



Pan-Cancer Analysis and Drug Formulation for GPR139 and GPR142

Aman Chandra Kaushik^{1†}, Aamir Mehmood^{2,3†}, Xiaofeng Dai^{1*} and Dong-Qing Wei^{2,3*}

¹Wuxi School of Medicine, Jiangnan University, Wuxi, China, ²School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, ³Peng Cheng Laboratory, Vanke Cloud City Phase I, Guangdong, China

GPR (G protein receptor) 139 and 142 are novel founding GPCRs (G protein-coupled receptors) in the class “A” of the GPCRs family and are suitable targets for various biological conditions. To engage these targets, validated pharmacophores and 3D QSAR (Quantitative structure-activity relationship) models are widely used because of their direct fingerprinting capability of the target and an overall accuracy. The current work initially analyzes GPR139 and GPR142 for its genomic alteration via tumor samples. Next to that, the pharmacophore is developed to scan the 3D database for such compounds that can lead to potential agonists. As a result, several compounds have been considered, showing satisfactory performance and a strong association with the target. Additionally, it is gripping to know that the obtained compounds were observed to be responsible for triggering pan-cancer. This suggests the possible role of novel GPR139 and GPR142 as the substances for initiating a physiological response to handle the condition incurred as a result of cancer.

Keywords: GPR142, molecular modeling, pharmacophore, 3D quantitative structure-activity relationship, 7TM, pan-cancer, the cancer genome atlas

OPEN ACCESS

Edited by:

Defang Ouyang,
University of Macau, China

Reviewed by:

Sudheer Kumar Ravuri,
Steadman Philippon Research
Institute, United States
Shuang-Xi Gu,
Wuhan Institute of Technology, China

*Correspondence:

Xiaofeng Dai
xiaofeng.dai@jiangnan.edu.cn
Dong-Qing Wei
dqwei@sjtu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 18 December 2019

Accepted: 23 November 2020

Published: 19 February 2021

Citation:

Kaushik AC, Mehmood A, Dai X and
Wei D-Q (2021) Pan-Cancer Analysis
and Drug Formulation for GPR139
and GPR142.
Front. Pharmacol. 11:521245.
doi: 10.3389/fphar.2020.521245

INTRODUCTION

Computational biology and immunoinformatic are the multidisciplinary fields that are progressing rapidly (Wadood et al., 2017). Besides, *in vitro*, and *in vivo* techniques are extremely demanding but still relatively challenging because of various factors such as resources, time of the experiment, experimental labor, cost, and environmental issues and safety as compared to the computational approaches. However, combining both these *in silico* and *in vitro/in vivo* techniques is highly essential for addressing a particular biological condition or targets such as the decryption of an immune response and vaccine design (Korber et al., 2006). The *in silico* drug designing, often referred to as Computer-Aided Drug Design is mainly grouped into two classes which are termed as structure-based and ligand-based. Thus, various techniques are practiced for this to computationally propose a drug or peptide for a particular biological condition. These techniques could either be for designing and discovering purposes or validation. For instance, to discover and recommend a drug for a unique or resistant viral infection, next to the background history and understanding the mechanism of action, a virtual screening (VS) approach is applied that scans various small compounds' databases (i.e. ZINC, PubChem, MAYBRIDGE, etc.) (Shemetulskis et al., 1995; Irwin and Shoichet, 2005; Kim et al., 2015) to discover those hits that are more likely to be engaged with the target. In the field of Computer-Aided Drug Design, the word “hits” is a term used for compounds that are hypothesized that they may have a stronger affinity with the target. Before this VS technique, a 3D ensemble of chemical and molecular features known as pharmacophore is designed (Schneider et al., 1999; Wadood et al., 2017). It is crucial because this 3D model can

recognize similar ligands or macromolecules from the enormous number of drugs like compounds provided in huge databases during the VS process (Sun, 2008).

Apart from VS, molecular docking (MD) (Huang et al., 2006) is a highly acceptable and practiced technique that considers the 3D conformation of a drug and its target in real-time and scores the performance of the drug along with providing various physical and chemical properties. Additionally, it explores how a drug interacts with the target that can be visualized in a 2D or 3D plane. There exists various offline and online software that are used for VS and MD. Few of the major software and servers are GOLD (Joy et al., 2006), AutoDock (Trott and Olson, 2010), PatchDock (Mehmood et al., 2019; Schneidman-Duhovny et al., 2005; Khan et al., 2018), Molecular Operating Environment (MOE) (Roy and Luck, 2007), Schrödinger suite (Bhachoo and Beuming, 2017), and FireDock (Andrusier et al., 2007). Drugs that perform satisfactorily are observed to be engaged with their targets are subjected to another technique that is termed as molecular dynamics simulation (MDS) which is also a real-time cellular system with an optimum human body pressure, temperature, water, and pH that is built inside a computer driven by a particular forcefield (FF) such as OPLS, AMBER96, GROMACS 431, etc. in software like AMBER (Junaid et al., 2019), GROMACS (Wang et al., 2019) or SCHRÖDINGER (Winstead and Ravaioli, 2003). Nevertheless, the successful outcomes of all these sophisticated techniques (MD and MDS) depend on the most crucial step in this whole process which is the selection of accurate compounds. This selection directly depends on the classical VS technique whose quality is dependent on the quality of the pharmacophore model. Similar to this, QSAR models also bears great importance. It is an approach of vital importance for chemistry and pharmacy that is created on the concept that the activity of a molecule can be altered by bringing amendments into its structural configuration. These structural amendments may be implemented for virtual computational operations, intentional *in vitro* projects such as synthetic research, or only meant to investigate available substances. This method involves the mapping of property and chemical spaces using modeling functions that are associated with a chemical structure to property or more simply a function that relates property to molecular descriptors. This allows us to competently design and propose new potent compounds that possess the required features and perform the desired function.

Researchers in drug designing areas frequently go for existing commercial software like the Schrodinger software suite. This package has in-built modules which are highly demanding like Phase (Dixon et al., 2006a; Dixon et al., 2006b), ConfGen (Watts et al., 2010), and MacroModel (Watts et al., 2010) for Pharmacophore hypotheses generation (Mitra et al., 2010) and 3D QSAR model development (Lee and Briggs, 2001; Kubinyi et al., 2006; Pissurlenkar et al., 2007; Jiang, 2010; Lauria et al., 2010; Puzyn et al., 2010; John et al., 2011). The G protein-coupled receptors (GPCRs) regulate a countless number of physiological signaling cascades in the body, constituting a rich source of targets for the pharmacological deterrence of various human ailments (Atanes et al., 2018). Most of the GPCRs are expressed in the pancreatic islets are still considered as

“orphan” which are poorly considered functional and or still have no recommended ligands and have not been used as promising antidiabetic targets (Amisten et al., 2017). A very limited quantity of functionally characterized GPCRs in the human body is the target for greater than 30% of all the current diseases. This stresses the reputation of the rest of GPCRs that lacks a functional chart yet in this regard. Analyses of such targets could reveal innovative ways to treat several biological conditions such as metabolic syndrome, cancer, and diabetes. Concerning the current study, the GPR139 and GPR142 are vital targets for numerous conditions and are therefore targeted in this work to discover potent compounds for them using valid pharmacophore and 3D QSAR studies that may have the ability to prevent us from the situation that is likely caused by these GPCRS.

METHODOLOGY

All the steps and techniques employed here are described in detail while the overall workflow of this work is diagrammatically represented in **Figure 1**.

Genomic Modifications Summary

The GPR139 and GP142 tumor samples were used to summarize the genomic alterations. For such analysis, wide-ranging CNA (amplifications and homozygous deletions) and color tagging were considered to review alterations in the gene expression. This was an initial approach to comprehend various forms of gene signaling in the Pan-cancer. The common exclusivity and co-occurrence between GPR139 and GPR142 were examined as well. Events associated with a particular cancer are high time differing in tumor clusters i.e., only a solitary biological incident is anticipated to happen in each cancerous sample. An additional situation is the concurrent existence of changes in several genes in the same sample. This was a preliminary way to collect information linked with various gene signaling in the pan-cancer.

GPR139 and GPR142 Mutations in Pan-Cancer

The locations and frequency of all the mutations within Pfam protein domains were detailed via mutations of GPR139 and GPR142. The whole extent of GPR139 and GPR142 is represented by colored bars while the base of every bar stands for the amino acids' amount. Colored regions are representing the protein's domain and the lines and points signify the location and amount of GPR139 and GPR142. The frameshift or nonsense mutations, missense mutations, and in-frames are visually represented in **Figure 2**.

Overall Survival Inspection

To highlight changes with time for prognosis purposes, the survival analysis bears great importance. In the current work, the Kaplan-Meier plots were used to evaluate differences in the overall survival among samples that were having more or equal to one alteration as that of the query gene(s). This was applied to those samples also that exhibits no alteration.

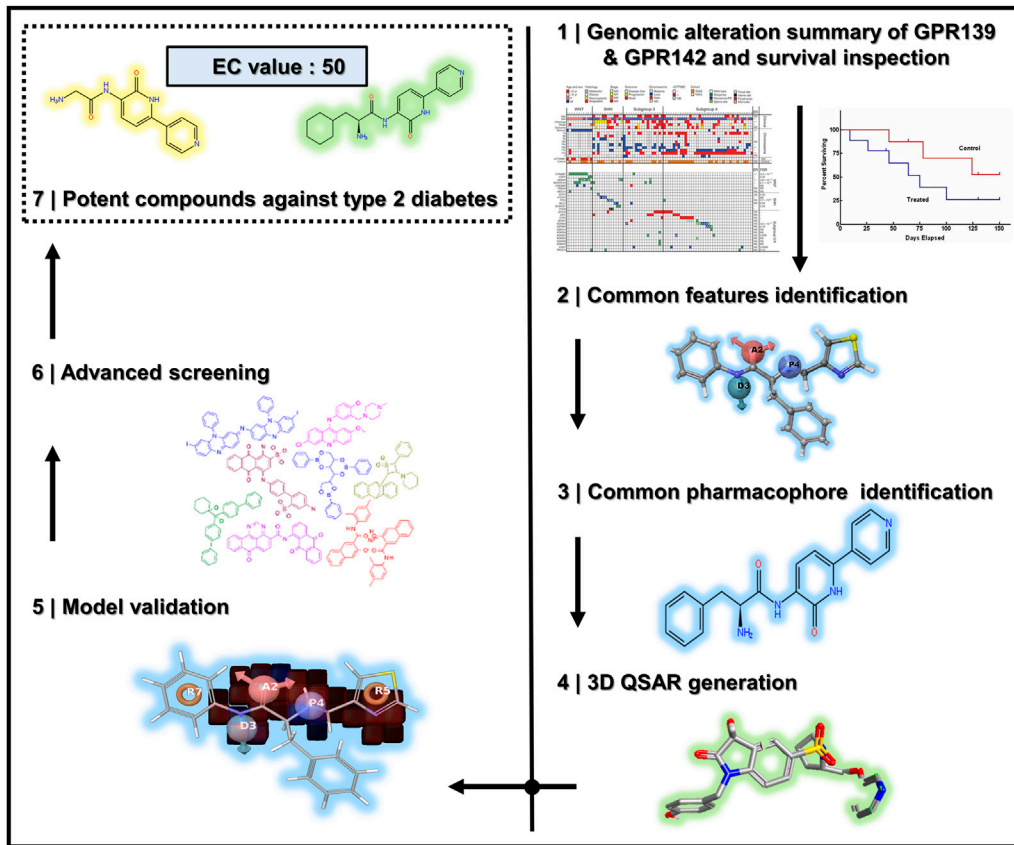


FIGURE 1 | The overall workflow of the steps taken for the successful pharmacophore's perception, 3D QSAR model development, and 3D database screening for GPR142.

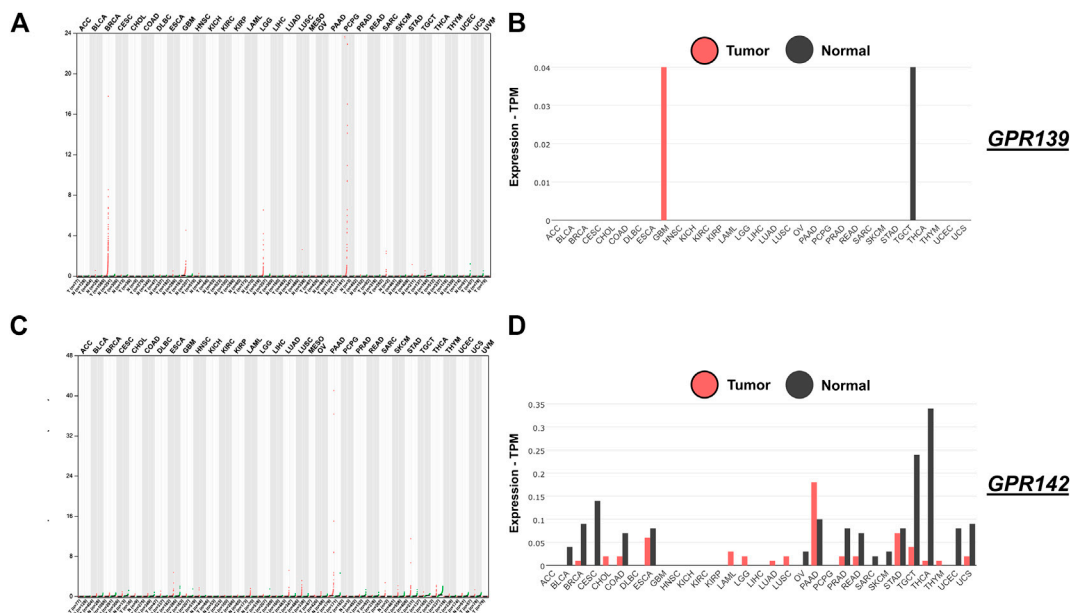
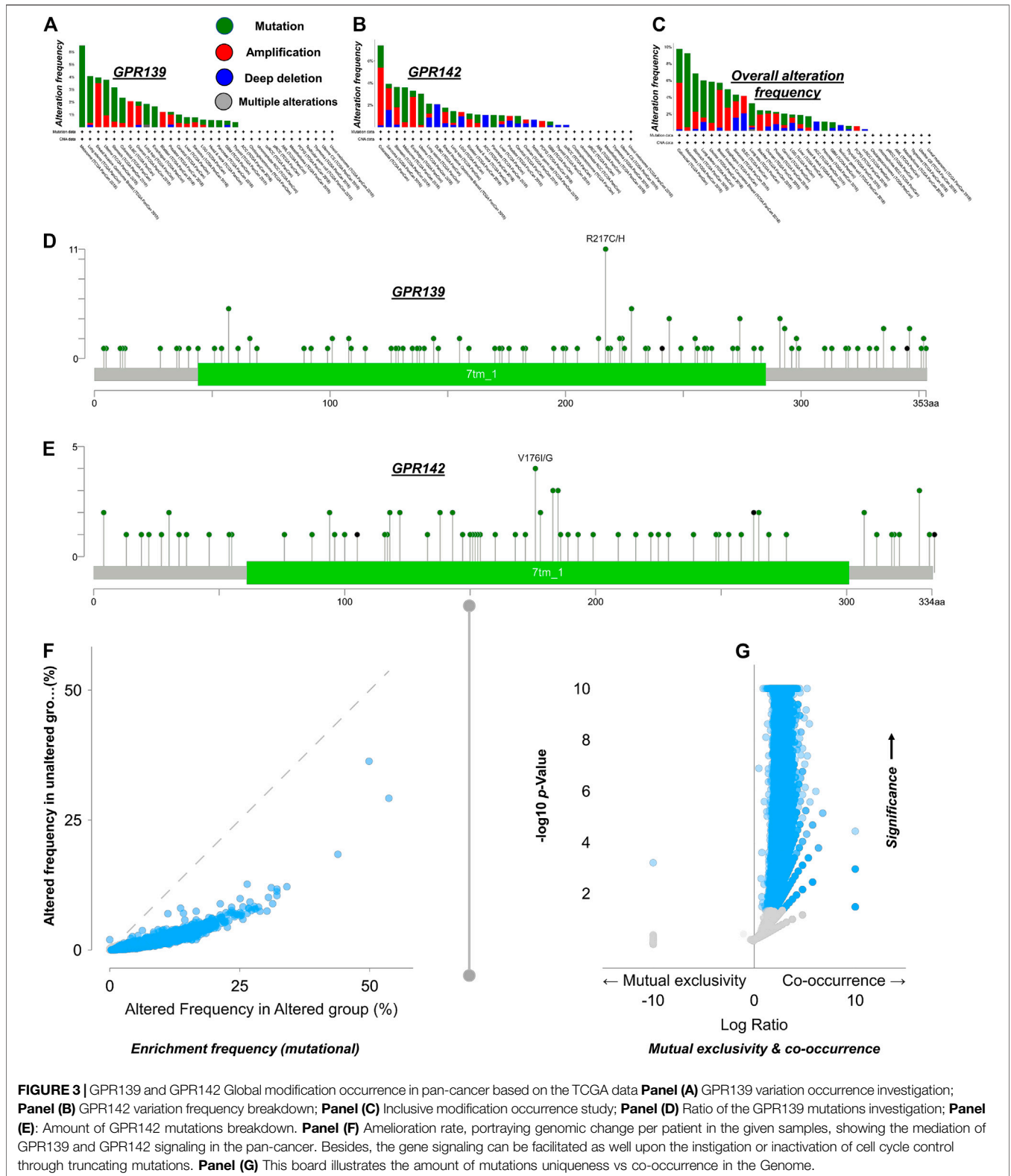


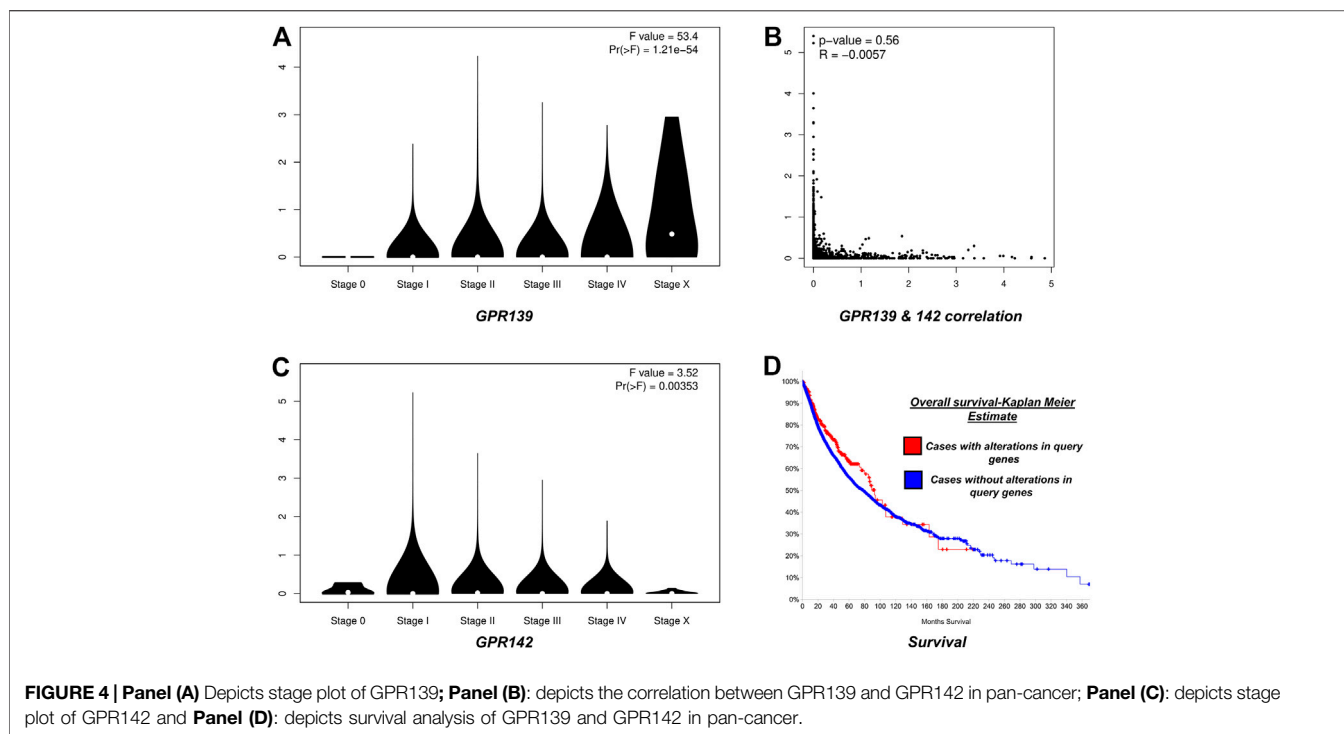
FIGURE 2 | Expression of GPR139 and GPR142 in pan-cancer from analyzed TCGA data. **Panel (A)** GPR139 expression breakdown; **Panel (B)** GPR139 tumor and normal expression breakdown; **Panel (C)** GPR142 expression study; **Panel (D)** GPR142 tumor and normal expression breakdown.



Ligands' Preparation

First of all, ligands were prepared before the construction of pharmacophore along with the development of a 3D Database

using LigPrep (Chen and Foloppe, 2010) which 3D protonates, convert the 2D structure into 3D, form stereoisomers, neutralize charged structures, balances the ionization state and pH using the



OPLS2005 force field (Jorgensen and Tirado-Rives, 1988; Jorgensen et al., 1996; Shivakumar et al., 2010).

Common Pharmacophore Hypotheses Development

A total of 63 compounds (see **Supplementary Table S2**) from the available literature (Du et al., 2012; Lizarzaburu et al., 2012) were chosen, having a particular EC₅₀ value ranging from 0.036 to 33.00. Ligands were prepared; Amide activity property was selected with all primary properties' subset. The performance is given as Activity = $-\log[1*\text{value}]$. The ligands and their various conformers were classified into sets to be used as input data (three datasets are used in this case). The ligands' conformer were generated having activity above 0.02 when Active, and activity below 0.01 when Inactive. Conformers were generated in such a way that the number of conformers per flexible bond was equal to 100. The maximum number of conformers per structure was equal to 1,000.

We employed ConfGen for the MacroModel search method that practices rapid sampling. Different conformers were used for Amide bonds which were pre-processed with a minimization step equal to 100 while the minimization step for high energy and redundant conformers was 50. The OPLS2005 force field was employed for this purpose. A maximum relative energy difference equal to 10.0 kcal/mol was maintained for the distance-dependent dielectric that eliminated redundant conformers using the RMSD of 1.0 Å. Structure cleaning- Stereoisomers retain specified chirality, meaning different chiral centers. Thus, the maximum number of stereoisomers kept was 32,

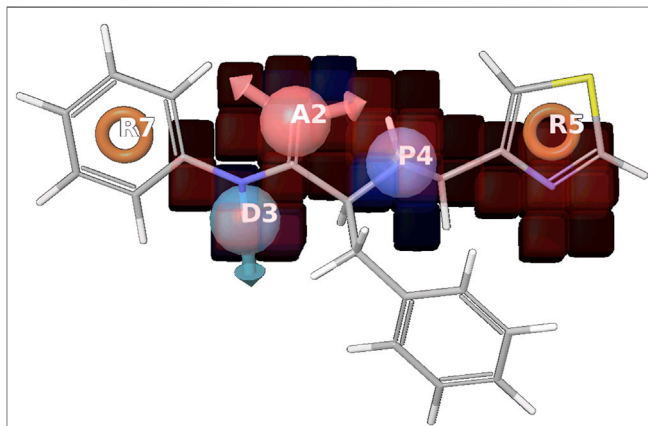


FIGURE 5 | Positive coefficient represented by dark blue color and Negative coefficient represented by the red color, Hydrogen bond donor (D), Hydrophobic/nonpolar (H), Electron-withdrawing (W) shown in the red cube, where R5 has a common pharmacophoric feature, responsible for the activity. A2 are in the negative coefficient and other R7, D3P4, and R5 are in the positive coefficient shown in the supplementary information (**Supplementary Figure S2**).

and Ionization states retained the original structure using phase Schrodinger suite software (Dixon et al., 2006a).

Creating Sites

For a given set of pharmacophoric features, sites of each feature in the given ligand conformations were identified and marked such as Acceptor (A), Donor (D), Hydrophobic (H), Negative (N), and

TABLE 1 | Pharmacophoric variant listing 21 out of 21 were selected, where the highest and lowest amount of positions are 7 and 4 correspondingly.

AAAADDP	AAAADDR	AAAADPR	AADPRRR	AAAADRR	DDPRRRR	AAPRRRR
AAAAPRR	AAAARRR	AAADDPR	AADRRRR	AAADRRR	ADPRRRR	ADDPRRR
AAADPRR	AAADRRR	AAAPRRR	AADPRRR	AAARRRR	ADDRRRR	AADRRRR

The selected variants were used for the generation of a maximum number of pharmacophoric hypotheses derived from a common pharmacophore model for the variants among the given active ligands. All the 7 out of 7 variants accompanying the number of maximum pharmacophoric features were selected.

Aromatic Rings (R). Features in the vector geometry were edited and a point was selected, in projected acceptor was selected, in point sp3, sp2, and sp were selected, 1 atom was used with distance.

First Dataset (High-Affinity EC50 Value)

The first dataset contained a total of sixty compounds. The activity thresholds were Active and Inactive if the compounds are above 0.036 and below 0.001 respectively. The maximum activity in the table was 0.930 while the minimum activity is observed as 0.036. The higher and lower number of sites were 7 and 3 and respectively that must resemble a minimum of 35 compounds out of 38 inactive or active's group. The hypotheses generation scoring, clustering and examined hypotheses, defining excluded volumes, selection of hypotheses for certain QSAR methods as well as search for the similarity or matches with the screened ligands were all carried out.

Second Dataset (Medium Affinity EC50 Value)

The second Dataset also contained the same number of compounds just like the first one, equal to 60. The activity threshold was Active and Inactive if the compounds are above 1.060 and below 1.000. On the other hand, the supreme movement was 6.6000 while a minimum activity of 1.060. The variants' list was defined with a total of 7 maximum sites and a minimum of 4 that must match at least 30 compounds out of 51 actives or in active's group. Score hypotheses generation scoring, clustering and examined hypotheses, defined excluded volumes, selection of hypotheses for certain QSAR methods as well as searching for similar compounds with screened ligands and identified best pharmacophoric featured compounds (Kaushik and Sahi, 2015).

Third Dataset (Low-Affinity EC50 Value)

The third Dataset was also comprised of 60 compounds, having an activity threshold of >0.036 and <0.035 for Active and Inactive respectively. The maximum activity in the table observed was 33.000 while the minimum activity score was found to be 0.036 that was maintained. Defined variants' list- the maximum number of sites was 7 and the minimum number of sites was 4 that had to resemble a minimum of twenty compounds with that of the total 60 actives or inactive set. Score hypotheses generation scoring, clustering and examined hypotheses, defining excluded volumes, selection of hypotheses for certain of QSAR methods as well as searching for matching with screened ligands were performed.

TABLE 2 | Selected and desired variants finding common pharmacophore for the variants among the given active ligands, 7 out of 7 selected variants can be seen in the variants' column.

Variant	Maximum number of hypotheses
ADRRR	1
AADPR	5
AAPRR	8
APRRR	7
DPRRR	6
ADPRR	22
AADRR	3

RESULTS

GPR139 and GPR142 Expression in Pan-Cancer

As a result of the expression analysis of GPR139, it was exposed that a substantial amount of up and down-regulation that explains the hotspots which are responsible for the activation and role in BRCA, GBM, LGG, PCPG, and SARC as shown in **Figure 2** while expression analysis of GPR142 revealed a substantial amount of up and down-regulation that explains the hotspots which are responsible for the activation and role in ESCA, LIHC, LUAD, LUSC, PAAD, STAD, TGCT, THCA, and UCS as shown in **Figure 2**.

Genomic Site and Variations Summary

Based on the obtained conclusions, most of the cases were found to be changing the GPR139 and GPR142. Upon further analysis, it was revealed that nearly all of the observed variations were missense mutations. Some deep deletions and few amplifications have also been noticed. Nevertheless, the remaining of the cases experienced alterations in GPR139 and GPR142 that are mainly exhibiting truncating and missense mutations. Examining the mutual exclusiveness suggests that events happened in GPR139 and GPR142 were responsible to occur again in pan-cancer (GPR139 exposed in BRCA, GBM, LGG, PCPG, and SARC) while GPR142 exposed in ESCA, LIHC, LUAD, LUSC, PAAD, STAD, TGCT, THCA, and UCS) as represented in **Figure 3**.

Analyzing Survival Rate

To inspect the rate of survival, the Kaplan-Meier approach was used to plot the complete survival analysis for the pan-cancer. Based on the global survival analysis, it was observed that mutations in the cell cycle control were concurrent and were not associated with the overall decreased survival (p -value =

TABLE 3 | Score hypotheses of ADPRR.18 where site, volume, and matches are listed with alignment for a hypothesis of ADPRR.18 where compound number 25 is observed to be a more valid and suitable one.

	Site	Vector	Volume	Selectivity	Match	Survival	Energy	Activity
ADPRR.18	0.71	0.913	0.611	2.273	31	3.237	0.000	3.300
Activity		Pharm set		Fitness		Site matched		Relative energy
3.300		Active		3.00		5		0.000

0.0615) as illustrated in **Figure 4**; while correlation analysis of clinical features observed GPR139 and GPR142 correlated in the progression of pathological stages.

Common Pharmacophoric Features Identification

Maintaining common features in a pharmacophore model is very important as they are the backbone of the parent molecules on which the functionality of the drug highly depends. All the common pharmacophoric features obtained from the 3D Database of GPR142 using phase Schrodinger suite software are summarized in the **Supplementary Table S1**. Out of 1,038 chemical structures from GPR142 3D Database; only 5 are chosen that are flexible chemical structures with an average reduction of up to 1.18 Å plus they are having a better common pharmacophoric match for pan-cancer target pharmacophore shown in **Figure 5**.

Common Pharmacophore Identification

To identify a common pharmacophore model, the selected and desired variants were used for the given active ligands (**Tables 1,2**).

Score Hypotheses

Score hypotheses generation was created by 30 out of 51. Apart from this, clustering and examination of the hypotheses, defining excluded volumes, and the selection of hypotheses for certain QSAR methods as well as the search for features with screened ligands were carried out. Score Actives; vector and site filtering maintained only those variants' that are having RMSD below 1.200 Å, vector score above 0.500 Å, and be the top 10%. The quantity should be in the range of 10 (minimum) and 50 (maximum). The survival score formula for vector score is 1.000, +1.000 for site score, +1.000 as volume score, -0.000 for the reference ligand relative conformational energy, +0.000 as selective score, +1.000 for the number of matches, and +0.000 in case of the reference ligand activity. All the values are listed in **Table 3**.

Building QSAR Model

In QSAR (Quantitative structure-activity relationship) model building, the predicted structure-activity relations for the matching ligands are investigated. The training set where the random seed was 0, keeping actives and inactive in the training set. The 3D QSAR model parameters were kept as standard. For example, the Grid sampling was 1.00 Å, the maximum PLS factor was 1, and the model type was atom-based (**Table 4**).

TABLE 4 | Training set where the random seed was 0, keeping actives and inactive in the training set where grid sampling is equal to 1.00 Å, maximum PLS factor was 1 and the model type is atom-based.

QSAR set (Training)	QSAR set (Test)
66, 64, 63, 62, 61, 60, 59, 58, 57, 56, 55, 53	54, 14, 13, 7, 6, 5, 4, 3, 2
51, 48, 45, 43, 42, 40, 39, 35, 34, 33, 32, 31	
30, 29, 27, 26, 24, 23, 22, 21, 20, 19, 18, 17	
16, 15, 52, 50, 49, 46, 41, 36, 28, 25, 11, 10, 9, 8, 1	

Model Validation

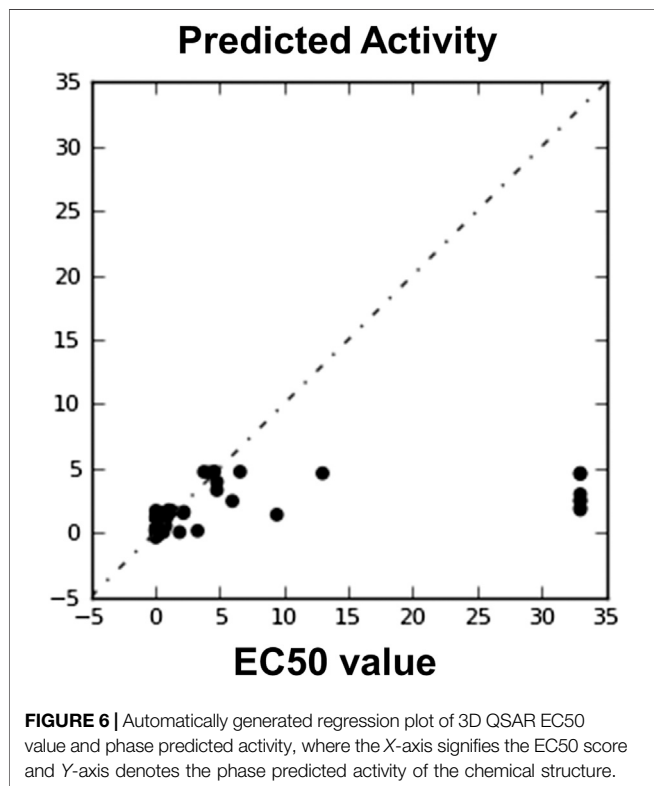
The model validation phase of the 3D QSAR models uses distinct training and testing sets for the cross-validation approach. The 3D QSAR models estimate outcome which is derived from the training set shown in **Figure 5**. The highly stable models are preferred as they don't rely completely on the training set. Regarding the statistical parameters for the training set, the m represents the number of PLS factors in the model, n represents the number of molecules in the training set, $df_1 = m + 1$ represents the degrees of freedom in the model, $df_2 = n - m - 2$ denotes the degree of freedom in the data, y represents the detected movement for a training set molecule i , \hat{y}_i signifies the predicted activity for training set molecule i and $R^2 = 1 - \sigma^2 / \sigma^2$ is the R-squared or coefficient of determination. Test set prediction was performed by phase defined parameters, where T signifies the test set of molecules, n_T denotes the number of molecules in T , Y_j is the observed activity for molecule $j \in T$, \hat{y}_j means the predicted activity for molecule $j \in T$, and $Q^2 = R^2(T)$ represents the Q-squared. The necessary parameters for the ADPRR are given in **Table 5**; **Figures 5,6**.

Screened Ligands and Known Compounds' Pharmacophore Analysis

The common pharmacophore using screened and known EC50 compounds was found along with a pharmacophore for those variants which are among the given active ligands. The variants' list was defined while the higher and the lower quantity regarding the sites were kept as 7 and 4 correspondingly, matching at least 20 compounds out of 137 active or inactive groups as shown in **Figure 7**. The variant list 269, after the identification of common pharmacophore, the ADHPRRR.35 was considered having maximum hypotheses as 16 with a survival rate of 2.960. The site score was 0.58, a vector was equal to 0.939, volume was 0.442, selectivity was 3.858 and the number of matches was 22. After the alignment of ADHPRRR.35 hypotheses, its fitness value turned out to be 2.02 and the number of site matches was 7 listed in **Table 6**; **Figures 7,8**.

TABLE 5 | 3D QSAR model parameters for ADPRR.18 pharmacophoric hypotheses, where R-squared value is 0.6478 which is considered a more favorable pharmacophoric feature in this case.

ID	SD	R-squared	F	P	Stability	RMSE	Q-squared	Pearson-R
ADPRR.18	1.0282	0.6478	90.1	1.097e-12	0.9449	26.7754	-7.674	0.0504

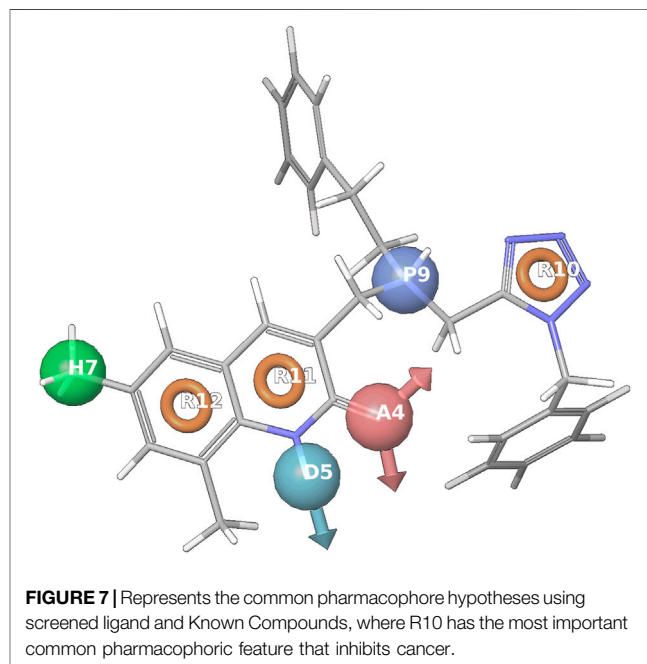


New Pharmacophore Hypotheses Generation Using Glide Docking Score

The compounds were docked and scored by the newly identified common pharmacophore to find common features for the variants present among the given active ligands. Defining the variants' list and the highest and lowest number of spots as 7 and 4 correspondingly that must resemble a minimum of 95 compounds out of 137 active/inactive groups was done and maintained. After the identification of a common pharmacophore model, the variants' list was 24 where R8 is common in all pharmacophoric analysis and is observed to be responsible for the native activity of compounds, given in the supplementary information (**Supplementary Figure S1**).

Building QSAR Model

The QSAR model was built and examined for the predicted structure-activity relations concerning the matching ligands. The random seed in the training set was 0, keeping both the actives and inactive in the training set, sampled uniformly over the activity coordinates. The random training set was only 50%, 3D QSAR model parameters Grid sampling was 1.00 Å, maximum PLS factor was 1, and the model type was atom-based. After the



alignment, Compound6 alignment score was observed to be 0.089967, vector score was equal to 0.996997, volume score was given as 0.428221, fitness value of 2.350246 and phase predicted score was -4.86423, Compound7 alignment score was 0.095122, vector score as 0.936649, volume score was observed to be 0.37549, fitness value was given as 2.232871 and phase predicted score was equal to -5.10154. Similarly, the Compound8 alignment score was 0.298618, vector score was 0.989711, volume score turned out to be 0.427929, fitness value as 2.168791 and phase predicted score equal to -5.16605 as shown in **Figure 9**.

Advanced Pharmacophore Screening of GPR142

Various hypotheses assisted in the identification of new matches and the conformers were generated automatically. For the conformer's generation and refinement, existing conformers were discarded, the number of structures per adjustable bond was kept as ten, the highest number of conformers per assembly was 100 with rapid sampling. The relative energy window was equal to 10.0 kcal/mol. Conformer's generation was skipped in case of rotatable bonds greater than fifteen. The matching tolerance was 2.0 Å in the inter-site and must match on at least 4 site positions out of 4. For the hit treatment purposes, hits were sorted out by decreasing fitness, that returned 1,000 hits

TABLE 6 | 3D QSAR model parameters for ADHPRRR.35 pharmacophoric hypotheses where grid sampling was 1.00 Å, maximum PLS factor was 1, and the model type was an atom-based pharmacophoric analysis.

ID	SD	R-squared	F	P	Stability	RMSE	Q-squared	Pearson-R
ADHPRRR.35	8.5856	0.3407	30	9.84e-07	0.931	7.6838	0	0

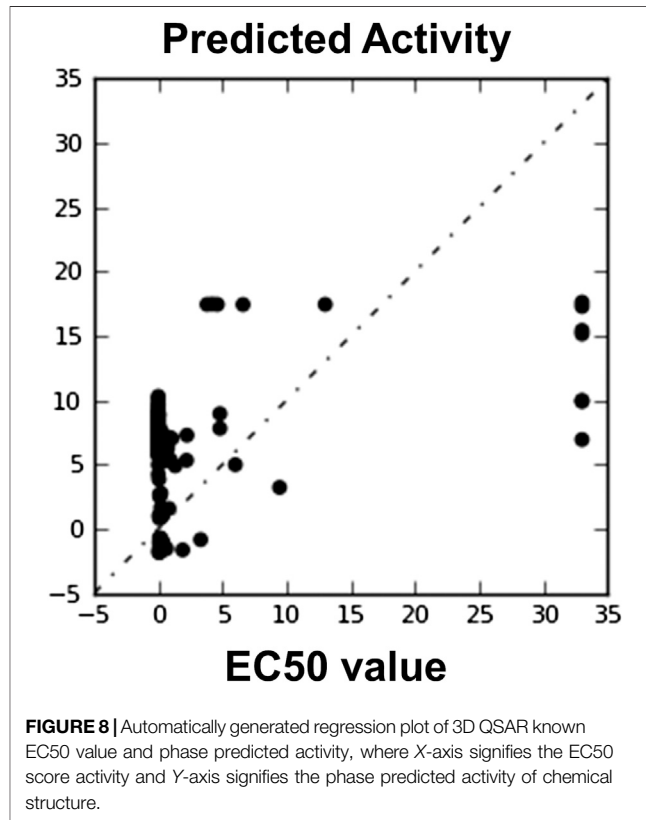
in total, i.e. 1 hit per molecule. Fitness of hit treatment was equal to 1.0 (for alignment score/1.2) + 1.0 (for vector score) and + 1.0 volume score. All hits with < -1.0 vector score and < 0.0 volume score were rejected. The **Compound1**, **Compound2**, and **Compound3** were obtained after screening from the 3D database of GPR142, having the same pharmacophoric features that inhibit cancer.

Advanced Pharmacophore Screening of Known EC50 Compounds and Screened Compounds

Advanced pharmacophoric screening of known EC50 compounds and Screened Compounds was done using the default parameters. The existing conformers were discarded, the number of structures per adjustable bond was kept ten. The highest number of conformers apiece structure was kept as 100 with rapid sampling. The relative energy window was 10.0 kcal/mol, skipping conformer generation for structures with >15 rotatable bonds. Matching in inter-site distance matching tolerance was maintained at 2.0 Å and must match on at least 4 site positions out of 4. For the case of hit treatment purpose, hits by decreasing fitness turned out 1,000 in total which is equivalent to 1 hit per molecule. Fitness of the hit treatment was 1.0 (for alignment score/1.2) + 1.0 (for vector score) + 1.0 volume score. All hits with < -1.0 vector score and < 0.0 volume score were totally rejected. **Compound4** and **compound5** were obtained after screening the 3D Database of GPR142, which has the same pharmacophoric features and common chemical structures that can inhibit cancer.

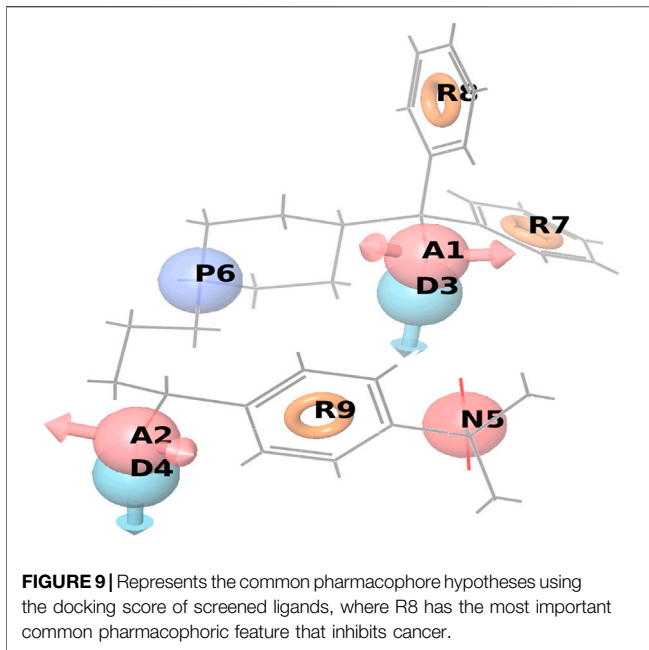
Advanced Pharmacophore Screening and Analysis of New Pharmacophore Model

Advanced pharmacophoric screening of the newly identified pharmacophore model was performed by the glide docking scoring system. The number of conformers was set per rotatable bond as 10, a maximum number of conformers per structure was 100 with rapid sampling. The relative energy window was 10.0 kcal/mol. All the conformers' generation in case of structures with >15 rotatable bonds were skipped. The matching in inter-site distance matching tolerance was kept as 2.0 Å and must match on four sites position out of 4. Hits were sorted by decreasing fitness returning at its most equal to 1000 hits in total that is 1 hit per molecule. The fitness of hit treatment was 1.0 (for alignment score/1.2) + 1.0 (for vector score) + 1.0 volume score. Reject hits with < -1.0 vector score and < 0.0 volume score. In all the considered compounds, **Compound6**, **compound7**, and **Compound8** were chosen after screening the 3D Database of GPR142. They all were having the same pharmacophoric features and can inhibit cancer.



Advanced Pharmacophore Screening of GPR139

The Paralog of the GPR142 gene is GPR139, and the similarity gene and protein between GPR139 and GPR142 is 50%, sharing suggested ligands. The advanced pharmacophoric screening was carried out by these protocols such as; finding matches from the generated features' search during conformers exploration that were automatically generated. All existing conformers were discarded, the conformers apiece adjustable bond was chosen to be 10. On the other hand, the maximum amount for conformers apiece structure was chosen to be 100, the relative energy window was equal to 10.0 kcal/mol. Similarly, conformers' generation for structures with greater than 15 rotatable bonds was skipped. Matching in inter-site distance matching tolerance was 2.0 Å and must match on at four site positions out of the total 4. All hits were sorted decreasing the fitness value, returning at the most given hits which are 1,000 in total that is 1 hit per molecule. Fitness of hit treatment was 1.0 (for alignment score/1.2) + 1.0 (for vector score) + 1.0 as the volume score. Hits were rejected with < -1.0 vector score and < 0.0 volume score. The **Compound9** and **Compound10** were selected as a result



of screening from the 3D Database of GPR139, which contains similar pharmacophoric features like GPR142 target compounds. Compound9 and Compound10 might be helpful in the proteins' dimerization activity and neuropeptide receptor activity as provided in the supplementary data (**Supplementary Table S3**).

CONCLUSION

Computational drug design mainly emphasizes structure-based techniques. The *in silico* screening of small compounds in large databases using developed pharmacophoric features is a good way to discover and identify the best chemical compounds by using common pharmacophoric hypotheses generation and 3D QSAR models. However, extra care should be taken during the development of such a model as the performance of the predicted compounds depends on the features considered. It is a more feasible, robust, and flexible way to use such models. The current study reports on genomic alterations of GPR139 and GPR142 through tumor samples, assisting in the identification of common drug compounds for GPR142 target by 3D database by scanning common pharmacophoric features and experimental EC₅₀ value which might be useful in the inhibition of cancer. We intend to subject these compounds to explore their interactions with the target and stability in the host region via long term molecular dynamics simulation and *in vitro* analysis.

REFERENCES

Amisten, S., Atanes, P., Hawkes, R., Ruz-Maldonado, I., Liu, B., Parandeh, F., et al. (2017). A comparative analysis of human and mouse islet G-protein coupled receptor expression. *Sci. Rep.* 7, 46600. doi:10.1038/srep46600

DATA AVAILABILITY STATEMENT

All the datasets analyzed during this study are available from the corresponding authors upon a reasonable request.

AUTHOR CONTRIBUTIONS

AC and D-QW designed the experiment, D-QW and AC performed the entire computational experiments and assisted in writing the manuscript, D-QW, AC, and AM analyzed the data and wrote the manuscript. AC, D-QW, XD, and AM read the manuscript and advised on method development. All authors have approved the final version of the manuscript.

FUNDING

This work is supported by the grants from the Key Research Area Grant 2016YFA0501703, 2018ZX10302205-004-002, and 2018ZX10302-205-004 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract No. 61832019, 61503244, and BK20161130), the State Key Lab of Microbial Metabolism and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The Six Talent Peaks Project in Jiangsu Province (Grant No. SWYY-128), Major Project of Science and Technology in Henan Province (Grant No.161100311400), The Technology Development Funding of Wuxi (Grant No. WX18IVJN017), Research Funds for the Medical School of Jiangnan University ESI special cultivation project (Grant No. 1286010241170320). These funding sources have no role in the writing of the manuscript or the decision to submit it for publication.

ACKNOWLEDGMENTS

The heavy-duty computational operations carried out in this work was supported by the Center for High-Performance Computing, Shanghai Jiao Tong University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2020.521245/full#supplementary-material>.

Andrusier, N., Nussinov, R., and Wolfson, H. J. (2007). FireDock: fast interaction refinement in molecular docking. *Proteins* 69 (1), 139–159. doi:10.1002/prot.21495

Atanes, P., Ruz-Maldonado, I., Hawkes, R., Liu, B., Zhao, M., Huang, G. C., et al. (2018). Defining G protein-coupled receptor peptide ligand expressomes and signalomes in human and mouse islets. *Cell. Mol. Life Sci.* 75 (16), 3039–3050. doi:10.1007/s00018-018-2778-z

- Bhachoo, J., and Beuming, T. (2017). "Investigating protein-peptide interactions using the Schrödinger computational suite, *Methods molecular biology, Modeling peptide-protein interactions*, Editors O. Schueler-Furman and N. London New York, NY: Humana Press, Vol. 1561, 235–254. doi:10.1007/978-1-4939-6798-8_14
- Chen, I. J., and Follpe, N. (2010). Drug-like bioactive structures and conformational coverage with the LigPrep/ConfGen suite: comparison to programs MOE and catalyst. *J. Chem. Inf. Model.* 50 (5), 822–839. doi:10.1021/ci100026x
- Dixon, S. L., Smondyrev, A. M., and Rao, S. N. (2006). PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug Des.* 67 (5), 370–372. doi:10.1111/j.1747-0285.2006.00384.x
- Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., and Friesner, R. A. (2006a). PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.* 20 (10–11), 647–671. doi:10.1007/s10822-006-9087-6
- Du, X., Kim, Y. J., Lai, S., Chen, X., Lizarzaburu, M., Turcotte, S., et al. (2012). Phenylalanine derivatives as GPR142 agonists for the treatment of type II diabetes. *Bioorg. Med. Chem. Lett.* 22 (19), 6218–6223. doi:10.1016/j.bmcl.2012.08.015
- Huang, N., Shoichet, B. K., and Irwin, J. J. (2006). Benchmarking sets for molecular docking. *J. Med. Chem.* 49 (23), 6789–6801. doi:10.1021/jm0608356
- Irwin, J. J., and Shoichet, B. K. (2005). ZINC— a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45 (1), 177–182. doi:10.1021/ci049714
- Jiang, Y.-K. (2010). Molecular docking and 3D-QSAR studies on beta-phenylalanine derivatives as dipeptidyl peptidase IV inhibitors. *J. Mol. Model.* 16 (7), 1239–1249. doi:10.1007/s00894-009-0637-4
- John, S., Thangapandian, S., Arooj, M., Hong, J. C., Kim, K. D., and Lee, K. W. (2011). Development, evaluation and application of 3D QSAR Pharmacophore model in the discovery of potential human renin inhibitors. *BMC Bioinf.* 12, S4. doi:10.1186/1471-2105-12-S14-S4
- Jorgensen, W. L., and Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110 (6), 1657–1666. doi:10.1021/ja00214a001
- Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118 (45), 11225–11236. doi:10.1021/ja9621760
- Joy, S., Nair, P. S., Hariharan, R., and Pillai, M. R. (2006). Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware. *Silico Biol.* 6 (6), 601–605.
- Junaid, M., Shah, M., Khan, A., Li, C.-D., Khan, M. T., Kaushik, A. C., et al. (2019). Structural-dynamic insights into the H. pylori cytotoxin-associated gene A (CagA) and its abrogation to interact with the tumor suppressor protein ASP2 using decoy peptides. *J. Biomol. Struct. Dyn.* 37 (15), 4035–4050. doi:10.1080/07391102.2018.1537895
- Kaushik, A. C., and Sahi, S. (2015). Boolean network model for GPR142 against Type 2 diabetes and relative dynamic change ratio analysis using systems and biological circuits approach. *Syst. Synth. Biol.* 9 (1–2), 45–54. doi:10.1007/s11693-015-9163-0
- Khan, A., Junaid, M., Kaushik, A. C., Ali, A., Ali, S. S., Mehmood, A., et al. (2018). Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches. *PLoS One* 13 (5), e0196484. doi:10.1371/journal.pone.0196484
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44 (D1), D1202–D1213. doi:10.1093/nar/gkv951
- Korber, B., LaBute, M., and Yusim, K. (2006). Immunoinformatics comes of age. *PLoS Comput. Biol.* 2 (6), e71. doi:10.1371/journal.pcbi.0020071
- Kubinyi, H., Folkers, G., and Martin, Y. C. (2006). *3D QSAR in drug design: recent advances*. Ludwigshafen, Germany: Springer Science & Business Media.
- Lauria, A., Ippolito, M., Fazzari, M., Tutone, M., Di Blasi, F., Mingoia, F., et al. (2010). IKK-beta inhibitors: an analysis of drug-receptor interaction by using molecular docking and pharmacophore 3D-QSAR approaches. *J. Mol. Graph. Model.* 29 (1), 72–81. doi:10.1016/j.jmgl.2010.04.008
- Lee, K. W., and Briggs, J. M. (2001). Comparative molecular field analysis (coMFA) study of epothilones-tubulin depolymerization inhibitors: pharmacophore development using 3D QSAR methods. *J. Comput. Aided Mol. Des.* 15 (1), 41–55. doi:10.1023/a:1011140723828
- Lizarzaburu, M., Turcotte, S., Du, X., Duquette, J., Fu, A., Houze, J., et al. (2012). Discovery and optimization of a novel series of GPR142 agonists for the treatment of type 2 diabetes mellitus. *Bioorg. Med. Chem. Lett.* 22 (18), 5942–5947. doi:10.1016/j.bmcl.2012.07.063
- Mehmood, A., Kaushik, A. C., and Wei, D. Q. (2019). Prediction and validation of potent peptides against Herpes Simplex Virus Type 1 (HSV-1) via immunoinformatic and systems biology approach. *Chem. Biol. Drug Des.* 94, 1868–1883. doi:10.1111/cbdd.13602
- Mitra, I., Saha, A., and Roy, K. (2010). Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers. *J. Mol. Model.* 16 (10), 1585–1596. doi:10.1007/s00894-010-0661-4
- Pissurlenkar, R. R., Shaikh, M. S., and Coutinho, E. C. (2007). 3D-QSAR studies of Dipeptidyl peptidase IV inhibitors using a docking based alignment. *J. Mol. Model.* 13 (10), 1047–1071. doi:10.1007/s00894-007-0227-2
- Puzyn, T., Gajewicz, A., Leszczynska, D., and Leszczynski, J. (2010). Nanomaterials—the next great challenge for QSAR modelers,” in *Recent advances in QSAR studies* (Gdansk, Poland: Springer), 383–409.
- Roy, U., and Luck, L. A. (2007). Molecular modeling of estrogen receptor using molecular operating environment. *Biochem. Mol. Biol. Educ.* 35 (4), 238–243. doi:10.1002/bmb.65
- Schneider, G., Neidhart, W., Giller, T., and Schmid, G. (1999). " Scaffold-hopping " by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed.* 38 (19), 2894–2896. doi:10.1002/(sici)1521-3773(19991004)38:19<2894::aid-anie2894>3.0.co;2-f
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33 (Suppl. 2), W363–W367. doi:10.1093/nar/gki481
- Shemetulskis, N. E., Dunbar, J. B., Dunbar, B. W., Moreland, D. W., and Humblet, C. (1995). Enhancing the diversity of a corporate database using chemical database clustering and analysis. *J. Comput. Aided Mol. Des.* 9 (5), 407–416. doi:10.1007/BF00123998
- Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010). Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theor. Comput.* 6 (5), 1509–1519. doi:10.1021/ct900587b
- Sun, H. (2008). Pharmacophore-based virtual screening. *Curr. Med. Chem.* 15 (10), 1018–1024. doi:10.2174/092986708784049630
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31 (2), 455–461. doi:10.1002/jcc.21334
- Wadood, A., Mehmood, A., Khan, H., Ilyas, M., Ahmad, A., Alarjah, M., et al. (2017). Epitopes based drug design for dengue virus envelope protein: a computational approach. *Comput. Biol. Chem.* 71, 152–160. doi:10.1016/j.cmpbiolchem.2017.10.008
- Wang, Q., Mehmood, A., Wang, H., Xu, Q., Xiong, Y., and Wei, D.-Q. (2019). Computational screening and analysis of lung cancer related non-synonymous single nucleotide polymorphisms on the human Kirsten rat sarcoma gene. *Molecules* 24 (10), 1951. doi:10.3390/molecules24101951
- Watts, K. S., Dalal, P., Murphy, R. B., Sherman, W., Friesner, R. A., and Shelley, J. C. (2010). ConfGen: a conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* 50 (4), 534–546. doi:10.1021/ci100015j
- Winstead, B., and Ravaioli, U. (2003). A quantum correction based on Schrodinger equation applied to Monte Carlo device simulation. *IEEE Trans. Electron. Dev.* 50 (2), 440–446. doi:10.1109/TED.2003.809431

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kaushik, Mehmood, Dai and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.