

Figure 5. Gene-concept network for CUPGS. The size of the GO term stands for the number of genes in the CUPGS that is annotated based on the term. Color scale of the gene name stands for the mean fold-change in the high GPR15 expression groups compared to the low GPR15 expression groups among COAD, HNSC, LUAD, and STAD.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 6226; doi:10.3390/ijms20246226

Based on all the aforementioned analyses, we hypothesized that GPR15 could be a novel target of cancer immunotherapy. Recent studies have shown that the known natural ligands of GPR15 are all agonist [22, 23]. We thus put more emphasize on the drug discovery specifically for STAD, to identify potential inhibitors of GPR15. There is no crystal structure available for GPR15. Thus, we performed homology modeling for 3D structure of human GPR15. Template-based modeling is the most common approach to explore the relationships between three-dimensional coordinates of unknown proteins and their homologs. The GPR15 sequence was searched against the PDB-BLAST for similar template selection, and type-1 angiotensin II receptor (PDB:4YAY) were selected with a sequence identity of 32.64%, and query coverage of 30 to 317 aa (**Figure S2**). A total of 10 models were generated and further validated by SAVE server. The best predicted model structures were further refined by calculation of probability density function (pdf) and discrete optimized potential energy (DOPE). The 3D model had DOPE score of -15495.15 which was lowest against the predicted other models. Also, Ramachandran plot showed 90.9% residues in the allowed region that depicted the stability of predicted model. The results of homology modelling of GPR15 is shown in **Figure 6**.

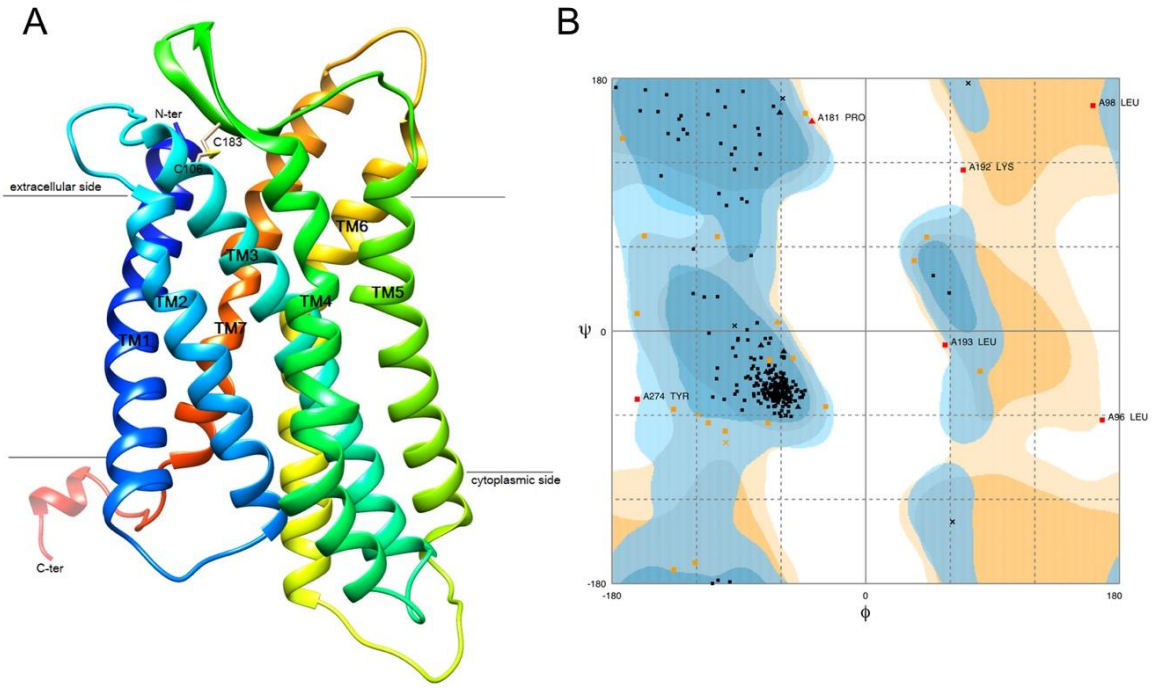

Figure 6 Homology modeling result of GPR15. (A) Predicted seven transmembrane structure (7TM) of GPR15. Different colors indicate various domains. In addition, there is a disulfide bond between Cys106 and Cys183. (B) The Ramachandran plot result of GPR15.

2.7 Structure-function relationship based binding site prediction

The structure-function relationship of GPR15 is helpful in drug design. We identified its structure-function relationship using Cofactor server[46]. We found that TRP89, SER109, ARG172, LYS180, CYS183,

TRP 193, PHE257 and LYS261 residues composed the active region in GPR15. Active site regions were largely located in extracellular regions of seven transmembrane domains where the potential leads can bind and play crucial role in signal transduction. A schematic representation of ligand binding site was shown in **Figure 7 A-B**. Also, cross-validation of the predicted residues at the active region was further supported by the results produced in Site Finder tool of MOE suite. Amino acid residues within 5 Å of active were used for the generation of receptor grid of GPR15 that was used for virtual screening.

2.8 Virtual screening and molecular docking results

We utilized the virtual screening technique to identify potential antagonists exhibiting adequate binding affinity. We started with a chemical database consisting of 62,500 small molecules and isolated a set of compounds satisfying the threshold of high docking score. After the first round of filtration, we obtained 733 compounds via shape based virtual screening. This shape-based screening approach utilizes the concept of shape of binding pockets and electrostatic potential resemblance to select new molecules, which may show similar binding modes into the binding pocket. These 733 molecules were subjected to re-docking. Docking of the selected ligands was achieved to obtain the top conformations of the selected 733 molecules into the predicted GPR15 binding site. Finally, based on the docking scores, with the threshold values fixed between −13.00 and −8.00, only top 8 screened compounds ranked based on lowest binding energy were identified as potential antagonists for GPR15. The interactions analysis for the 8 hits is given in the **Table 5** and **Figure S3**, and their 2D structures are given in **Figure S4**.

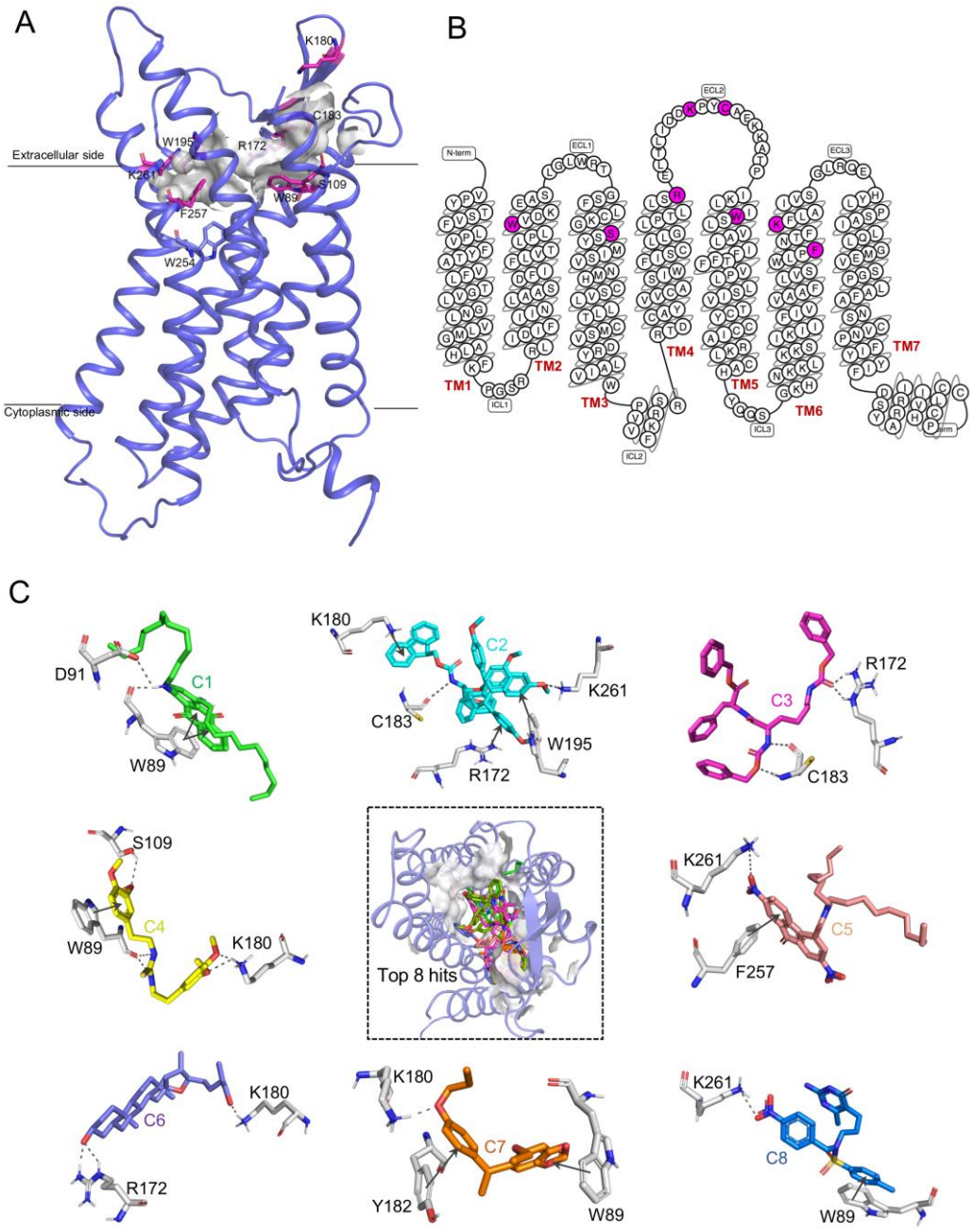

Figure 7. Ligand binding landscape of GPR15 (A) The schematic graph of predicted binding site of GPR15. The predicted binding site in our 3D structure is a traditional orthosteric binding site in the vicinity of the highly conserved residue (TRP254, $W^{6.48}$) of family A GPCRs[47]. The key residues were shown in purple red sticks. (B) The snake diagram, generated via GPCRdb[48], of predicted active side residues (purple) interacting with the ligands. (C) The predicted binding mode between GPR15 and ligands at the active site pocket (dashed box). The protein–ligand interactions of representative docking poses of top 8 hits were displayed around. Different ligands are represented by different color sticks. Hydrogen bonds are illustrated by purple lines, and Pi–pi and Pi-cation interactions are marked by black line.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 6226; doi:10.3390/ijms20246226

Table 3. The docking score and predicted protein-ligand interaction of top 8 compounds selected in virtual screening

Compound No.	Molecular formula	Weight(g/mol)	Docking score	Noncovalent interactions	Residues
C1	$C_{38}H_{58}O_2N_2$	576.91	-11.63	2 Pi-pi，2 H-bond	TRP89, ASP91
C2	$C_{60}H_{55}O_8N_1$	918.09	-11.15	1 Pi-pi, 2 Pi-cation, 1 H-Bond	LYS180, ARG172, TRP195, LYS261
C3	$C_{38}H_{41}O_7N_3$	653.77	-10.79	2 H-Bond	CYS183, ARG172
C4	$C_{21}H_{28}O_4N_2S$	404.53	-10.28	1 Pi-pi, 2 H-bond	TRP89, SER109, LYS180
C5	$C_{34}H_{47}O_6N_3$	593.76	-10.11	1 Pi-pi, 1 Salt-bridge	PHE257, LYS261
C6	$C_{27}H_{46}O_3$	418.66	-8.72	2 H-bond	ARG172, LYS180
C7	$C_{20}H_{24}O_4$	328.41	-8.3	2 Pi-pi, 1 H-bond	TRP89, TYR182, LYS180
C8	$C_{22}H_{27}O_5N_5S$	473.55	-8.29	1 Salt-bridge,1 Pi-pi	LYS261, TRP89

2.9 MD simulations and binding free energy analysis

We performed MD simulation of the top 8 potential complexes to measure the stability of the protein-ligand complex. RMSD profiles of the protein were shown in **Figure 7A**, which indicates that all systems are stable during the entire simulation run and can be used for further analysis. RMSD of ligand heavy atoms was also conducted to predict the stability of the atoms in docked complexes (**Figure 7B**). The compounds 5−8, exhibited a consistently lower value of RMSD (< 2.1Å), suggesting that these compounds formed stable complexes with GPR15. We selected 4 hits with lower ligand RMSD value for further interaction analysis and explored the ligand binding mode in protein based on occupancy of hydrogen throughout simulation time. Compound 5 showed more than 90% salt bridge interaction with Lys261 (**Figure S5**) in MD trajectories. The fluctuation in RMSD was further supported by the MM/PBSA results (**Table S3**), which shown that Compound 5 ($C_{34}H_{47}O_6N_3$) had stronger binding affinity (lowest binding free energy) among the hits with consistently lower RMSD value. Combining all the structural analyses, we found compound 5 might be promising candidate for GPR15 inhibition.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 July 2019 doi:10.20944/preprints201907.0258.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 6226; doi:10.3390/ijms20246226

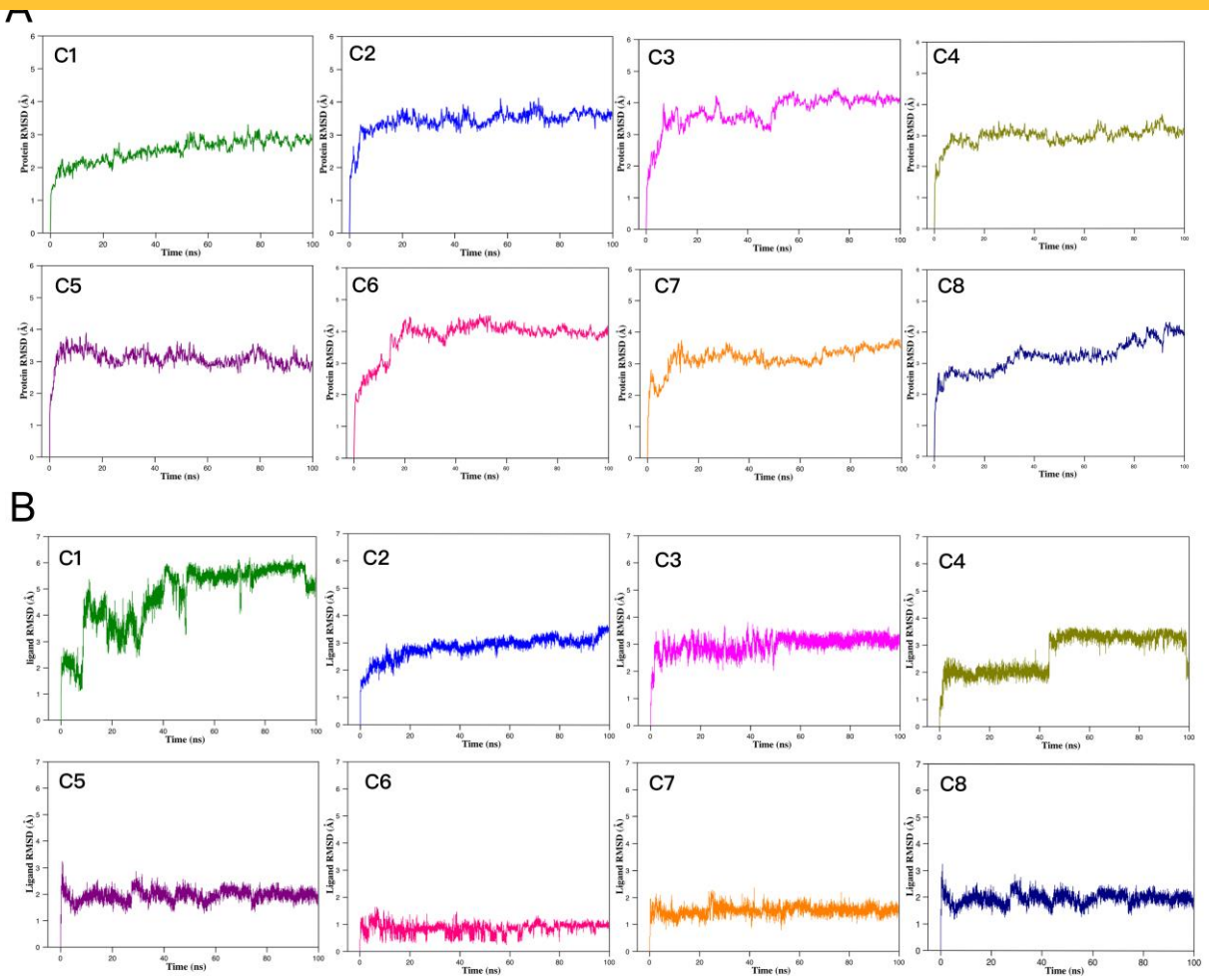

Figure 7. RMSD results of the GPR15-ligand complex systems as a function of time during the MD simulations. (A) RMSDs of backbone atoms for GPR15-ligand complexes. All the complexes are stable and attain the stability soon after reaching to 20ns. (B) RMSDs of heavy atoms for screened ligands. Compound 5, 6, 7 and 8 showed lower fluctuations among 8 hits in RMSD.

3. DISCUSSION AND CONCLSIONS

GPCRs are well-established crucial participants in various signal transduction pathways and are major fountain of targets in drug design. Until now more than 134 GPCRs are targets for drugs approved in the United States or European Union[49]. Although the endogenous ligand is not known, O-GPCRs are still popular targets with specificity in many therapeutic approaches. There is a broad range of indications linked to orphan GPCRs including cancers, thus O-GPCRs may be utilized as clinical therapeutic targets in cancer therapy[50].

In this study, we demonstrated a novel integrative pan-cancer analysis workflow and conducted a comprehensive analysis from upstream omics to downstream drug discovery of GPR15 in cancer. Our study reported GPR15 expression and mutation levels across all cancers, correlation between its expression and

cancer prognosis, investigation of genes with similar GPR15 expression patterns in COAD, HNSC, LUAD and STAD, 3D structure modeling of GPR15 to virtual screen its antagonists.

Our study provided evidence of the associations between GPR15 expression and cancer immunity. We analyzed CUPGS in COAD, HNSC, LUAD and STAD, to investigate functions of co-expression genes with similar GPR15 expression pattern. Nearly all CUPGS were enriched in immune-related function. GPR15 was proved to mediates regulatory T cells (Tregs) to migrate to the large intestine and reduce inflammation in the mouse model[8], but it is preferentially expressed on human effector T cells[13]. Further research supported GPR15-dependent human $CD8^+$ T cells can migrate into inflamed gut, and GPR15 can also help dendritic epidermal T cells migrate into skin[15]. Mounting evidence have suggested that GPR15 may play a role in the pathological process of chronic inflammatory diseases[51]. In normal tissue, inflammation may facilitate tumor cells to evade the defense from the immune system[18]. However, in the malignant cancer tissue, immune infiltration, which is often depicted as "inflammation" suggests better prognosis[52]. We found that GPR15 expression is significantly associated with the prognoses of COAD, HNSC, and LUAD positively and STAD negatively. Based on clinical and transcriptomic analyses, we can hypothesize that GPR15 could influence cancer prognosis through downstream immunological effectors.

Recent study has shown that GPR15 was deorphanized and its known natural ligands are all agonist[22, 23]. From another point of view, designing inhibitors for GPR15 could provide some clues for the treatment of STAD and help the functional study of GPR15 at the molecular level for experimental biologists. Therefore, we performed virtual screening for GPR15 antagonists and predicted protein-ligand interaction of top 8 compounds. MD simulation and free energy calculation conducted on the top 8 compounds led to the discovery of best compound 5 ($C_{34}H_{47}O_6N_3$), which could be a hit for novel drugs targeting STAD.

Together, our analysis functionally annotated GPR15 expression in a pan-cancer manner and identified potential inhibitory agents that target GPR15. Our study provided evidence of the associations between GPR15 expression and cancer immunity. Our results provide new clues of GPR15's role in carcinogenesis and new insights of cancer therapy target. Also, this novel comprehensive omics-based workflow could be utilized for hypothesis generation of new targets in cancer.

4. METHODS

4.1 Pan-cancer mutational data retrieval

We retrieved the mutation annotation format (MAF) files of the Cancer Genome Atlas (TCGA)[28] cohorts using the R package "TCGAbiolinks"[53] on October 15, 2018. The TCGA program provided us with multiple versions of somatic mutation data sets, which were generated using different workflows, and we selected the data set "MuTect2 Variant Aggregation and Masking"[54] because it encompassed more mutations than the others. Besides, according to a comparison study of mutation caller[55], MuTect2 has

the highest recall and robustness. A total number of 33 cancer types and 9914 cancer samples were included in this study.

4.2 Pan-cancer GPR15 expression profile analysis

We used GEPIA[56] to interactively analyze the expression profile of *GPR15* among 33 cancer types, 9736 tumors, and 8587 paired normal samples from the TCGA and the Genotype-Tissue Expression (GTEx) project[34]. We used the limma[45] backend with the threshold of log2 fold-change > 1 and *q*-value < 0.05 to detect cancer types exhibiting differential expressed *GPR15* compared to matched normal samples.

4.3 Integrated network analysis of *GPR15*

To obtain more functional insights on *GPR15*, we used GeneMANIA[35] and Cytoscape[57] to perform integrated network analysis. The integrated network consists of co-expression data from Gene Expression Omnibus[58] (GEO), physical and genetic interaction data from BioGRID[37], predicted protein interaction data based on orthologue using the Interologous Interaction Database[59] (I2D), and pathway molecular interaction data from BioGRID. In an algorithmic perspective, the integrated network analysis can be divided into two parts. Firstly, we used a linear regression-based algorithm that to calculate a single composite functional association network from multiple network data sources (co-expression, physical interaction, genetic interaction, shared protein domains, pathway data and so on.). Second, a variation of the Gaussian field label propagation algorithm was utilized to assign a score (the discriminant value) to each node in the network. This score reflects the computed strength of association between gene pairs[60]. We also used Reactome[61] platform to conduct pathway analysis of the predicted genes.

4.4 Survival analysis of *GPR15* expression

We used OncoLnc[43] to determine the cancer type of which the prognosis is potentially associated with *GPR15* expression. Then, we used the R package “TCGAbiolinks” to retrieve the corresponding clinical and expression data. We stratified patients in each associated cohort into “high” and “low” groups based on the median expression value of *GPR15*. The Kaplan-Meier method was utilized to estimate the survival function, and we used the log-rank test to evaluate the significance between two groups. Survival analysis and corresponding visualization were performed using the R package “survival” and “survminer”[62], respectively.

4.5 Gene differential expression analysis

We used standalone limma and voom[44] pipeline to identify differentially expressed genes associated with *GRP15* expression, comparing tumor samples with high expression of *GPR15* to low-expression ones and applied the threshold of $\log_2$ fold-change >1 and adjusted *p*-value < 0.01 to select for biological significant genes.

4.6 Commonly upregulated gene set identification and annotation

We extracted differential expressed genes (DEGs) that were upregulated and shared among COAD, HNSC, LUAD, and STAD and visualized these using the R package "UpsetR"[63] and "vennerable"[64]. These genes were collectively defined as the commonly upregulated gene set (CUPGS). We used the R package "clusterProfiler"[65] to perform Gene Ontology[66] and KEGG[67] annotation of the CUPGS.

4.7 3D structure prediction and validation of GPR15

We used BLASTP[68], which is implemented in NCBI (https://www.ncbi.nlm.nih.gov/) , to search and align the best templates for 3D structural modeling of GPR15. We used BLAST-p to align GPR15 with similar PDB structures and protein sequences retrieved from the UniProt database (https://www.uniprot.org/). A template was identified from the NCBI-BLASTp program. Homology modeling was performed using MODELLER program[69], where ten models were built through the aligned templates, and Python scripts were executed for loop modeling and model refinement. Model selection was based on the parameters of optimized loop, side-chain conformations, DOPE, Q-mean, Z-score, Maximum deviation. Structure refinement of the modeled GPR15 was performed using the KoBaMIN[70], a web server in order to obtain the best conformation of modeled structures resulting from MODELLER. Validation of the predicted model was performed using the Ramachandran Plot generated by Structure Analysis and Verification Server (SAVES) server (https://servicesn.mbi.ucla.edu/SAVES/).

4.8 Active site prediction

Structure to function relationship for the predicted GPR15 model was established by exploring the active-site residues using Cofactor server[46]. This approach was used to identify the biochemical function of the predicted GPR15 model and the potential binding region for its antagonists. The cross-validation of the predicted binding site was also conducted by Site Finder tool of MOE software.

4.9 Screening of potential compounds targeted GPR15

We screened the chemical library using the best homology model of GPR15 as receptor structure. We performed the shape-based virtual screening in first round by Surflex-Dock[71] module of SYBYL software. The receptor was optimized and prepared for virtual screening with hydrogen atoms and charges added. A library of 62,500 compounds obtained from the Maybridge Library (55,975) and in-house compound library (6,525) were used for the screening of compounds. The dataset compounds were converted to 3D coordinates and then minimized via the Powell method using 1000 iterations with a Tripose force field. We detected the surface in the predicted active sites and mapped an idealized active site ligand (called a protomol). Then, we applied five maximum conformations per fragment and five maximum poses per ligand with a 0.05Å minimum (root-mean-square deviation) RMSD to dock in the defined Protomol region. The search area was

set to 5Å in the grid. We restricted the cutoff value (total score < 0) in the docking scoring function to eliminate false positive results.

4.10 Molecular docking

Docking experiments were performed via Surflex-Dock. Top hits were selected for docking and the receptor grid box was confined around the 5 Å area of the predicted active site radiuses. The Lamarckian genetic algorithm, a well-known docking algorithm, was used to conduct docking by setting the default parameters with 150 initial populations with randomly placed individuals and the maximum number of generations set to 27,000. A shortlist based on consensus scoring function (Chem score + G Score + D Score + PMF Score) was generated. We then applied a cutoff (total score < -8) for the docking score function to eliminate false positive results. The lowest free binding energy was set as criterion for the selection of the top poses.

4.11 Molecular dynamics (MD) simulations

We selected the top 8 hits based on stability and protein-ligand interaction analyses. The initial structures for MD simulations were originated from the representative docking pose form virtual screening. The GROMACS V5.1.3 package[72] was used to perform biophysical simulation of the 8 complexes for 100,000 ps (100 ns), respectively. Membrane systems were constructed using the CHARMM-GUI Membrane Builder[73]. Proteins and lipids were presented using the CHARMM36 force field[74], and ligands were assigned to the CHARMM CGenFF[75]. All the systems were solvated in cubic water box with the distance of 10 Å between protein and a TIP3P model[76] was used for water molecules. Counter ions (0.15M NaCl) were used to keep electrically neutral of each system followed by a steepest descent energy minimization (~5000 steps). Subsequently, the minimized system was equilibrated into NVT and NPT phases for 500 ps, and all bond lengths were restrained by using the LINCS method[77] with time steps of 2 fs. The temperature was set to 310 K and the pressure was maintained at 1.01325×10^5 Pa (1 air pressure) using Langevin piston method[78], which were controlled by Nose-Hoover thermostat and Parrinello-Rahman Barostat[79], individually. Particle Mesh Ewald algorithm (PME)[80] was utilized to compel long-range electrostatic interactions, and a 1.4 nm cut off for short-range van der Waals interactions was utilized. Sampling of the MD trajectories was carried out every 2.0 ps. Finally, 100 ns MD simulations for each system were performed for further analysis. Detailed simulations conditions are listed in **Table S4**.

4.12 MD trajectories analysis

The time course of the root-mean-square deviation (RMSD) from the respective initial structures were used to assess the stability of the proteins and ligand in different simulations. Hydrogen bonds were defined with hydrogen-acceptor at a distance less than 3.5 Å and donor-hydrogen-acceptor angle with more than 135°. Salt bridges are defined by oppositely-charged atoms that are within 5 Å. All of the analyses were

performed using the analysis tools implemented in GROMACS. Totally, 100 ns trajectories (10,000 structure) of each MD system were analyzed after eliminating the rotational and translational movements. The trajectory images were visualized and analyzed with PyMol[81] and VMD[82].

4.13 Binding free energy calculations

To estimate the corresponding relative binding affinities, binding free energy for selected complexes was calculated using the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method as implemented in the g_mmpbsa tool[83], which integrates functions from GROMACS and APBS[83]. For a protein-ligand complex, the lower the binding free energy, the higher the binding affinity. The calculation is based on the following equation:

$$\begin{cases} \Delta G_{bind} = G_{complex} - G_{protein} - G_{ligand} \\ \Delta G_{bind} = \Delta E_{MM} + \Delta G_{PB} + \Delta G_{nonpolar} - T\Delta S \\ \Delta G_{bind} = G_{ele} + G_{vdw} + G_{SA} + G_{PA} \end{cases}$$

where ΔE_{MM} is the sum of van der Waals and electrostatic energy, ΔG_{PA} is polar solvation energy; ΔG_{SA} is non-polar solvation energy. The final, binding energy ΔG_{bind} is relative value rather than the absolute value because the vibrational entropy contribution ($T\Delta S$) was not included in our calculation. Total 100 snapshots at an interval of 10 ps from last 10 ns trajectories during stable phase were extracted as a sampling for calculations.

Conflicts of interest

The authors declare no conflicts of interest in this work

ACKNOWLEDGMENT

This work was supported by the funding from National Key Research Program (Contract No.2016YFA0501703), National Natural Science Foundation of China (Grant No. 31601074, 61872094, 61832019), and Shanghai Jiao Tong University School of Medicine (Contract No. YG2017ZD14).

REFERENCES

1. Rosenbaum DM, Rasmussen SG, Kobilka BK. The structure and function of G-protein-coupled receptors. Nature. 2009;459(7245):356-63. Epub 2009/05/22. doi: 10.1038/nature08144. PubMed PMID: 19458711; PubMed Central PMCID: PMCPMC3967846.

2. Deng HK, Unutmaz D, KewalRamani VN, Littman DR. Expression cloning of new receptors used by simian and human immunodeficiency viruses. Nature. 1997;388(6639):296-300. Epub 1997/07/17. doi: 10.1038/40894. PubMed PMID: 9230441.

3. Lynch JR, Wang JY. G Protein-Coupled Receptor Signaling in Stem Cells and Cancer. Int J Mol Sci. 2016;17(5). Epub 2016/05/18. doi: 10.3390/ijms17050707. PubMed PMID: 27187360; PubMed Central PMCID: PMCPMC4881529.

Curr Opin Cell Biol. 2014;27:126-35. Epub 2014/02/11. doi: 10.1016/j.ceb.2014.01.005. PubMed PMID: 24508914; PubMed Central PMCID: PMCPMC4021379.

5. Li S, Huang S, Peng SB. Overexpression of G protein-coupled receptors in cancer cells: involvement in tumor progression. Int J Oncol. 2005;27(5):1329-39. Epub 2005/10/08. PubMed PMID: 16211229.

6. Gugger M, White R, Song S, Waser B, Cescato R, Riviere P, et al. GPR87 is an overexpressed G-protein coupled receptor in squamous cell carcinoma of the lung. Dis Markers. 2008;24(1):41-50. Epub 2007/12/07. PubMed PMID: 18057535; PubMed Central PMCID: PMCPMC3850521.

7. Jin Z, Luo R, Piao X. GPR56 and its related diseases. Prog Mol Biol Transl Sci. 2009;89:1-13. Epub 2009/01/01. doi: 10.1016/S1877-1173(09)89001-7. PubMed PMID: 20374731.

8. Kim SV, Xiang WV, Kwak C, Yang Y, Lin XW, Ota M, et al. GPR15-mediated homing controls immune homeostasis in the large intestine mucosa. Science. 2013;340(6139):1456-9.

9. Prossnitz ER, Maggiolini M. Mechanisms of estrogen signaling and gene expression via GPR30. Mol Cell Endocrinol. 2009;308(1-2):32-8. Epub 2009/05/26. doi: 10.1016/j.mce.2009.03.026. PubMed PMID: 19464786; PubMed Central PMCID: PMCPMC2847286.

10. Bar-Shavit R, Maoz M, Kancharla A, Nag JK, Agranovich D, Grisaru-Granovsky S, et al. G Protein-Coupled Receptors in Cancer. Int J Mol Sci. 2016;17(8). Epub 2016/08/17. doi: 10.3390/ijms17081320. PubMed PMID: 27529230; PubMed Central PMCID: PMCPMC5000717.

11. Heiber M, Marchese A, Nguyen T, Heng HHQ, George SR, O'Dowd BF. A novel human gene encoding a G-protein-coupled receptor (GPR15) is located on chromosome 3. Genomics. 1996;32(3):462-5.

12. Blaak H, Boers PHM, Gruters RA, Schuitemaker H, Van Der Ende ME, Osterhaus A. CCR5, GPR15, and CXCR6 are major coreceptors of human immunodeficiency virus type 2 variants isolated from individuals with and without plasma viremia. Journal of virology. 2005;79(3):1686-700.

13. Nguyen LP, Pan J, Dinh TT, Hadeiba H, O'Hara E, 3rd, Ebtikar A, et al. Role and species-specific expression of colon T cell homing receptor GPR15 in colitis. Nat Immunol. 2015;16(2):207-13. Epub 2014/12/23. doi: 10.1038/ni.3079. PubMed PMID: 25531831; PubMed Central PMCID: PMCPMC4338558.

14. Seong Y, Lazarus NH, Sutherland L, Habtezion A, Abramson T, He XS, et al. Trafficking receptor signatures define blood plasmablasts responding to tissue-specific immune challenge. JCI Insight. 2017;2(6):e90233. Epub 2017/03/30. doi: 10.1172/jci.insight.90233. PubMed PMID: 28352656; PubMed Central PMCID: PMCPMC5358486 exists.

15. Lahl K, Sweere J, Pan J, Butcher E. Orphan chemoattractant receptor GPR15 mediates dendritic epidermal T-cell recruitment to the skin. Eur J Immunol. 2014;44(9):2577-81. Epub 2014/05/20. doi: 10.1002/eji.201444628. PubMed PMID: 24838826; PubMed Central PMCID: PMCPMC4165750.

16. Kim SV, Xiang WV, Kwak C, Yang Y, Lin XW, Ota M, et al. GPR15-mediated homing controls immune homeostasis in the large intestine mucosa. Science. 2013;340(6139):1456-9. Epub 2013/05/11. doi: 10.1126/science.1237013. PubMed PMID: 23661644; PubMed Central PMCID: PMCPMC3762262.

17. Habtezion A, Nguyen LP, Hadeiba H, Butcher EC. Leukocyte Trafficking to the Small Intestine and Colon. Gastroenterology. 2016;150(2):340-54. Epub 2015/11/10. doi: 10.1053/j.gastro.2015.10.046. PubMed PMID: 26551552; PubMed Central PMCID: PMCPMC4758453.

18. Nakamura K, Smyth MJ. Targeting cancer-related inflammation in the era of immunotherapy. Immunol Cell Biol. 2017;95(4):325-32. Epub 2016/12/22. doi: 10.1038/icb.2016.126. PubMed PMID: 27999432.

Expression in Relation to Smoking. Biomolecules. 2018;8(3). Epub 2018/08/22. doi: 10.3390/biom8030074. PubMed PMID: 30127295; PubMed Central PMCID: PMCPMC6163736.

20. Koks G, Uudelepp ML, Limbach M, Peterson P, Reimann E, Koks S. Smoking-induced expression of the GPR15 gene indicates its potential role in chronic inflammatory pathologies. Am J Pathol. 2015;185(11):2898-906. Epub 2015/09/09. doi: 10.1016/j.ajpath.2015.07.006. PubMed PMID: 26348578.

21. Koks S, Koks G. Activation of GPR15 and its involvement in the biological effects of smoking. Exp Biol Med (Maywood). 2017;242(11):1207-12. Epub 2017/04/21. doi: 10.1177/1535370217703977. PubMed PMID: 28423922; PubMed Central PMCID: PMCPMC5478000.

22. Ocón B, Pan J, Dinh TT, Chen W, Ballet R, Bscheider M, et al. A Mucosal and Cutaneous Chemokine Ligand for the Lymphocyte Chemoattractant Receptor GPR15. Frontiers in Immunology. 2017;8:1111.

23. Pan W, Cheng Y, Zhang H, Liu B, Mo X, Li T, et al. CSBF/C10orf99, a novel potential cytokine, inhibits colon cancer cell growth through inducing G1 arrest. Sci Rep. 2014;4:6812. Epub 2014/10/30. doi: 10.1038/srep06812. PubMed PMID: 25351403; PubMed Central PMCID: PMCPMC4212244.

24. Suply T, Hannedouche S, Carte N, Li JP, Grosshans B, Schaefer M, et al. A natural ligand for the orphan receptor GPR15 modulates lymphocyte recruitment to epithelia. Sci Signal. 2017;10(496). doi: ARTN eaal0180 10.1126/scisignal.aal0180. PubMed PMID: WOS:000410370300001.

25. Mishra S, Kaddi CD, Wang MD. Pan-cancer analysis for studying cancer stage using protein and gene expression data. Conf Proc IEEE Eng Med Biol Soc. 2016;2016:2440-3. Epub 2017/03/09. doi: 10.1109/EMBC.2016.7591223. PubMed PMID: 28268818.

26. Morris LG, Riaz N, Desrichard A, Senbabaoglu Y, Hakimi AA, Makarov V, et al. Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. Oncotarget. 2016;7(9):10051-63. Epub 2016/02/04. doi: 10.18632/oncotarget.7067. PubMed PMID: 26840267; PubMed Central PMCID: PMCPMC4891103.

27. Ng JCF, Quist J, Grigoriadis A, Malim MH, Fraternali F. Pan-cancer transcriptomic analysis dissects immune and proliferative functions of APOBEC3 cytidine deaminases. Nucleic Acids Res. 2019. Epub 2019/01/10. doi: 10.1093/nar/gky1316. PubMed PMID: 30624727.

28. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. Nature genetics. 2013;45(10):1113.

29. Chen H, Fu W, Wang Z, Wang X, Lei T, Zhu F, et al. Reliability of Docking-Based Virtual Screening for GPCR Ligands with Homology Modeled Structures: A Case Study of the Angiotensin II Type I Receptor. ACS Chem Neurosci. 2018. Epub 2018/09/29. doi: 10.1021/acschemneuro.8b00489. PubMed PMID: 30265513.

30. Varano F, Catarzi D, Falsini M, Vincenzi F, Pasquini S, Varani K, et al. Identification of novel thiazolo[5,4-d]pyrimidine derivatives as human A1 and A2A adenosine receptor antagonists/inverse agonists. Bioorg Med Chem. 2018;26(12):3688-95. Epub 2018/06/09. doi: 10.1016/j.bmc.2018.05.048. PubMed PMID: 29880250.

31. Cooke RM, Brown AJ, Marshall FH, Mason JS. Structures of G protein-coupled receptors reveal new opportunities for drug discovery. Drug Discov Today. 2015;20(11):1355-64. Epub 2015/08/26. doi: 10.1016/j.drudis.2015.08.003. PubMed PMID: 26303408.

32. Becker OM, Shacham S, Marantz Y, Noiman S. Modeling the 3D structure of GPCRs: advances and application to drug discovery. Curr Opin Drug Discov Devel. 2003;6(3):353-61. Epub 2003/07/02. PubMed PMID: 12833668.

discovery in 3D. Proc Natl Acad Sci U S A. 2004;101(31):11304-9. Epub 2004/07/28. doi: 10.1073/pnas.0401862101. PubMed PMID: 15277683; PubMed Central PMCID: PMCPMC509175.

34. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. Science. 2015;348(6235):660-5.

35. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research. 2010;38(suppl_2):W214-W20. doi: 10.1093/nar/gkq537.

36. Takihara Y, Matsuda Y, Hara J. Role of the beta isoform of 14-3-3 proteins in cellular proliferation and oncogenic transformation. Carcinogenesis. 2000;21(11):2073-7. Epub 2000/11/04. PubMed PMID: 11062170.

37. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic acids research. 2006;34(suppl_1):D535-D9.

38. Razick S, Magklaras G, Donaldson IM. iRefIndex: A consolidated protein interaction database with provenance. BMC Bioinformatics. 2008;9(1):405. doi: 10.1186/1471-2105-9-405.

39. Tateno H, Matsushima A, Hiemori K, Onuma Y, Ito Y, Hasehira K, et al. Podocalyxin Is a Glycoprotein Ligand of the Human Pluripotent Stem Cell-Specific Probe rBC2LCN. Stem Cell Transl Med. 2013;2(4):265-73. doi: 10.5966/sctm.2012-0154. PubMed PMID: WOS:000317551800004.

40. Hannenhalli S, Putt ME, Gilmore JM, Wang JW, Parmacek MS, Epstein JA, et al. Transcriptional genomics associates FOX transcription factors with human heart failure. Circulation. 2006;114(12):1269-76. doi: 10.1161/Circulationaha.106.632430. PubMed PMID: WOS:000240556700008.

41. Coelho SG, Yin LL, Smuda C, Mahns A, Kolbe L, Hearing VJ. Photobiological implications of melanin photoprotection after UVB-induced tanning of human skin but not UVA-induced tanning. Pigm Cell Melanoma R. 2015;28(2):210-6. doi: 10.1111/pcmr.12331. PubMed PMID: WOS:000352821100013.

42. Scholtysik R, Kreuz M, Hummel M, Rosolowski M, Szczepanowski M, Klapper W, et al. Characterization of genomic imbalances in diffuse large B-cell lymphoma by detailed SNP-chip analysis. Int J Cancer. 2015;136(5):1033-42. doi: 10.1002/ijc.29072. PubMed PMID: WOS:000346350500029.

43. Anaya J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. PeerJ Computer Science. 2016;2:e67.

44. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology. 2014;15(2):R29. doi: 10.1186/gb-2014-15-2-r29.

45. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. Epub 2015/01/22. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510.

46. Zhang CX, Freddolino PL, Zhang Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Research. 2017;45(W1):W291-W9. doi: 10.1093/nar/gkx366. PubMed PMID: WOS:000404427000043.

47. Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, et al. Structural basis for allosteric regulation of GPCRs by sodium ions. Science. 2012;337(6091):232-6. Epub 2012/07/17. doi: 10.1126/science.1219218. PubMed PMID: 22798613; PubMed Central PMCID: PMCPMC3399762.

structure models and ligands. Nucleic acids research. 2017;46(D1):D440-D6.

49. Sriram K, Insel PA. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? Mol Pharmacol. 2018;93(4):251-8. doi: 10.1124/mol.117.111062. PubMed PMID: 29298813; PubMed Central PMCID: PMCPMC5820538.

50. Andradas C, Caffarel MM, Perez-Gomez E, Salazar M, Lorente M, Velasco G, et al. The orphan G protein-coupled receptor GPR55 promotes cancer cell proliferation via ERK. Oncogene. 2011;30(2):245.

51. Cartwright A, Schmutz C, Askari A, Kuiper JH, Middleton J. Orphan receptor GPR15/BOB is up-regulated in rheumatoid arthritis. Cytokine. 2014;67(2):53-9. doi: 10.1016/j.cyto.2014.02.015. PubMed PMID: 24725539; PubMed Central PMCID: PMCPMC3996549.

52. Sato E, Olson SH, Ahn J, Bundy B, Nishikawa H, Qian F, et al. Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. Proceedings of the National Academy of Sciences. 2005;102(51):18538-43.

53. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic acids research. 2015;44(8):e71-e.

54. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010.

55. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. Nature communications. 2015;6:10001.

56. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45(W1):W98-W102. Epub 2017/04/14. doi: 10.1093/nar/gkx247. PubMed PMID: 28407145; PubMed Central PMCID: PMCPMC5570223.

57. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.

58. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research. 2002;30(1):207-10.

59. Brown KR, Jurisica I. Online Predicted Human Interaction Database. Bioinformatics. 2005;21(9):2076-82. doi: 10.1093/bioinformatics/bti273.

60. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. 2008;9 Suppl 1:S4. Epub 2008/07/22. doi: 10.1186/gb-2008-9-s1-s4. PubMed PMID: 18613948; PubMed Central PMCID: PMCPMC2447538.

61. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic acids research. 2013;42(D1):D472-D7.

62. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using'ggplot2'. R package version 03. 2017;1.

63. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938-40.

64. Swinton J. Vennerable: Venn and Euler area-proportional diagrams. R package version. 2011;2(0):1-33.

65. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16(5):284-7.

biology. Nature genetics. 2000;25(1):25.

67. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30.

68. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403-10.

69. Sanchez R, Sali A. Evaluation of comparative protein structure modeling by MODELLER-3. Proteins. 1997;Suppl 1:50-8. Epub 1997/01/01. PubMed PMID: 9485495.

70. Rodrigues JP, Levitt M, Chopra G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. Nucleic Acids Res. 2012;40(Web Server issue):W323-8. Epub 2012/05/09. doi: 10.1093/nar/gks376. PubMed PMID: 22564897; PubMed Central PMCID: PMCPMC3394243.

71. Jain AN. Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. J Comput Aided Mol Des. 1996;10(5):427-40. Epub 1996/10/01. PubMed PMID: 8951652.

72. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701-18. Epub 2005/10/08. doi: 10.1002/jcc.20291. PubMed PMID: 16211538.

73. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. Biophys J. 2009;97(1):50-8. Epub 2009/07/08. doi: 10.1016/j.bpj.2009.04.013. PubMed PMID: 19580743; PubMed Central PMCID: PMCPMC2711372.

74. Huang J, MacKerell AD, Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013;34(25):2135-45. Epub 2013/07/09. doi: 10.1002/jcc.23354. PubMed PMID: 23832629; PubMed Central PMCID: PMCPMC3800559.

75. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010;31(4):671-90. Epub 2009/07/04. doi: 10.1002/jcc.21367. PubMed PMID: 19575467; PubMed Central PMCID: PMCPMC2888302.

76. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics. 1983;79(2):926-35.

77. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry. 1997;18(12):1463-72. doi: Doi 10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.3.Co;2-L. PubMed PMID: WOS:A1997XT81100004.

78. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: the Langevin piston method. The Journal of chemical physics. 1995;103(11):4613-21.

79. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied physics. 1981;52(12):7182-90.

80. Darden T, York D, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics. 1993;98(12):10089-92.

81. DeLano WL. PyMOL. 2002.

82. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996;14(1):33-8, 27-8. PubMed PMID: 8744570.

MM-PBSA calculations. J Chem Inf Model. 2014;54(7):1951-62. Epub 2014/05/23. doi: 10.1021/ci500020m. PubMed PMID: 24850022.