


12-2018

Targeting PH domain proteins for cancer therapy

Zhi Tan

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the [Bioinformatics Commons](#), [Medicinal Chemistry and Pharmaceutics Commons](#), [Neoplasms Commons](#), and the [Pharmacology Commons](#)

Recommended Citation

Tan, Zhi, "Targeting PH domain proteins for cancer therapy" (2018). *UT GSBS Dissertations and Theses (Open Access)*. 910.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/910

This Dissertation (PhD) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact laurel.sanders@library.tmc.edu.

TARGETING PH DOMAIN PROTEINS
FOR
CANCER THERAPY

by

Zhi Tan

APPROVED:

Advisory Professor, Shuxing Zhang, Ph.D.

Paul Chiao, Ph.D.

Jian Kuang, Ph.D.

Gabriel Lopez-Berestein, M.D.

Sanjay Shete, Ph.D.

APPROVED:

Dean, The University of Texas

Graduate School of Biomedical Sciences at Houston

TARGETING PH DOMAIN PROTEINS FOR CANCER THERAPY

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

M.D. Anderson Cancer Center

Graduate School of Biomedical Sciences

In Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Zhi Tan, M.D.

Houston, Texas

December 2018

Dedication

To my bright and beautiful wife, Rui Liang and adorable daughter, Serena Tan, who have loved, inspired, and trusted me all the time. To my parents, Rongfang Tan and Xiaoqiu Li, who are always supportive in my scientific career. To my brother Zheng Tan, who supports every decision I have ever made. To Dr. Shuxing Zhang, the greatest mentor I ever met.

Acknowledgements

I acknowledge my dear colleagues in Dr. Shuxing Zhang's lab, including Beibei Huang, Lu Chen, Luo Song, Diana Machado, John Morrow, and Rajan Chaudhari. I also thank Dr. Jangsoon Lee and great researchers in Dr. Yubin Zhou lab for their dedicate contributions to experimental validation results.

I owe great acknowledgement to my advisory committee, Drs. Shuxing Zhang, Paul Chiao, Sanjay Shete, Peng Huang, Jian Kuang, and Gabriel Lopez- -Berestein for their valuable suggestions on this thesis. I could not make it happen without their support.

Special thanks to the University of Texas M.D. Anderson Cancer Center and University of Austin for providing state-of-the-art HPC resources. Appreciate TACC for free academic license of Gromacs.

TARGETING PH DOMAIN PROTEINS FOR CANCER THERAPY

By Zhi Tan, M.D.

Advisor: Shuxing Zhang, Ph.D.

Targeted therapy has been one of the most promising treatment options for cancer during the past decade. Discoveries of potent and selective small molecule inhibitors are critical to new and promising targeted therapy. Pleckstrin Homology (PH) domain proteins are one of the biggest protein families in human proteome. However, no drugs have been achieved to the late development stages, let alone getting to the market. Thus, a deeper understanding about this protein family is required and there is an urgent need to develop novel small molecule compounds targeting these proteins.

Studies of PH domains began around two decades ago and a lot of efforts have been focused on their structures and functions. However, not much is known about their role in cancers, except a few proteins such as AKT. In order to delineate the roles of PH domain proteins in cancers, we performed a comprehensive analysis of 313 PH domain proteins using 13 types of most common cancers in TCGA. From this analysis, we identified the most frequently upregulated and mutated PH domain proteins. Interestingly, we found Tiam1, a guanine nucleotide exchange factor (GEF) specific for Rac1 activation, was overexpressed in several cancers, particularly neuroendocrine prostate cancer.

Targeting PH domain proteins remains to be a significant challenge for multiple reasons. First, the binding pockets of most PH domain proteins are unknown due to lacking of PH-PIPs complex crystal structures. Second, these binding pockets are positively charged, which makes it really difficult to design small molecule inhibitors targeting these sites. In order to address these issues, we performed structural sequence alignment of available PH domain structures to identify conserved residues. Also, ensemble docking was performed in order to address the flexibility of the proteins. Through these efforts, we identified two scaffolds as Tiam1 small molecule inhibitors. These inhibitors showed binding affinity to the PH domain using surface plasmon resonance (SPR) assay and inhibition of Rac1 activation in prostate cancer cells. Also, these compounds inhibited prostate cancer cell proliferation and migration *in vitro*.

Table of Contents

Approval page.....	i
Title page	ii
Dedication.....	iii
Acknowledgements.....	iv
Abstract.....	v
Table of Contents.....	vii
List of Tables	xiii
Abbreviations.....	xiv
Chapter 1: Introduction.....	1
1.1 PH domain as a drug target.....	1
1.2 PH domain proteins in cancer	2
1.3 Current situation of developing small molecule inhibitors targetin PH domains	3
Chapter 2: Genomics, structural and PIPs binding specificity analysis of PH domain proteins.....	5
2.1 Introduction.....	5
2.2 Methods and materials	6
2.2.1 KEGG pathway analysis of PH domain genes.....	6
2.2.2 Somatic mutation analysis	6
2.2.3 Unsupervised clustering.....	7
2.2.4 Curation of PH domain proteins from PDB.....	7
2.2.5 Structural sequence alignment and weblogo generation.....	7
2.2.6 Datasets curation and database generation.....	8
2.2.7 Classification of PIP binding using convolutional neural network.....	8
2.3 Results.....	9
2.3.1 PH domain genes were overrepresented in multiple pathways.....	9

2.3.2 A 20-gene signature resulted in five clusters of clinical samples across 13 cancer types	12
2.3.3 Tiam1 is one of the most frequently mutated genes across 13 cancer types in PH domain genes.	15
2.3.4 Structural analysis reveals conserved residues in PH domains.....	22
2.3.5 Clustering of the available PH domain proteins.	26
2.3.6 A heatmap of PIPs binding specificity based on published data.....	28
2.3.7 Collecting PH-PIPs binding data from PubMed using text mining and building a classification model based on Convolutional neural network (CNN).....	30
2.4 Discussion.....	66
Chapter 3: In silico discovery of small molecule inhibitors targeting Tiam1	68
3.1 Introduction.....	68
3.2 Methods and materials	71
3.2.1 Sequence alignment and sequence logo generation	71
3.2.2 Ensemble docking.....	71
3.2.3 Chemical dataset and Virtual screening.....	71
3.2.4 Pharmacophore Modeling	72
3.2.5 Surface plasmon resonance (SPR) assays.	72
3.2.6 Culture of prostate cancer and normal prostate cell lines.	73
3.2.7 Rac1 activation assay.....	73
3.2.8 Wound healing assay.	73
3.2.9 Lamellipodia formation assay.....	73
3.2.10 Colony formation assay.	74

3.2.11 Matrigel invasion assay.....	74
3.3 Results.....	76
3.3.1 Structural analysis of cPH domain of Tiam1 protein.....	76
3.3.3 <i>In silico</i> discovery of inhibitors of Tiam1 binding to cPH domain.....	87
3.3.4 TPH3 inhibits prostate cancer cell proliferation and migration.	94
3.3.6. Complex structure refinement and Prediction of molecular interaction between TPH3 and cPH Tiam1.	99
3.3.7. Bound ligands induce Tiam1 cPH domain conformational changes and stabilize protein complexes.	101
3.4 Discussion.....	104
Chapter 4 Summary and future directions	107
4.1 Summary of PH domain protein features.....	107
4.2 Summary of small molecule inhibitors targeting Tiam1.....	107
4.3 Future directions in small molecule inhibitors targeting Tiam1	108
4.3 Future directions in drug discoveries targeting PH domain proteins.....	109
Bibliography	110
Vita.....	116

List of Illustrations

Chapter 1 (no illustrations)

Chapter 2

Figure 2.1: KEGG pathway analysis of PH domain proteins	9
Figure 2.2: Visualization of overrepresented signaling pathways	10
Figure 2.3: Clustering of clinical samples across 13 most common cancer types based on expression level of all 313 PH domain proteins.	12
Figure 2.4: Clustering of clinical samples across 13 most common cancer types based on Top20 most differentiated expressed PHGs.	13
Figure 2.5 PHG mutations across 13 major cancer types in the TCGA study:	15
Figure 2.6: Tiam1 is one of the most frequently mutated gene across 13 cancer types	16
Figure 2.7: Plot of OG and TSG mutation of PH domain proteins.	17
Figure 2.8: Mutation of Tiam1 status across cancer types.	18
Figure 2.9: Plot of GOF and LOF mutation of Tiam1.....	19
Figure 2.10: Accumulated mutations in PH domains.....	20
Figure 2.11: Multiple sequence alignment of PH domain proteins.	22
Figure 2.12: Conserved residues of PH domain mapped to AKT PH domain crystal structure	23
Figure 2.13: Multiple sequence alignment of all PH domain proteins annotated in TCGA.	24
Figure 2.14: Clustering of all Crystal structure of PH domain proteins.	26
Figure 2.15: A heatmap of PIPs binding specificity.	28
Figure 2.16: A snapshot of the database PH domain and their PIPs binding information.	30
Figure 2.17 Confusion matrix and ROC curve of the best prediction model.	31
Figure 2.18 Display of the webserver for PH domain proteins.	32

Chapter 3

Figure 3.1: Workflow of in silico screen process.	69
Figure 3.2: Root mean square fluctuations of C-alpha Atoms of residues in Tiam1 protein.	77
Figure 3.3: Selected snapshots from MD simulation for ensemble docking study.	78
Figure 3.4: RMSD between each pair of structures.	79
Figure 3.5: Structural sequence alignment of Tiam1 protein crystal structures.....	83
Figure 3.6: Electrostatic potential surface maps of Tiam1 cPH domain reveal positive charge around the putative binding pocket.....	84
Figure 3.7: Predicted PI3P and PI5P binding poses with best docking scores.	85
Figure 3.8: A structure-based pharmacophore was generated using GRID method as a filter of the virtual screening hits.....	88
Figure 3.9: Putative binding site of TPH3 in Tiam1 cPH domain.	89
Figure 3.10: Chemical structures of the hits.	90
Figure 3.11: TPH3 binds to the Tiam1 cPH domain.	91
Figure 3.12: TPH3 inhibits Rac1 activity in cells.	92
Figure 3.13: TPH3 inhibits wound healing in prostate cancer cell line.	94
Figure 3.14: TPH3 inhibited lamellipodia formation in prostate cancer cells..	95
Figure 3.15: TPH3 inhibits prostate cancer cell invasion in invasion assay.	96
Figure 3.16: TPH3 inhibits prostate cancer cell proliferation in colony assay.	97
Figure 3.17: Complex structure refinement and Prediction of molecular interaction between TPH3 and cPH Tiam1.....	99
Figure 3.18: Bound ligands induce Tiam1 cPH domain conformational changes and stabilize protein complexes.	101

Chapter 4 (no illustrations)

List of Tables

Chapter 1 (no tables)

Chapter 2

Table 2.1: All 313 PH domain proteins.....33

Table 2.2: Parameters of Convolutional neural network.38

Table 2.3: List of proteins predicted to bind to PIPs.39

Table 2.4: Classification of mutations among all 313 PH domain proteins.40

Chapter 3

Table 3.1: Structures used in structural multiple sequence alignment.74

Chapter 4 (no tables)

Abbreviations

μ L: microliter

μ M: micromolar

PI3K: phosphatidylinositol-3-kinase

GFP: green fluorescent protein

PDB: Protein Data Bank

TCGA: The Cancer Genome Atlas

ROC: Receiver operating characteristics

PH domain: Pleckstrin Homology domain

AUC: Area under the curve

RMSD: Root mean square deviation

GEF: Guanine nucleotide factor

RMSE: Root mean square error

Tiam1: T-lymphoma invasion and metastasis-inducing protein 1

H-bond: Hydrogen bond

PI(3)P: phosphatidylinositol (3)-phosphate

PI(4)P: phosphatidylinositol (4)-phosphate

PI(5)P: phosphatidylinositol (5)-phosphate

PI(4,5)P: phosphatidylinositol (4, 5)-bisphosphate

PI(3,4,5)P: phosphatidylinositol (3,4, 5)-trisphosphate

Chapter 1: Introduction

1.1 PH domain as a drug target

PH domain was first noted in pleckstrin, which contains two regions with high sequence similarity [1, 2]. As one of the most common protein domains in the human proteome, PH domains have very conserved secondary structures: seven beta sheets and one alpha helix at the C terminus, although with relatively low sequence identity. This protein domain containing about 120 amino acids is involved in intracellular signaling or serve as critical constituents of the cytoskeleton.

One of the most important features of these proteins is that they bind to phosphatidylinositol lipids (such as phosphatidylinositol (4, 5)-bisphosphate and phosphatidylinositol (3,4,5)-trisphosphate) and recruit proteins to the membranes of different cellular organelles. PIs consist of a water-soluble Myo-inositol head group linked by a glycerol moiety to two different fatty acid chains, usually a saturated C18 residue in the 1-position and a tetra-unsaturated C20 residue in the 2-position [3]. Unphosphorylated phosphatidylinositols (PtdIns), usually synthesized in the endoplasmic reticulum, are transported to other membranes via PtdIns transfer protein [4, 5]. PIs bind to different cell membranes via two lipid tails. They also directly interact with proteins and regulate their functions via the water-soluble inositol head group.

The first type of PH domains bind to cytoplasmic membrane via PI(4,5)P₂. Phospholipase C-delta (PLC- δ) was the first PH domain protein that demonstrated the binding specificity to PI(4,5)P₂ [6, 7]. Such interactions were found to be required in the recruitment of PH domain

proteins to the cytoplasmic membrane using green fluorescent protein (GFP) label [8, 9]. Later, PH domains binding to PI(3,4,5)P₃ and PI(3,4)P₂ were also found to be recruited to cytoplasmic membranes [10, 11] For example, AKT PH domain recognizes PI(3,4,5)P₃ and PI(3,4)P₂, but does not bind to PI(4,5)P₂ [10, 12]. AKT will be recruited to the cytoplasmic membrane with the presence of these PIs. Other PH domain proteins that recognize PI3K products include BTK and GRP1. In contrast to these PIPs, the binding specificity to PI3P, PI5P and PI(3,5)P₂ is far less well studied. The C-terminal TAPP1 PH domain may bind to these monophosphate PIPs, but shows relatively weak binding affinities [13]. Interestingly, binding of PI4P has been reported to specifically target the Golgi apparatus [14], although such binding alone may be not strong enough to drive the targeting and require assistance of other proteins like Arf1p [15]. OSBP and FAPP1 PH domains are examples of proteins targeting Golgi apparatus. Also, PH domains are known to mediate signaling transduction through protein-protein interactions [16]. In summary, PH domain proteins are implicated in multiple signaling pathways and they are potentially important targets for drug discovery.

1.2 PH domain proteins in cancer

Membrane recruitment has been noticed to be related to carcinogenesis. One of the best studied is the PIP₃ signaling [17]. Through phosphorylation of PI (4, 5)P₂ by the phosphatidylinositol-3-kinase (PI3K), PIP₃ is accumulated in the cytoplasmic membrane and recruits PIP₃ specific binding PH domain proteins such as AKT and PDK1 to the cell membrane [18, 19]. The concentration of PIP₃ is upregulated by oncogenes like Ras and degraded by PTEN, which dephosphorylates PIP₃ to PI(4,5)P₂ [20, 21]. Mutations in the PH domain was first systematically reported in 2005 [22]. Carpten *et al.* identified E17K

mutation, which was located in the PIP3 binding pocket of the AKT PH domain in 9 out of 162 cancer patients. This mutation increased the PIP3 binding affinity through replacing a negatively charged residue to a basic residue. Moreover, it was also observed that this mutation decreased the sensitivity to allosteric kinase inhibitors. Later, more and more driver mutations were reported in the PIP3 signaling pathways such as the RAS-PI3K-AKT axis [23]. For example, a mutation in the PDK1 PH domain causes inhibition of AKT and insulin resistance [24].

1.3 Current situation of developing small molecule inhibitors targeting PH domains

The initial interest of developing small molecule inhibitors of PH domains was to develop safe and potent AKT inhibitors. Although being one of the most critical oncogenes in the human genome, safe and selective AKT drugs have not been developed although intensively studied. Then researchers switched their interest to see if they can find small molecule inhibitors that bind to AKT PH domain. Initially, lipid-based derivatives were synthesized to mimic PI analogs [25-27]. However, scientists quickly realized that these compounds had poor solubility and pharmacokinetics, although they showed some effect in cells [28]. After that, researchers recognized that novel chemical scaffolds were required to develop small molecule inhibitors targeting these domains. Mahadevan et.al. identified compounds that selectively bind to AKT PH domain [29]. In 2010, Miao *et.al.* identified two compounds that actively bind to AKT with $K_d \approx 43.2 \mu\text{M}$ through screening of 50,000 compounds. Recently, another scaffold was reported to inhibit AKT PH domain with $K_D = 3.08 \pm 0.49 \mu\text{mol/L}$ [30]. Also, several compounds targeting PH domains other than AKT were reported [31]. The activity of all these compounds are in

micromolar range and thus still far away from clinical use. As a result, there is an urgent need to discover more potent inhibitors to target these domains.

Chapter 2: Genomics, structural and PIPs binding specificity analysis of PH domain proteins

2.1 Introduction

Although it has been known that several PH domain proteins are involved in cancer mechanisms, lots of information about other PH domain proteins are still elusive such as the total number of PH domain proteins in the human proteome, frequency, and types of the genetic alterations of these proteins in cancer patients, and PIPs binding specificity.

Herein, we extracted all proteins with annotations as PH domain proteins from the InterPro database and Uniprot website, which generated 313 PH domain proteins in total. Then we downloaded the expression level and mutation data of 13 types of most common cancer in The Cancer Genome Atlas (TCGA) dataset to explore the genetic changes of these 313 proteins. KEGG pathway analysis was performed to analyze which pathways these genes are significantly overrepresented. Clustering analysis was performed to identify the expression pattern of these genes across 13 types of cancers. Somatic mutation analysis of the 313 genes was performed to identify most frequently mutated PH domain genes in different types of cancers. Especially, mutations within PH domains were extracted and discussed separately.

Then all the PH domain proteins in the PDB database were downloaded to perform structural multiple sequence alignment to identify the recognition pattern of the PIPs binding specificity. Also, all PH domain proteins annotated in the TCGA database were also aligned to identify conserved residues of PIPs binding.

Due to the limitation of the amount of data related to PH-PIPs binding affinity, we downloaded all abstracts published on PubMed to extract all PH-PIPs binding information to build a model to predict the binding specificity of all PH domain proteins using the convolutional neural network. A database with the PH domain protein information and PIPs binding specificity was generated and made available to the public online.

2.2 Methods and materials

2.2.1 KEGG pathway analysis of PH domain genes

Although numerous studies about PH domain have been reported, the number of PH domain proteins is inconclusive in human proteome due to their diverse and integrative nature. We extracted all proteins annotated as PH domains in the InterPro web server [32]. Then duplicate proteins were removed based on their Gene IDs. As a result, a total of 313 PH domain proteins and their gene IDs were retrieved (**Table 2.1**). Some of the proteins have not been reviewed by the UniProt consortium [33]. However, these proteins were still kept in our list because they were annotated to contain PH domains and they comprised only a small part of the whole protein list. Then we examined the distribution of these PH domain genes within KEGG Pathways. The gene list was uploaded to the DAVID web server and converted to official gene symbols. Also, an overrepresentation test of these proteins among KEGG pathway was performed.

2.2.2 Somatic mutation analysis

We obtained the somatic mutation data from TCGA Pan-cancer effort on <https://www.synapse.org>. To decrease the noise from passenger mutations, samples with more than 500 somatic mutations (hypermutators) were removed from our study. Samples with no

somatic mutations were removed as well, resulting in mutations from 1,511 tumors for the following clustering analysis. Only non-silent somatic mutations were included in the analysis. SomInaClust[34] was used to identify genes with mutation patterns which resemble either those found in oncogene or tumor suppressor gene at a q value of 0.1. Hotspot mutation was defined as in-frame or missense mutations at the same amino acid in more than two samples.

2.2.3 Unsupervised clustering

Expression levels of 313 PH domain genes across the 3,281 tumors were collected. Matrix (sample \times gene) with mutation status and gene expression levels were constructed and passed to perform complete-linkage hierarchical clustering using R function ‘hclust’. Also, heatmaps with dendrograms were visualized using R function “heatmap.2” in the gplots package.

2.2.4 Curation of PH domain proteins from PDB

Crystal structures of PH domain proteins were downloaded from the PDB website and duplicate structures were removed. In total, 34 unique structures were used to build a maximum likelihood (ML) tree based on their structure-based sequence alignment.

2.2.5 Structural sequence alignment and weblogo generation

All PH-PIPs structure complexes available in PDB (**Table 5.1**) were collected to perform multiple sequence alignment based on their secondary structures using STRAP [35]. The output of the alignment was then used to generate the signatures of conserved residues involved in PIP binding using Weblogo web server[36].

2.2.6 Datasets curation and database generation

PH-PIPs binding data was curated through text mining of all abstracts published on PubMed. First, all abstracts were downloaded from PubMed; then all the abstracts were split into single sentence; finally, all the sentences include “PH domain”, “Pleckstrin homology domain”, “bind” and “bound” were extracted and saved for the following analysis. All the extracted sentences were manually checked and put into a database include the following information: Protein name, PIP binding affinity, reference, PubMed ID of the literature and annotation.

2.2.7 Classification of PIP binding using convolutional neural network

A convolutional neural network is a feed-forward artificial neural network which has been widely used in identifying patterns and classifying images. We used Keras and python3 to build deep neural network models. And our final model was comprised of two layers of convolution layers and two layers of maxpooling layers. The detailed description of the model setting was described in **Table 2.1**.

2.3 Results

2.3.1 PH domain genes were overrepresented in multiple pathways

Among all the 313 PH domain genes, only 105 genes were annotated in David KEGG pathways. Consistent with previous reports, the Ras signaling pathway, actin cytoskeleton, phagocytosis, and chemokine signaling pathways were the most significantly overrepresented pathways among PH domain genes. Interestingly, the most significant overrepresented pathway was endocytosis, which has not been reported widely. Also, immune system pathways such as B cell and T cell receptor signaling pathways were also overrepresented among these PH domain genes (**Figure 2.1**). Visualization of the pathways reveals that these genes are also significantly involved in the protein metabolism, the small molecule transportation, and the cell cycle pathways (**Figure 2.2**). These discoveries suggested future research directions for PH domain proteins, such as small molecule transportation and immune functions.

Figure 2.1 KEGG pathway analysis of PH domain proteins

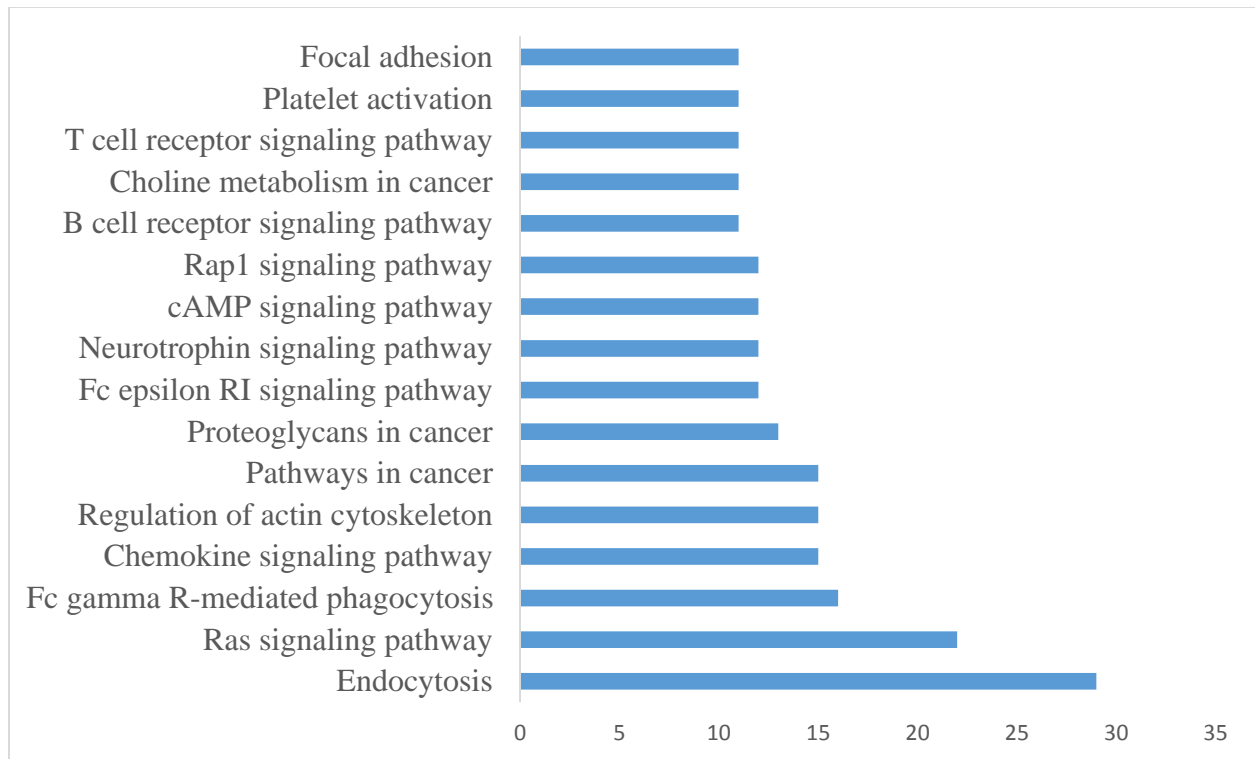
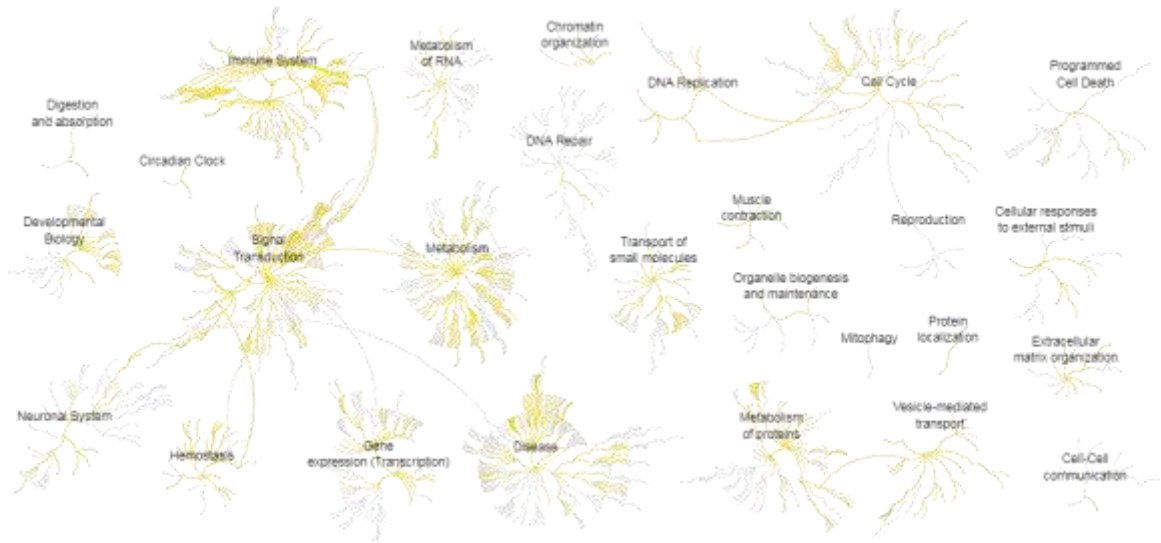


Figure 2.2 Visualization of overrepresented signaling pathways



2.3.2 A 20-gene signature resulted in five clusters of clinical samples across 13 cancer types

To further investigate the expression levels of these genes across different cancer types, we downloaded the mRNA expression level and mutation data of 13 types of cancer from <https://www.synapse.org>. Unsupervised hierarchical clustering of the clinical samples based on expressional levels of all 313 PH domain genes (PHGs) did not reveal obvious distribution patterns (**Figure 2.3**). Next, we selected the top 20 PHGs with the largest standard deviation (SD) to perform the clustering analysis again, which resulted in five main clusters (**Figure 2.4**). Interestingly, KIF1A, CADPS, PLEKHN1, STAP1 and MCF2 were upregulated in Cluster 1; GRB14, DOK5, STAP1, RASAL1, and MCF2 were upregulated in Cluster2, which was comprised by Colon adenocarcinoma (COAD) samples; CADPS, STAP1, and MCF2 were upregulated in Cluster 4, which mainly consists of Heck and Neck Squamous Cell Carcinoma (HNSC) and Glioblastoma (GBM); CADS, PLEKHN1, CNKSR2, MCF2, RTKN2, DOK2, and RASGRF1 were upregulated in Cluster 5, which was mainly comprised of Kidney Renal Clear Cell Carcinoma (KIRC).

Figure 2.3 Clustering of clinical samples across 13 most common cancer types based on expression level of all 313 PH domain proteins.

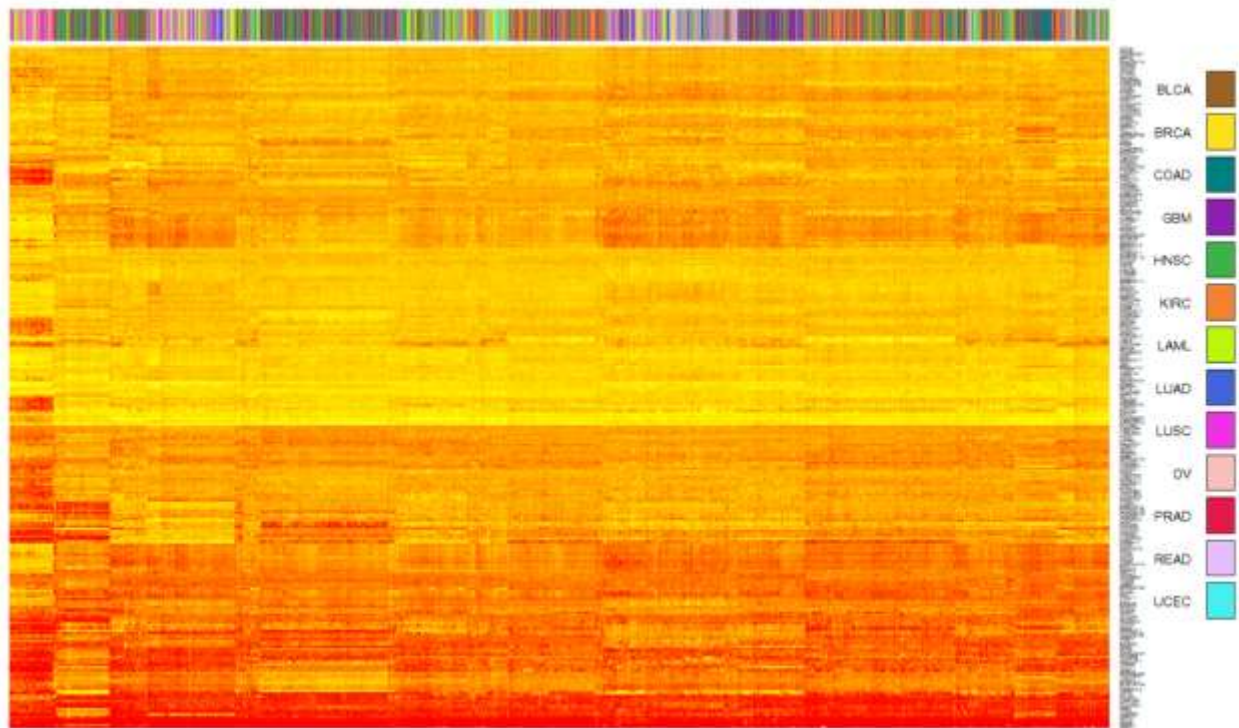
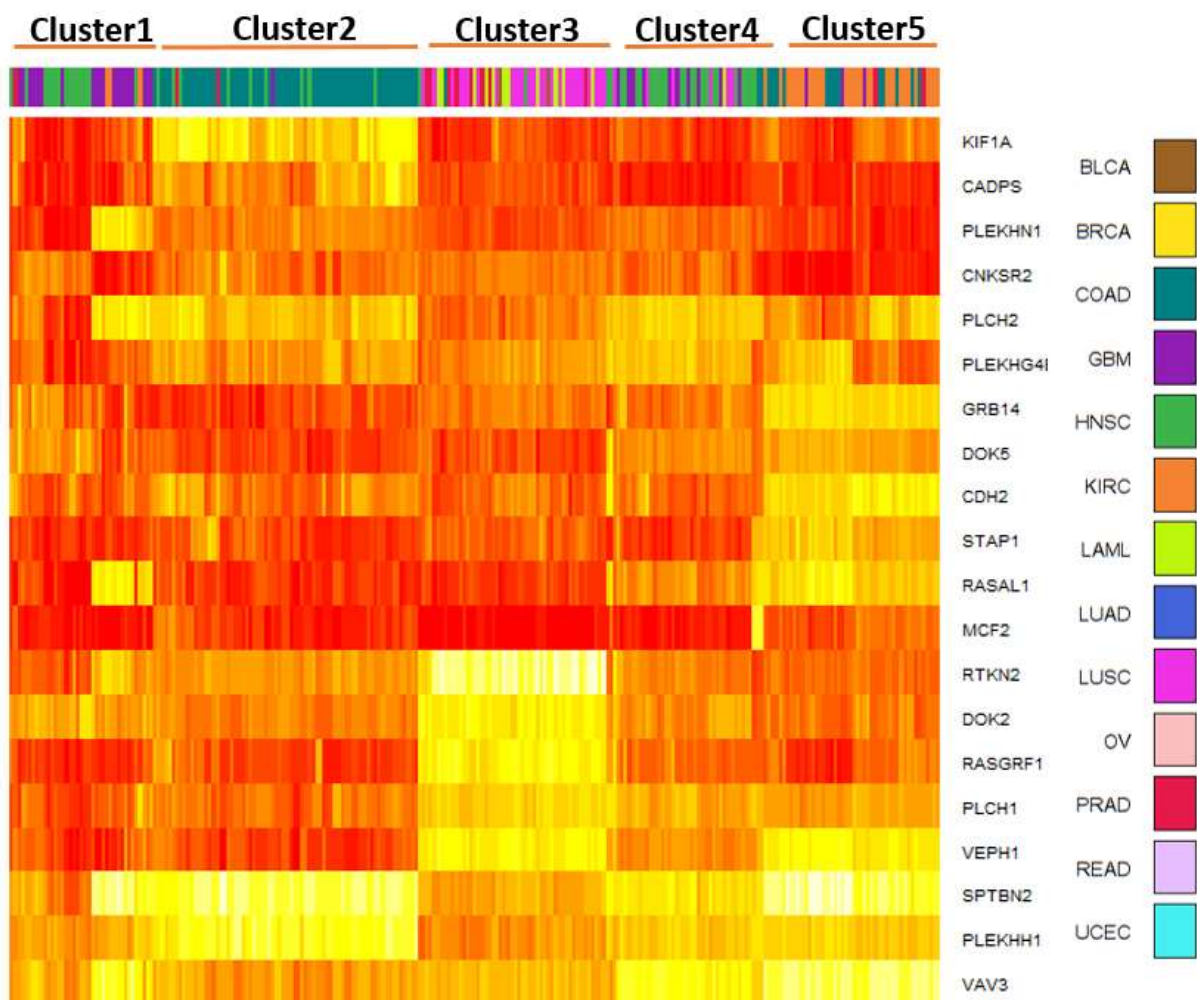


Figure 2.4 Clustering of clinical samples across 13 most common cancer types based on Top20 most differentiated expressed PHGs.



2.3.3 Tiam1 is one of the most frequently mutated PH domain genes across 13 cancer types.

Next, we wanted to investigate the mutation status of the PH domain proteins across the 13 cancer types. Only non-silent mutations were considered in the analysis. To this end, we first annotated genetic alterations to these genes. Whole exome sequencing identified 12768 non-silent coding mutations. Uterine Corpus Endometrial Cancer, Lung Squamous Cell Carcinoma and Lung adenocarcinoma have the highest mutation frequency of all PH domains (**Figure 2.5**). The top 40 most frequently mutated PH domain genes were further used to build a phylogenetic tree on the basis of the sequence similarity (**Figure 2.6**). Consistent with previous reports, AKT1 was one of the most frequent mutated genes not only among PH domain genes but also among whole human genome in different cancer types [37]. Interestingly, we found Tiam1, a Guanine nucleotide factor (GEF) specifically activates Rac1, was one of the top 10 most frequently mutated genes. It was most frequently mutated in lung adenocarcinoma, uterine corpus endometrial cancer as well as head and neck squamous cell carcinoma (**Figure 2.7-2.9**). Then we asked which PHDGs had the most frequent mutations in PH domains. We extracted all the mutations mapped to PH domains and measure the accumulated mutations at each residue on the basis of multiple structural sequence alignment (**Figure 2.10**). Interestingly, Tiam1 was among the most frequently mutated PHGs. Other frequently mutated PHGs include AKT1, PLEK, SKAP2, APBB1IP, ITK, PLEKHA6, ARHGAP15, RASGRF1, OSBPL8, ARHGAP24, DOCK11, AKT3, ARAP1, and ARHGEF7. All the calculated data used to predict OGs and TSGs were listed in **Table 2.4**.

Figure 2.5 The number of PHG mutations per patient across 13 major cancer types in the TCGA study. As observed in the figure, uterine, colon, bladder, lung, and head and neck cancer patients tend to carry more mutations on PHDG. On the contrary, breast, glioblastoma, kidney, ovarian, prostate, and leukemia patients rarely carry mutations on PHDG.

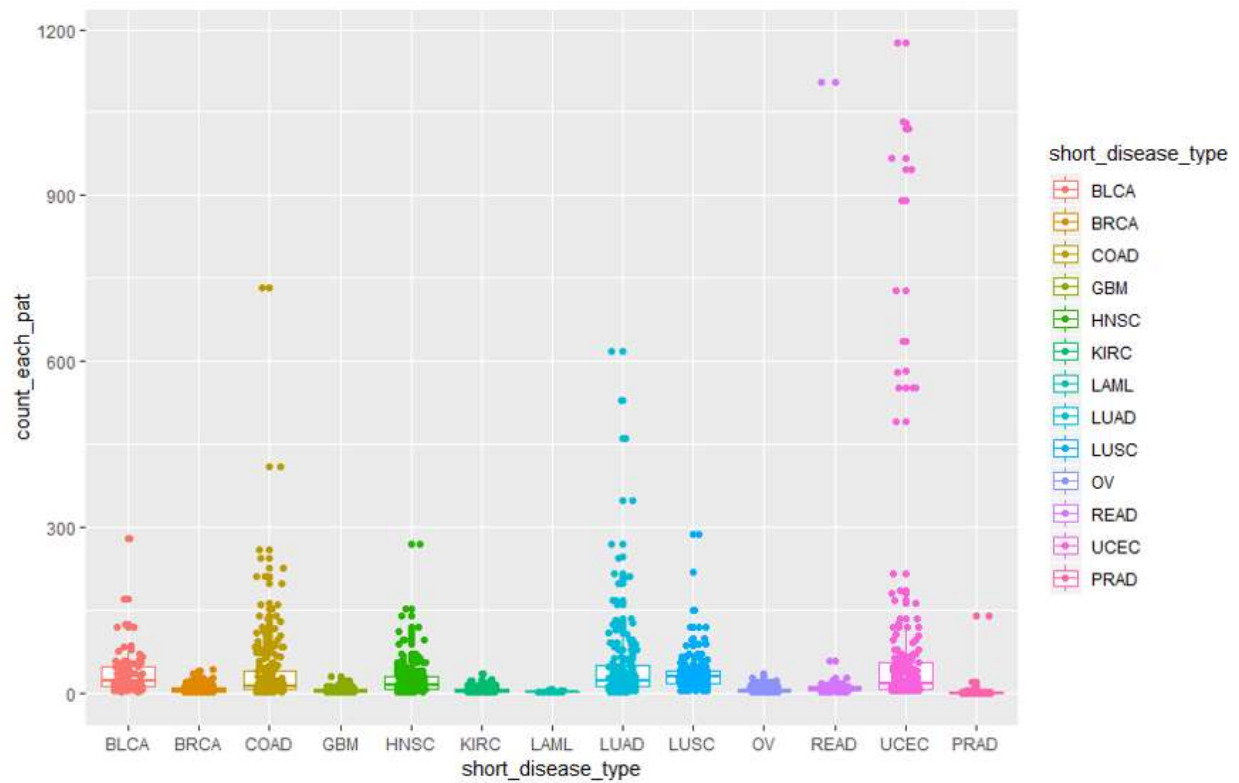


Figure 2.6 Tiam1 is one of the most frequently mutated gene across 13 cancer types.

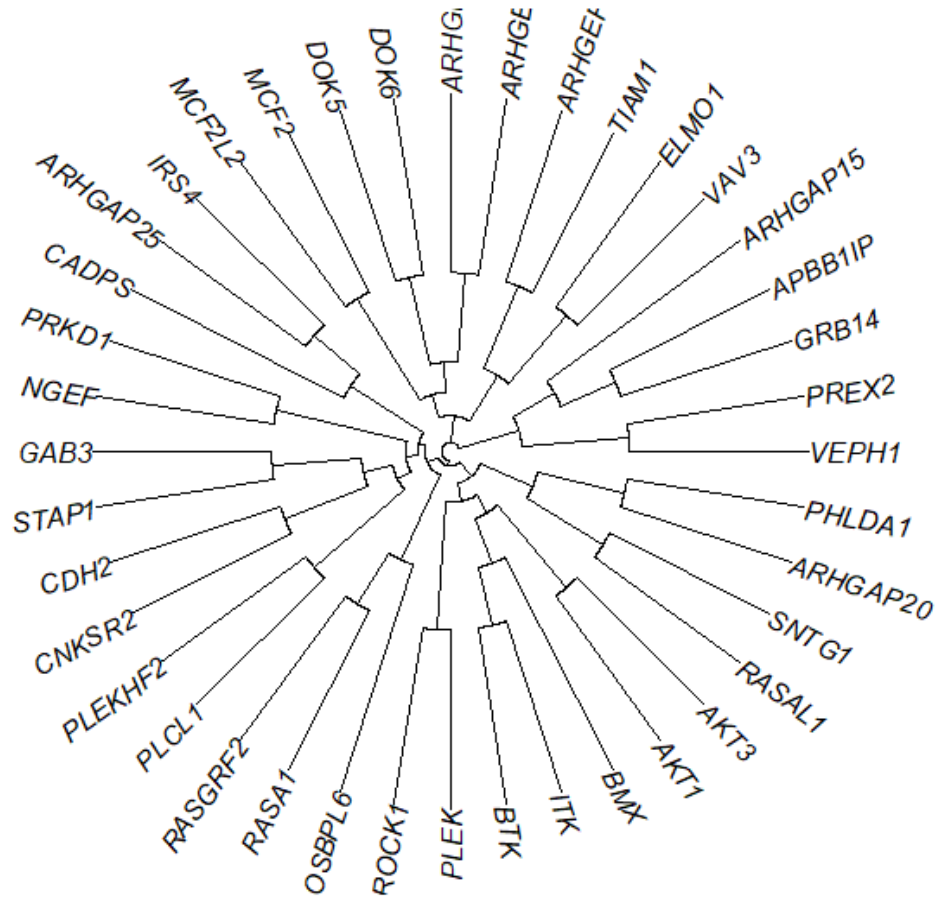


Figure 2.7 Two of the PHDGs, AKT1 and SOS1, had somatic mutation pattern which significantly resembles that of oncogenes (OG, $q < 0.1$). On the other hand, 80 of the PHGs had tumor-suppressor-gene-like (TSG-like) pattern.

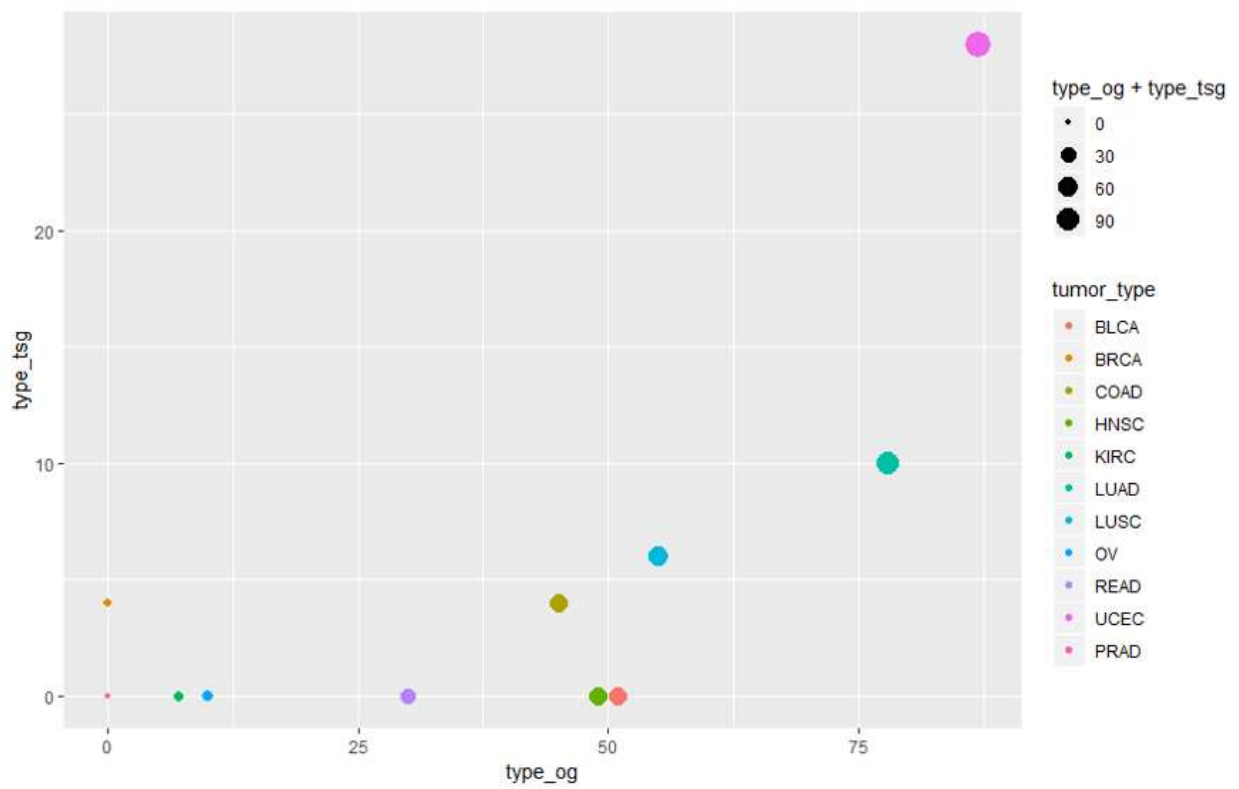


Figure 2.8 Mutation of Tiam1 status across cancer types.

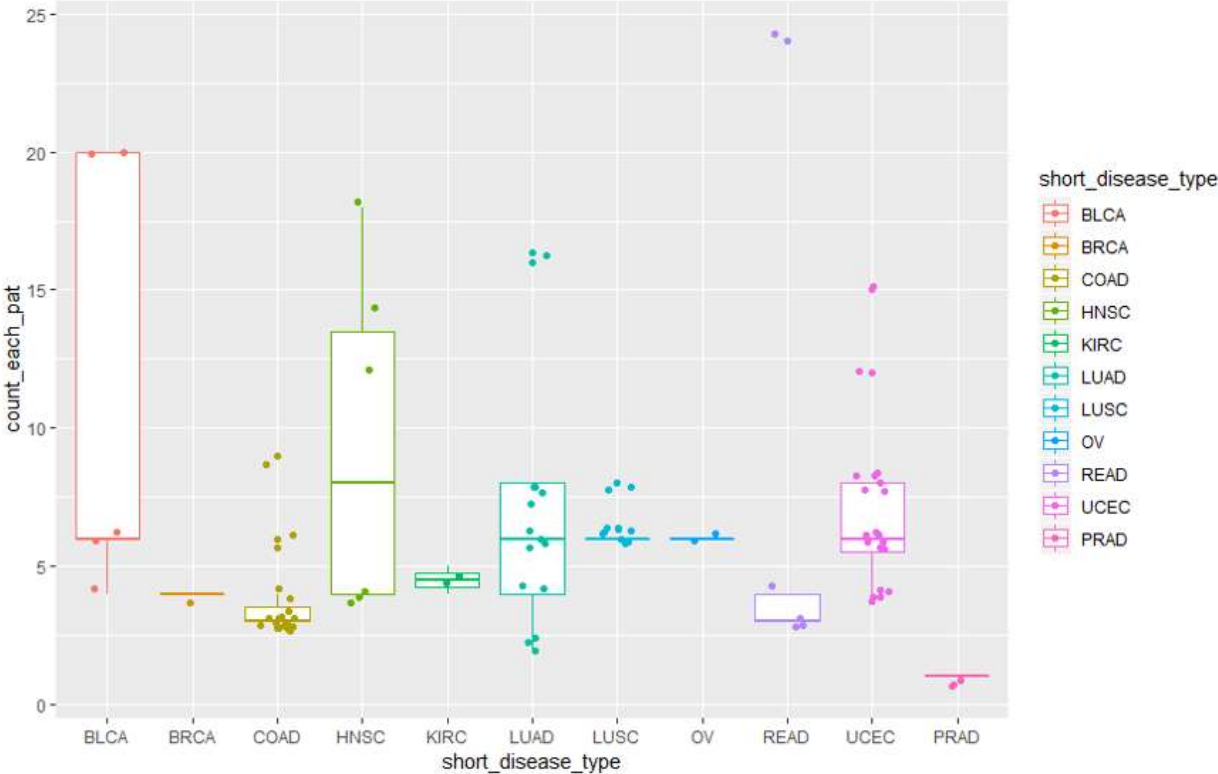


Figure 2.9 Gain-of-function (GOF) and loss-of-function (LOF) mutations of Tiam1 gene across 13 types of cancers. Uterine cancer had most GOF and LOF mutations. Tiam1 mutations in ovarian, kidney, rectal, bladder, head and neck cancers were exclusively GOF while the mutations of Tiam1 in breast cancers were all LOF.

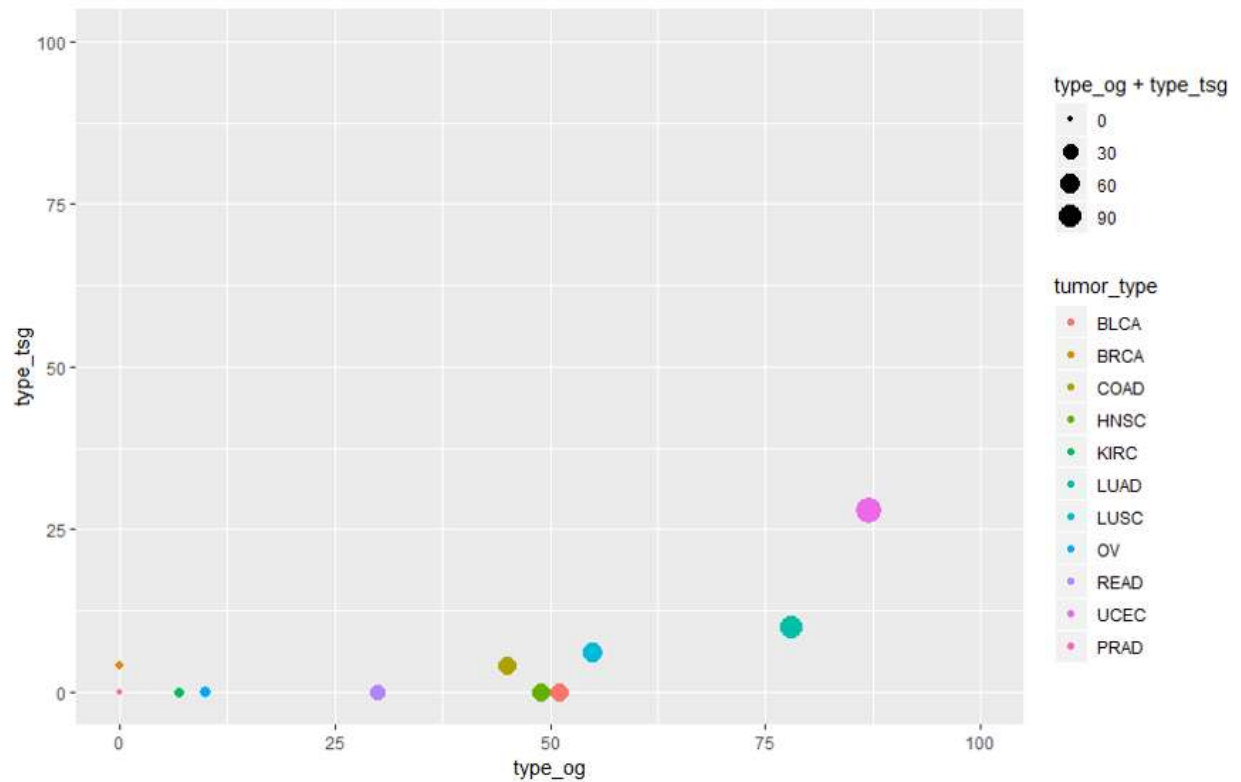
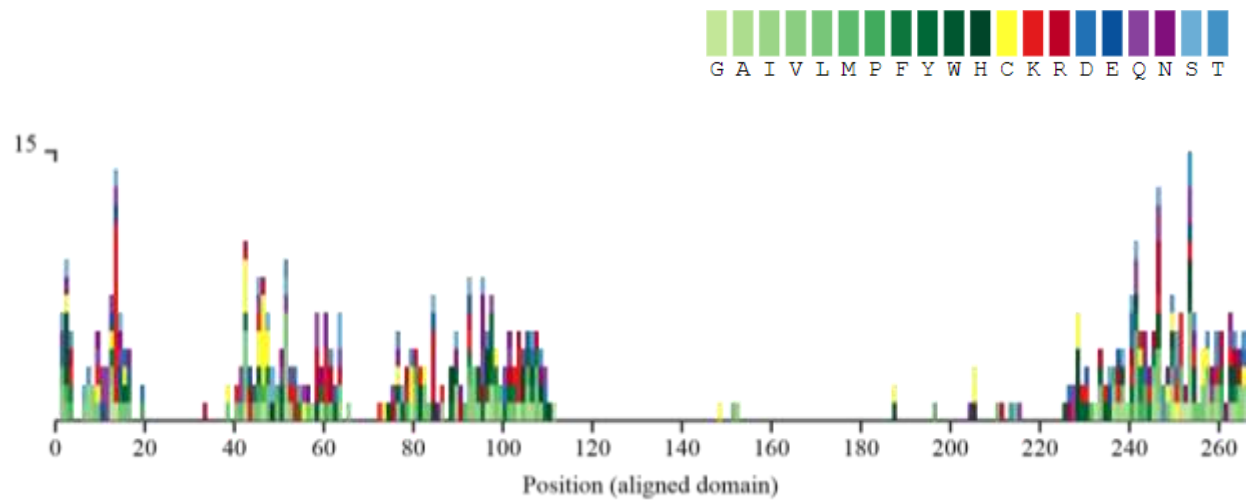


Figure 2.10 Accumulated mutations in PH domains.



2.3.4 Structural analysis reveals conserved residues in PH domains.

The PH domains have highly conserved secondary structure, although they only share 20-40% sequence identity. We would like to explore the structural features of the PH domain proteins further and to see if they share any similar properties in common. We performed multiple sequence alignment of all PH domain proteins using MutationAligner [38] in cBioPortal web server [39]. Interestingly, several conserved residues were observed in the alignment. For example, Trp11, Lys14, Arg23, Tyr26 (in the nomenclature of AKT1) were residues highly conserved in the alignments of PH domain proteins (**Figure 2.11**). These conserved residues were mapped to the AKT1 structure in **Figure 2.12** Detailed multiple sequence alignment of all PH domain proteins annotated in TCGA was presented in **Figure 2.13**.

Figure 2.11 Multiple sequence alignment of PH domain proteins.

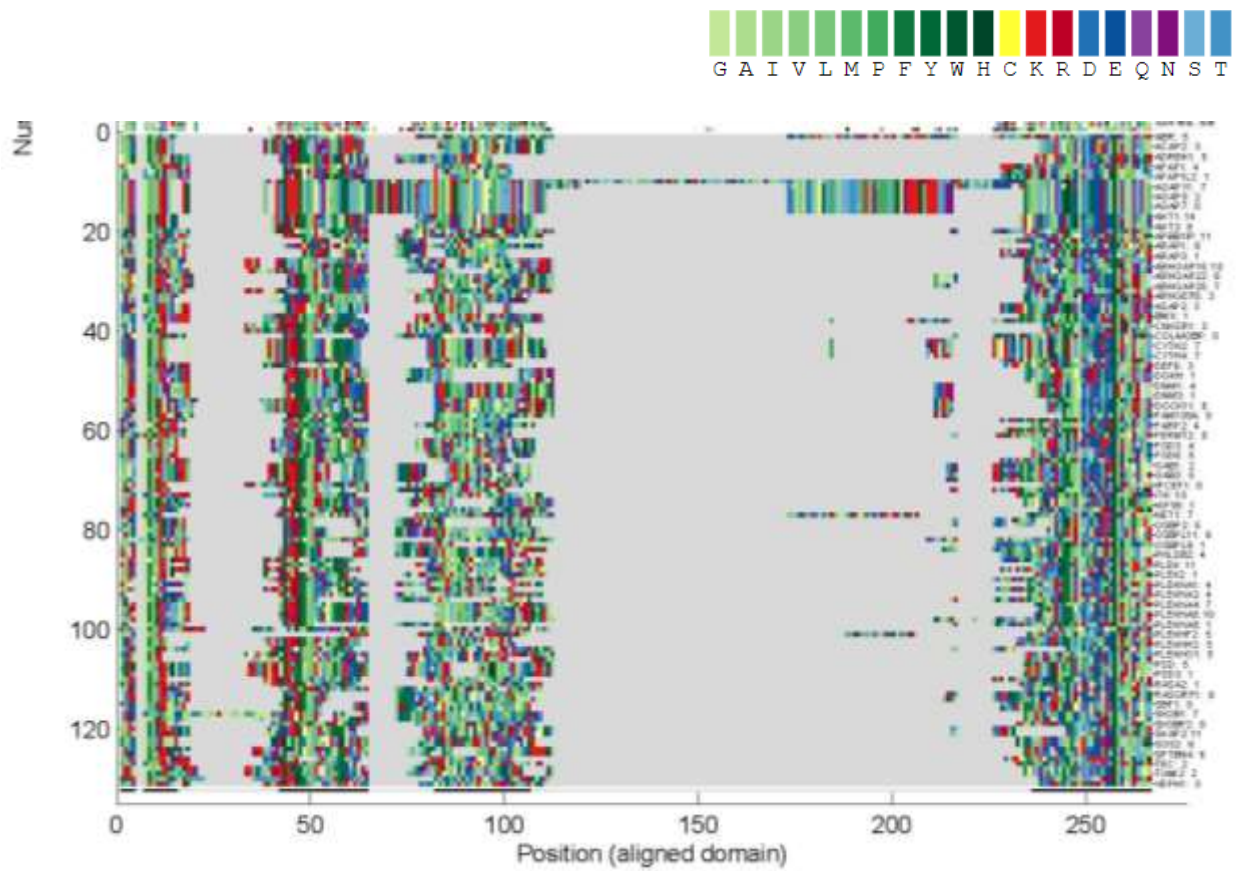


Figure 2.12 Conserved residues of PH domain mapped to AKT PH domain crystal structure.

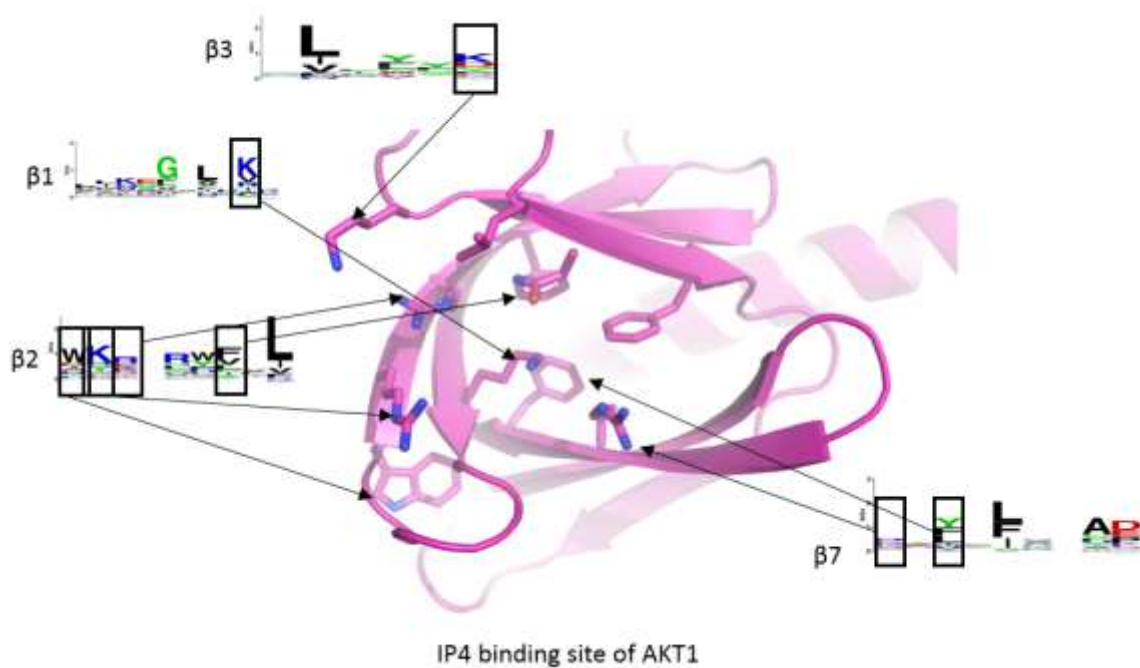
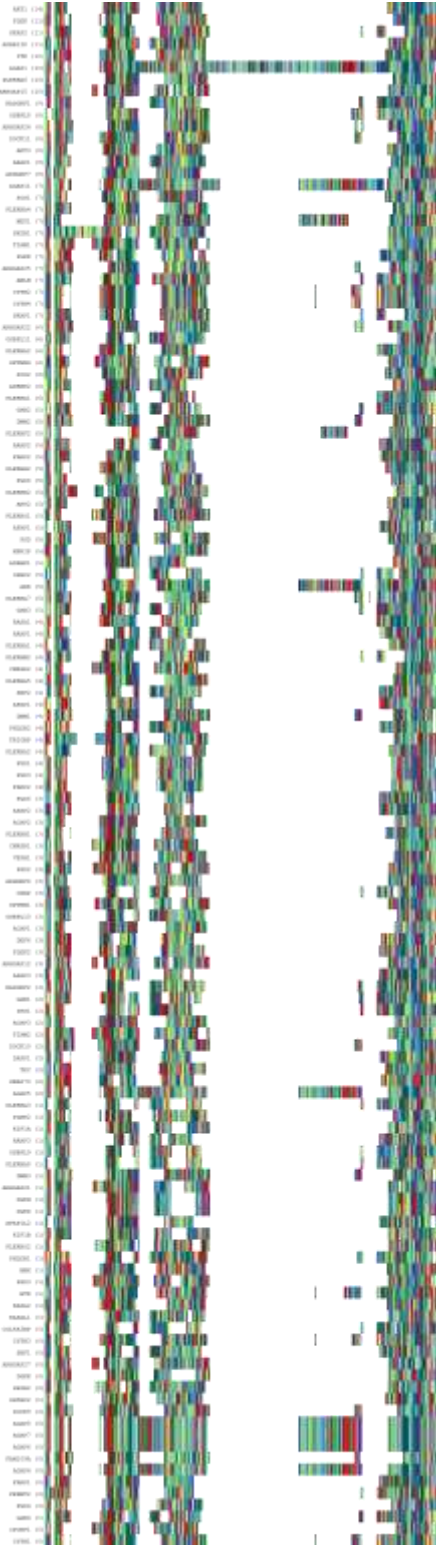


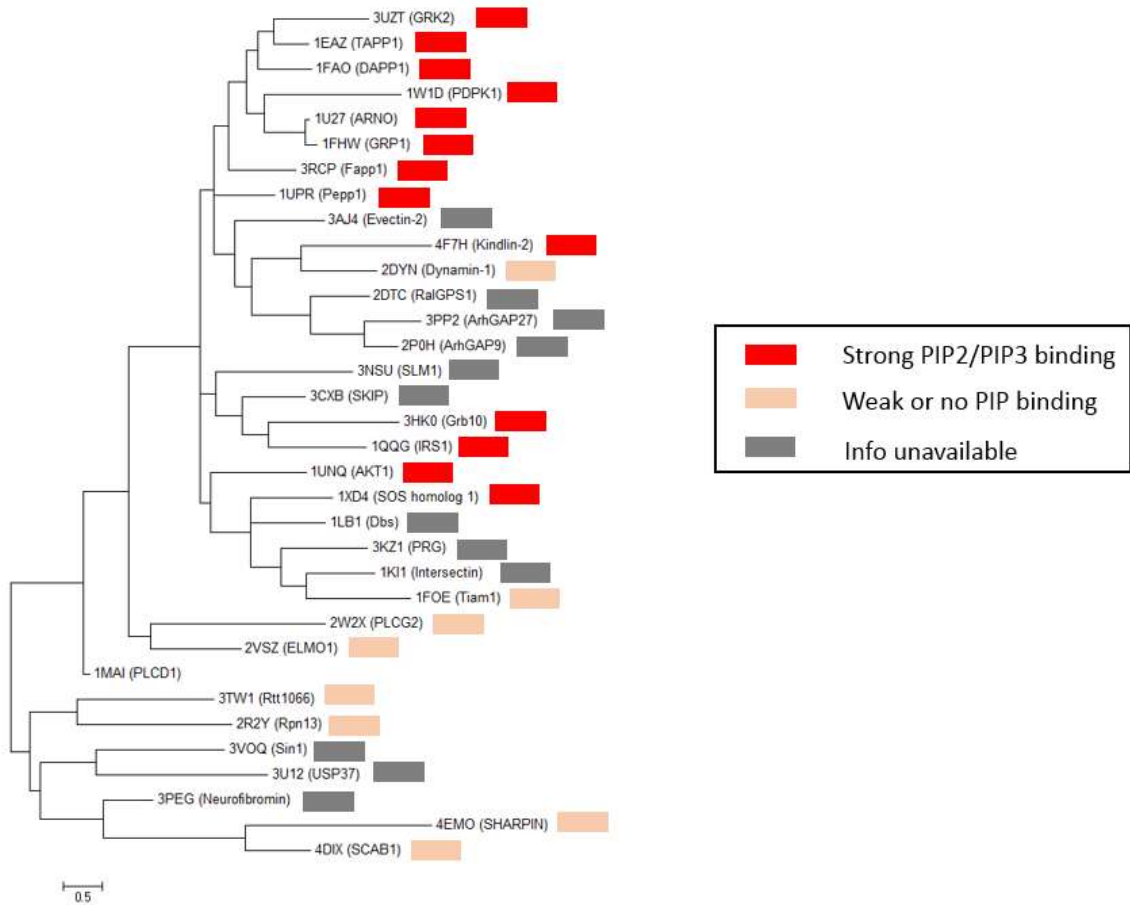
Figure 2.13 Multiple sequence alignment of all PH domain proteins annotated in TCGA.



2.3.5 Clustering of the available PH domain proteins.

Due to the difficulties of crystallization, structures of most PH domain proteins have not been determined so far. Structures of PH-PIPs protein complexes were even rarer in the PDB database. We collected all the available structures of PH domain proteins (duplicate proteins were removed) and performed cluster analysis based on their structure-based sequence alignment (**Figure 2.14**). Interestingly, GRK2, TAPP1, DAPP1, PDPK1, ARNO, GRP1, FAPP1, PEPP1, 4F7H were clustered in close groups and all of these proteins bound to PIPs. However, these proteins do not seem to have similar functions.

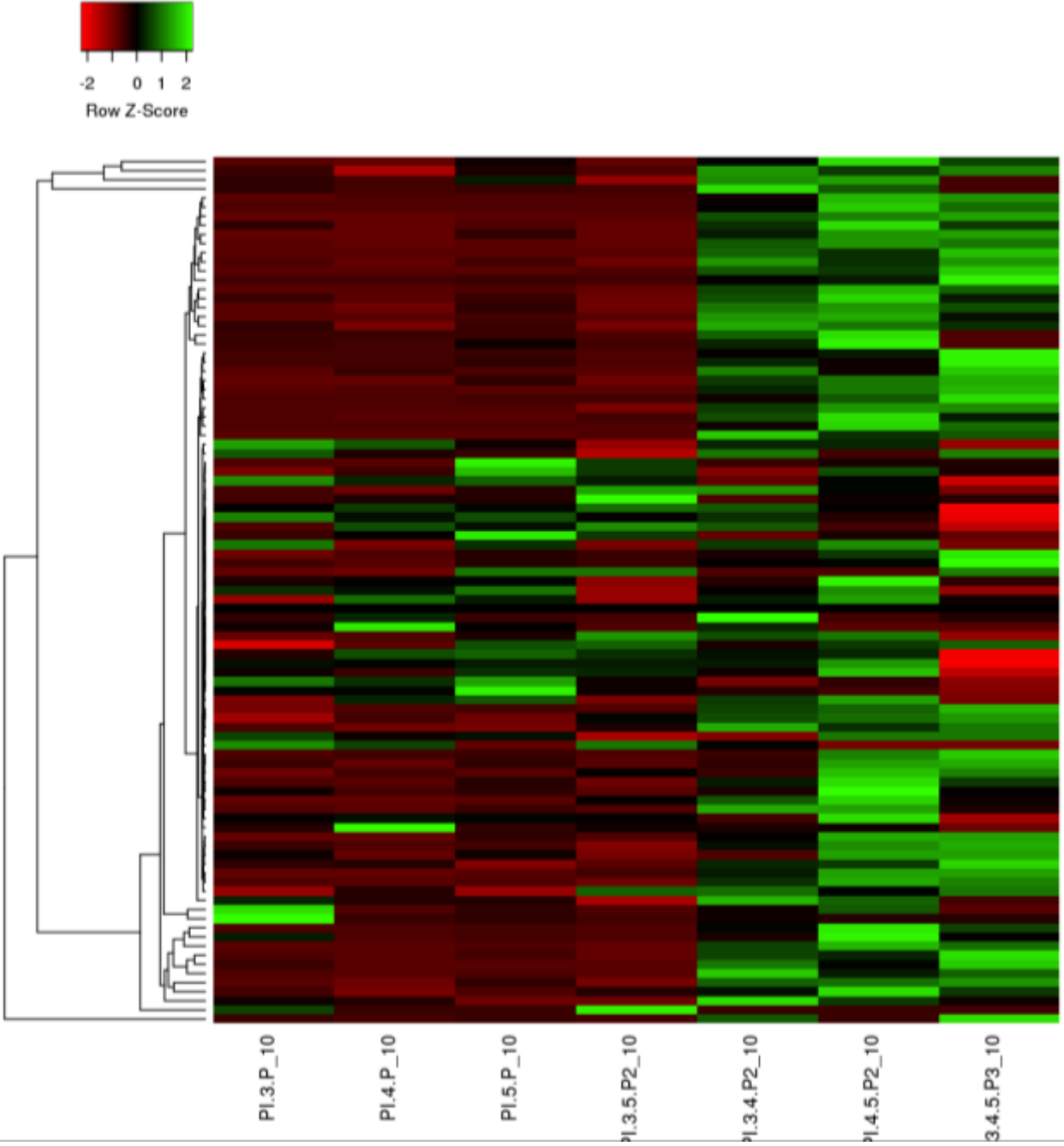
Figure 2.14 Clustering of all Crystal structure of PH domain proteins.



2.3.6 A heatmap of PIPs binding specificity based on published data

Due to the limited availability of PH domain protein structures, we collected published PH-PIPs binding data [40] and generated a heatmap of the PIPs binding affinity of 95 PH domain proteins (**Figure 2.15**). Interestingly, we found that proteins which bind to PI3P also bind to PI4P and PI5P. Moreover, these proteins prefer binding to PI(3,5)P₂, but not other types of PIPs. In other words, PI(3,5)P₂ has similar binding patterns similar to that of PI3P, PI4P, and PI5P. In contrast, proteins which bind to PI(3,4,5)P₃ more likely bind to PI(4,5)P₂ and PI(4,5)P₂. These data considerably raised our interest to explore the PIPs binding selectivity among all the PH domain proteins.

Figure 2.15 A heatmap of PIPs binding specificity.



2.3.7 Collecting PH-PIPs binding data from PubMed using text mining and building a classification model based on Convolutional neural network (CNN).

As described in the previous section, we downloaded all abstracts published in PubMed and extracted all sentences containing PIPs binding information of PH domain proteins. All binding information, along with binding affinity (if available), PubMed ID were used to build a database for the easy access later on. A part of the database was shown in **Figure 2.16**.

With the data collected, we built CNN based classification models to predict the binding ability of the 313 PH domain proteins. In total, eight models with prediction accuracy larger than 80% were generated. Then these models were applied to predict the other proteins on the list. The proteins predicted to bind to PIPs were listed in **Table 2.3** in descending of confidence. 41 out of 44 PH domain proteins that found to bind to PIPs were correctly predicted (sensitivity = 93.18%); and the absence of binding for 47 out of 49 PH domain proteins was correctly predicted (specificity = 95.92%). Confusion matrix which showed the prediction result was shown in **Figure 2.17**. ROC curve was plotted and the AUC is 0.98, indicating reliable prediction performance of our model. All these data are available in our web server and a snapshot of the webs server was shown as **Figure 2.18**.

Figure 2.16 A snapshot of the database containing PH domain and their PIPs binding information.

Protein	PIP	Binding Affinity	Sentence	Pubmed ID	Annotation
Tiam1	PtdIns(3)P	KD<5-10nM	While the PH domains of Intersectin and Dbs promiscuously bind several multiphosphorylated phosphoinositides, Tiam1 selectively interacts with phosphatidylinositol 3-phosphate (KID) approximately 5-10 microm)	11577097	
Tiam1	PtdIns(3,4,5)P²-PtdIns(3,4)P²->PtdIns(4,5)P			10908360	Bind to N-terminal PH domain
Tiam1	PtdIns5P		Restricted Rac1 activation results from the binding of Tiam1 DH-PH domains to PtdIns5P	24905281	Bind to DH-PH domain
Tiam1	PtdIns(3,4,5)P		We showed that liposomes of "resting cell membrane" composition (less than 20 mol % zwitterionic phospholipids), supplemented with 1 mol % of polyvalent anionic phospholipids) supplemented with 1 mol % of polyvalent anionic phosphatidylinositol 3,4,5-triphosphate [PtdIns(3,4,5)P(3)] in conjunction with constitutively active forms of the guanine nucleotide exchange factors (GEFs) for Rac, Tric, or Tiam1 and a non-hydrolyzable GTP analogue, cause dissociation of Rac1(GDP),RhoGDI complexes. GDP to GTP exchange on Rac1, and binding of Rac1(GTP) to the liposomes.	18505730	
Tiam1	PtdIns(4,5)P and PtdIns(3,4,5)P		Finally, the data demonstrate that PtdIns(4,5)P(2) and PtdIns(3,4,5)P(3) bind to the same pleckstrin homology domain in Tiam1 and that soluble inositol phosphates appear to compete with lipids for this binding.	15242348	

Figure 2.17 Confusion matrix and ROC curve of the model with best prediction ability.

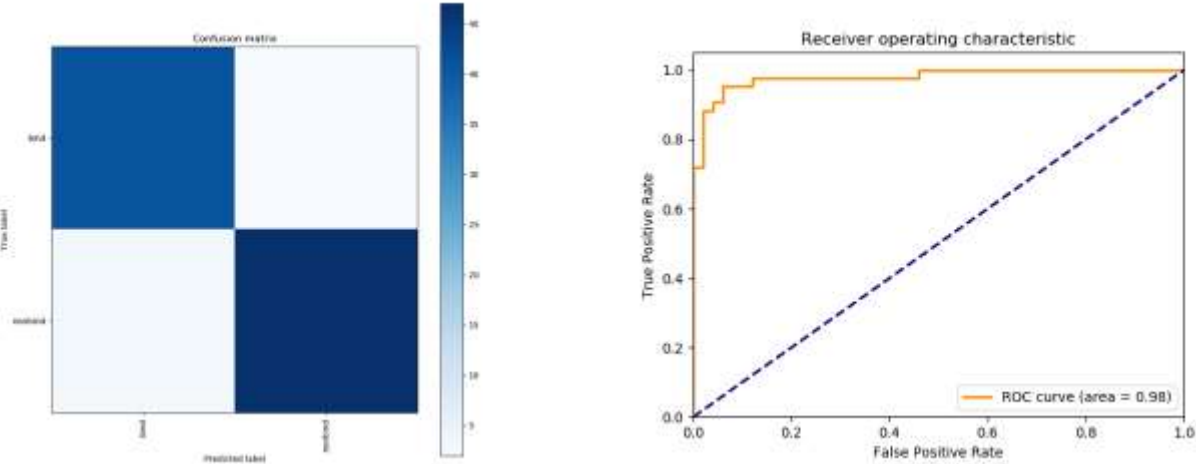


Figure 2.18 Display of the webserver for PH domain proteins.

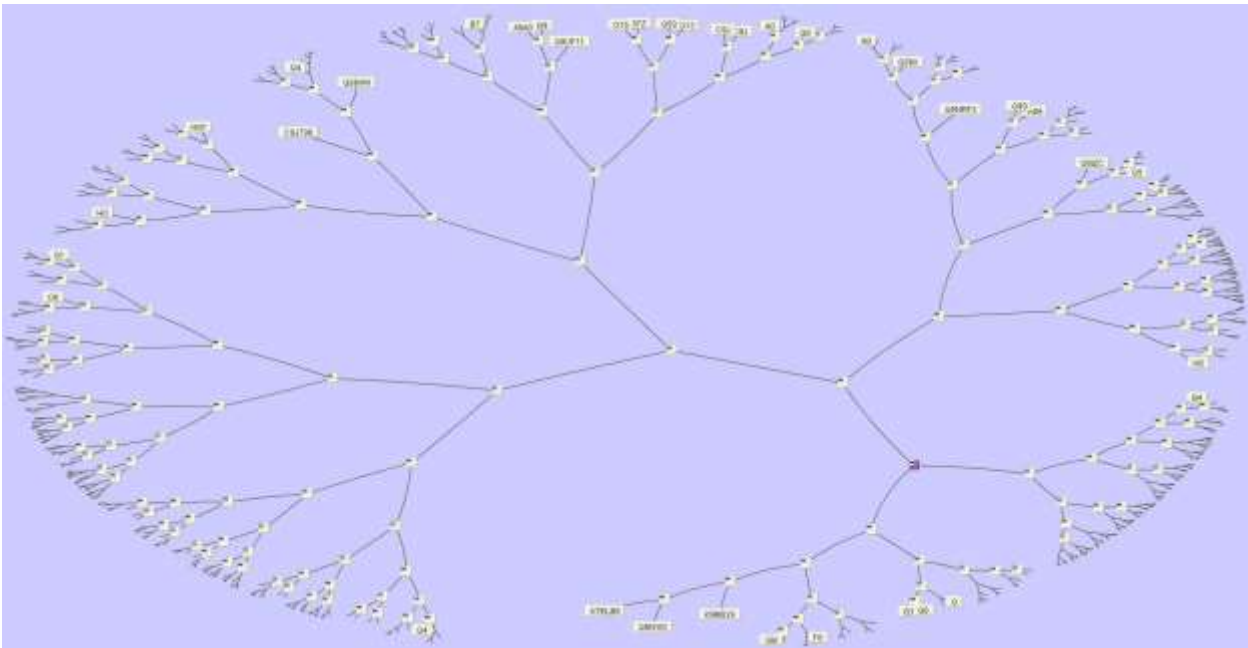


Table 2.1 All 313 PH domain proteins.

Protein ID	Gene ID	Protein ID	Gene ID	Protein ID	Gene ID	Protein ID	Gene ID
Q12979	ABR	Q99490	CENTG1	V9HWC8	HEL-S-308	Q6ZR37	PLEKHG7
Q15027	ACAP1	Q4ZG22	CENTG2	V9HW03	HEL-S-81p	Q9ULM0	PLEKHH1
Q15057	ACAP2	Q96P47	CENTG3	Q8WWN9	IPCEF1	Q8IVE3	PLEKHH2
Q96P50	ACAP3	O14578	CIT	Q6DN90	IQSEC1	Q7Z736	PLEKHH3
O75689	ADAP1	G9CGD6	CNK3/IPC EF1	Q5JU85	IQSEC2	K7EIZ3	PLEKHJ1
Q9NPF8	ADAP2	Q969H4	CNKSR1	P35568	IRS1	Q9Y4G2	PLEKHM1
P25098	ADRBK1	Q8WXI2	CNKSR2	Q96RG5	IRS2	Q8IWE5	PLEKHM2
P35626	ADRBK2	Q9Y5P4	COL4A3BP	O14654	IRS4	Q6ZWE6	PLEKHM3
Q8N556	AFAP1	Q15438	CYTH1	Q08881	ITK	Q494U1	PLEKHN1
Q8TED9	AFAP1L1	Q99418	CYTH2	Q15811	ITSN1	Q53GL0	PLEKHO1
Q8N4X5	AFAP1L2	O43739	CYTH3	Q9NZM3	ITSN2	Q8TD55	PLEKHO2
E7EUN2	AGAP1	Q9UIA0	CYTH4	J3QSW6	KALRN	Q5SXH7	PLEKHS1
Q8TF27	AGAP11	Q5VWQ8	DAB2IP	Q53R16	KIAA0053	Q8TCU6	PREX1
Q99490	AGAP2	Q9UN19	DAPP1	Q5W9H1	KIAA0142	Q70Z35	PREX2

Q96P47	AGAP3	Q53RS3	DDEF2	Q5W9G0	KIAA063 8	F8WBA3	PRKD1
Q96P64	AGAP4	Q8TDY4	DDEFL1	Q12756	KIF1A	Q9BZL6	PRKD2
A6NIR3	AGAP5	Q9H4E7	DEF6	O60333	KIF1B	O94806	PRKD3
Q5VW22	AGAP6	Q16760	DGKD	Q4R9M9	KIF1Bbet a	A5PKW4	PSD
Q5VUJ5	AGAP7P	Q86XP1	DGKH	P10911	MCF2	Q9BQI7	PSD2
Q5VTM2	AGAP9	Q5KSL6	DGKK	O15068	MCF2L	Q9NYI0	PSD3
Q12802	AKAP13	Q658P8	DKFZp313 N0632	Q86YR7	MCF2L2	Q8NDX1	PSD4
P31749	AKT1	Q9NTG0	DKFZp434 G2016	B9EGI2	MPRIP	Q5JS13	RALGPS1
P31751	AKT2	Q9UFY1	DKFZp434 N101	U6FSN9	Mprip- Ntrk1	Q86X27	RALGPS2
Q9Y243	AKT3	Q5HYM3	DKFZp686 C0249	A0A0A0 MQX1	MYO10	C9K0J5	RAPH1
Q9NQW6	ANLN	Q5HYD7	DKFZp686 K101	Q7Z628	NET1	P20936	RASA1
Q7Z5R6	APBB1IP	Q5HYB0	DKFZp686 P1738	Q8N5V2	NGEF	Q15283	RASA2
Q9UKG1	APPL1	Q9H3X2	DKFZp761 E2216	A6NGQ3	OBSCN	Q14644	RASA3

			DKFZp762				
Q8NEU8	APPL2	Q69YP8	A083	O60890	OPHN1	O43374	RASA4
Q96P48	ARAP1	Q05193	DNM1	P22059	OSBP	C9J798	RASA4B
Q8WZ64	ARAP2	P50570	DNM2	Q969R2	OSBP2	O95294	RASAL1
Q8WWN8	ARAP3	Q9UQ16	DNM3	Q9BXB5	OSBPL10	Q9UJF2	RASAL2
A1A4S6	ARHGAP10	Q96BY6	DOCK10	Q9BXB4	OSBPL11	Q13972	RASGRF1
Q8IWW6	ARHGAP12	A6NIW2	DOCK11	Q9BXW6	OSBPL1A	O14827	RASGRF2
		A0A0A0MS					
Q53QZ3	ARHGAP15	Y4	DOCK9	Q9H4L5	OSBPL3	Q13464	ROCK1
Q9P2F6	ARHGAP20	Q99704	DOK1	Q9H0X9	OSBPL5	O75116	ROCK2
Q5T5U3	ARHGAP21	O60496	DOK2	Q9BZF3	OSBPL6	Q9BST9	RTKN
Q7Z5H3	ARHGAP22	Q7L591	DOK3	Q9BZF2	OSBPL7	Q8IZC4	RTKN2
Q9P227	ARHGAP23	H3BQ19	DOK4	Q9BZF1	OSBPL8	O95248	SBF1
Q8N264	ARHGAP24	Q9P104	DOK5	B1AKJ6	OSBPL9	Q86WG5	SBF2
P42331	ARHGAP25	Q6PKX4	DOK6	Q8WV24	PHLDA1	B7Z5R3	SCAP2
Q9UNA1	ARHGAP26	Q18PE1	DOK7	Q53GA4	PHLDA2	Q9NRF2	SH2B
Q6ZUM4	ARHGAP27	Q54A15	DTGCU2	Q9Y5J5	PHLDA3	Q9NRF2	SH2B1
A6NI28	ARHGAP42	Q92556	ELMO1	Q86UU1	PHLDB1	O14492	SH2B2
J3KPQ4	ARHGAP9	Q96JJ3	ELMO2	Q86SQ0	PHLDB2	Q9UQQ2	SH2B3
M0QZR4	ARHGEF1	F8W9E7	ELMO3	Q6NSJ2	PHLDB3	P78314	SH3BP2
O15085	ARHGEF11	Q8IYI6	EXOC8	O60346	PHLPP1	Q86WV1	SKAP1
Q9NZN5	ARHGEF12	Q8N4B1	FAM109A	P51178	PLCD1	O75563	SKAP2
Q5VV41	ARHGEF16	Q6ICB4	FAM109B	Q8N3E9	PLCD3	Q13424	SNTA1

Q6ZSZ5	ARHGEF18	Q96TA1	FAM129B	C9JEA7	PLCD4	Q13884	SNTB1
Q8IW93	ARHGEF19	Q86XR2	FAM129C	Q4LE43	PLCG1	Q13425	SNTB2
V9GYM8	ARHGEF2	C9JME2	FARP1	P16885	PLCG2	Q9NSN8	SNTG1
Q86VW2	ARHGEF25	O94887	FARP2	Q4KWH8	PLCH1	Q07889	SOS1
Q96DR7	ARHGEF26	Q9BQL6	FERMT1	O75038	PLCH2	Q07890	SOS2
Q8N1W1	ARHGEF28	Q96AC1	FERMT2	Q15111	PLCL1	Q96N96	SPATA13
E9PG37	ARHGEF3	Q86UX7	FERMT3	Q9UPRO	PLCL2	P11277	SPTB
Q8N4T4	ARHGEF39	P98174	FGD1	Q13393	PLD1	P11277	SPTBN1
E7EV07	ARHGEF4	Q7Z6J4	FGD2	O14939	PLD2	O15020	SPTBN2
Q8TER5	ARHGEF40	Q5JSP0	FGD3	P08567	PLEK	Q9H254	SPTBN4
Q12774	ARHGEF5	F8VWL3	FGD4	Q9NYT0	PLEK2	Q9NRC6	SPTBN5
Q15052	ARHGEF6	Q6ZNL6	FGD5	Q9HB21	PLEKHA1	Q9ULZ2	STAP1
Q14155	ARHGEF7	Q6ZV73	FGD6	Q9HB19	PLEKHA2	Q9UH65	SWAP70
O43307	ARHGEF9	Q9NXY1	FLJ00004	Q9HB20	PLEKHA3	Q96PV0	SYNGAP1
Q9ULH1	ASAP1	Q6ZMK7	FLJ00312	Q9H4M7	PLEKHA4	Q9BYX2	TBC1D2
O43150	ASAP2	Q6ZMJ9	FLJ00357	Q9HAU0	PLEKHA5	B9A6J8	TBC1D2B
Q8TDY4	ASAP3	Q86YU9	FLJ00414	Q9Y2H5	PLEKHA6	P42680	TEC
Q86XR2	BCNP1	D3DS14	FLJ10357	E9PKC0	PLEKHA7	Q13009	TIAM1
P11274	BCR	Q13480	GAB1	J3KQS5	PLEKHA8	Q8IVF5	TIAM2
A2RQD7	BCR-ABL1	Q9UQC2	GAB2	Q9UF11	PLEKHB1	O75962	TRIO
A9UF07	BCR/ABL	Q8WWW8	GAB3	B7WPA5	PLEKHB2	Q9H2D6	TRIOBP
P51813	BMX	Q2WGN9	GAB4	Q96AC1	PLEKHC1	P15498	VAV1
Q9ULZ2	BRDG1	Q13322	GRB10	A6NEE1	PLEKHD1	P52735	VAV2

Q06187	BTK	Q14449	GRB14	Q96S99	PLEKHF1	Q9UKW4	VAV3
A0A024R8 72	C9orf88	Q14451	GRB7	Q9H8W4	PLEKHF2	Q14D04	VEPH1
Q9ULU8	CADPS	P25098	GRK2	Q9ULL1	PLEKHG1		
A0A087X1 P3	CADPS2	P35626	GRK3	Q9H7P9	PLEKHG2		
Q5VT25	CDC42BPA	D3DWE7	hCG_1994 053	A1L390	PLEKHG3		
Q9Y5S2	CDC42BPB	D3DU33	hCG_2002 091	Q58EX7	PLEKHG4		
Q6DT37	CDC42BPG	D3DSB1	hCG_2013 210	A0A1B0G W72	PLEKHG4 B		
C9J126	CDH2	A0A024RB A8	hCG_2015 932	O94827	PLEKHG5		
Q2V6Q1	CENTA2	D6W646	hCG_2225 3	Q3KR16	PLEKHG6		

Table 2.2 Parameters of Convolutional neural network.

Layer (type)	Output Shape	Parameter #
embedding_1 (Embedding)	(None, 330, 8)	184
conv1d_1 (Conv1D)	(None, 330, 32)	1568
max_pooling1d_1 (MaxPooling1D)	(None, 165, 32)	0
conv1d_2 (Conv1D)	(None, 165, 32)	3104
max_pooling1d_2 (MaxPooling1D)	(None, 82, 32)	0
flatten_1 (Flatten)	(None, 2624)	0
dense_1 (Dense)	(None, 128)	336000
dense_2 (Dense)	(None, 2)	258

Total parameters: 341,114

Table 2.3 Proteins predicted to bind to PIPs.

Q9NRF2	B7Z5R3	O60333	Q12756	Q54A15	Q7Z736
P35626	B9A6J8	O60346	Q12774	Q58EX7	Q86SQ0
P11277	B9EGI2	O60496	Q12802	Q5HYB0	Q86UU1
Q96AC1	C9J126	O60890	Q13009	Q5HYD7	Q86UX7
Q86XR2	C9J798	O75038	Q13322	Q5HYM3	Q86VW2
Q96TA1	C9JEA7	O75116	Q13393	Q5JS13	Q86WG5
DDEFL1	C9JME2	O75563	Q13424	Q5JSP0	Q86WV1
Q9HAU0	C9K0J5	O75689	Q13425	Q5JU85	Q86X27
Q99490	D3DS14	O75962	Q13464	Q5KSL6	Q86XP1
A0A024RBA8	D3DSB1	O94806	Q13480	Q5SXH7	Q86XR2
A0A024RBK8	D3DU33	O94827	Q13884	Q5T5U3	Q86YR7
Q9ULZ2	D3DWE7	O94887	Q13972	Q5VT25	Q86YU9
Q01082	D6W646	O95248	Q14155	Q5VTM2	Q8IVE3
Q6DN90	E7EUN2	P08567	Q14449	Q5VUJ5	Q8IVF5
Q96P64	E7EV07	P10911	Q14451	Q5VV41	Q8IW93
Q86UW7	E9PG37	P11274	Q14644	Q5VW22	Q8IWE5
Q96P47	E9PKC0	P15498	Q14D04	Q5VWQ8	Q8IWW6
Q9HD67	F8VWL3	P16885	Q15027	Q5W9G0	Q8IYI6
Q15283	F8W9E7	P20936	Q15052	Q5W9H1	Q8IZC4
Q9BZ29	F8WBA3	P22059	Q15057	Q658P8	Q8N1W1
O43307	G9CGD6	P25098	Q15111	Q69YP8	
P25098	H3BQ19	P31749	Q15438	Q6DT37	
Q9Y2H5	J3KPQ4	P31751	Q15811	Q6ICB4	
Q96PX9	J3KQS5	P35568	Q16760	Q6NSJ2	
Q12979	J3QSW6	P35626	Q18PE1	Q6PKX4	
Q6ZR37	K7EIZ3	P42331	Q2V6Q1	Q6ZMJ9	
A1A4S6	M0QZR4	P42680	Q2WGN9	Q6ZMK7	
A1L390	O14492	P50570	Q3KR16	Q6ZNL6	
A2RQD7	O14578	P51178	Q494U1	Q6ZSZ5	
A5PKW4	O14654	P51813	Q4KWH8	Q6ZUM4	
A6NEE1	O14827	P52735	Q4LE43	Q6ZV73	
A6NGQ3	O14939	P78314	Q4R9M9	Q6ZWE6	
A6NI28	O15020	P98174	Q4ZG22	Q70Z35	
A6NIR3	O15068	Q05193	Q53GA4	Q7L591	
A6NIW2	O15085	Q06187	Q53GL0	Q7Z5H3	
A9UF07	O43150	Q07889	Q53QZ3	Q7Z5R6	
B1AKJ6	O43374	Q07890	Q53R16	Q7Z628	
B7WPA5	O43739	Q08881	Q53RS3	Q7Z6J4	

Table 2.4 Annotations of mutations among all 313 PH domain proteins.

CDS	n_m ut	n_ clu st	n_ O G	n_mut _in_clu st	min_cl ustersi ze	corr_fa ctor_O G	n_T SG	n_TSG_ nonsen se	corr_fa ctor_TS G	n_si l	OG_ scor e	TSG_ sco re	OG_ p	log_ OG_ p	TSG_ _p	log_ TSG_ _p	DG_ _p	log_ DG_ p	OG_ _q	TSG_ _q	DG_ q	
AKT1	14 43	43	1	43	20	NA	0.85 583 4	0	0	0	0	0.45 945 9	0	3.08 E-44	100 .18 9	1	0	3.08 E-44	100 .18 9	8.22 E- 42	1	8.22 E- 42
RAS A1	31 44	89	0	56	0	NA	0.29 859 5	33	21	0.93 100 7	0	0	0.37 349 4	1	0	1.34 E-21	48. 062 7	1.34 E-21	48. 062 7	1	3.57 E- 19	3.57 E- 19
ROC K1	40 65	10 5	0	69	0	NA	0.33 514 2	36	13	0.86 029 2	0	0	0.34 444 4	1	0	2.20 E-20	45. 265 5	2.20 E-20	45. 265 5	1	2.93 E- 18	2.93 E- 18
ARA P3	46 35	10 1	0	66	0	NA	0	35	7	0.56 309 4	0	0	0.35 087 7	1	0	7.46 E-14	30. 226 3	7.46 E-14	30. 226 3	1	6.64 E- 12	6.64 E- 12
ROC K2	41 67	65	0	43	0	NA	0.24 957 8	22	7	0.77 985 1	0	0	0.33 333 3	1	0	1.42 E-11	24. 980 1	1.42 E-11	24. 980 1	1	9.46 E- 10	9.46 E- 10
DOC K11	62 22	11 6	0	87	0	NA	0.20 356 3	29	20	0.66 943 1	0	0	0.24 359 1	1	0	4.08 E-10	21. 619 8	4.08 E-10	21. 619 8	1	2.03 E- 08	2.03 E- 08
ITSN 2	50 94	81	0	59	0	NA	0.16 590 6	22	10	0.76 313 4	0	0	0.27 419 4	1	0	4.56 E-10	21. 508 3	4.56 E-10	21. 508 3	1	2.03 E- 08	2.03 E- 08
ARA P2	51 15	95	0	70	0	NA	0	25	14	0.66 551 9	0	0	0.26 984 1	1	0	6.00 E-10	21. 233 9	6.00 E-10	21. 233 9	1	2.29 E- 08	2.29 E- 08

MCF 2	27 78	90	4	68	2	NA	0.34 421 5	22	14	0.76 218 3	0	0.04 347 8	0.24 637 7	0.03 259 8	- 3.4 235	2.78 E-09	- 19. 701 5	9.06 E-11	- 23. 125	1	9.27 E- 08	9.27 E- 08
TRIO	92 94	14 5	0	110	0	NA	0	35	12	0.41 809 2	0	0	0.24 590 2	1	0	2.42 E-08	- 17. 537 8	2.42 E-08	17. 537 8	1	6.85 E- 07	6.85 E- 07
MCF 2L2	33 45	90	0	70	0	NA	0.31 682 9	20	15	0.77 686 6	0	0	0.22 857 1	1	0	2.57 E-08	- 17. 477 9	2.57 E-08	17. 477 9	1	6.85 E- 07	6.85 E- 07
PSD3	31 44	68	0	47	0	NA	0.06 773 8	21	6	0.56 716 4	0	0	0.30 769 2	1	0	3.96 E-08	- 17. 045 6	3.96 E-08	17. 045 6	1	9.60 E- 07	9.60 E- 07
DAB 2IP	31 98	42	0	24	0	NA	0	18	4	0.48 507 5	0	0	0.45	1	0	4.38 E-08	- 16. 943 9	4.38 E-08	16. 943 9	1	9.74 E- 07	9.74 E- 07
IRS4	37 74	11 5	0	85	0	NA	0	30	9	0.38 495	0	0	0.27 272 7	1	0	1.75 E-07	- 15. 556 2	1.75 E-07	15. 556 2	1	3.60 E- 06	3.60 E- 06
MYO 10	61 77	99	0	69	0	NA	0	30	15	0.37 853 8	0	0	0.29 729 7	1	0	2.16 E-07	- 15. 346 4	2.16 E-07	15. 346 4	1	4.13 E- 06	4.13 E- 06
PLEK HA5	33 51	58	0	42	0	NA	0.15 687 9	16	5	0.65 671 6	0	0	0.28 947 4	1	0	2.93 E-07	- 15. 043 5	2.93 E-07	15. 043 5	1	5.21 E- 06	5.21 E- 06
OSB PL11	22 44	38	0	26	0	NA	0.46 811 4	12	8	0.82 835 8	0	0	0.32 258 1	1	0	3.28 E-07	- 14. 931	3.28 E-07	14. 931	1	5.47 E- 06	5.47 E- 06

APPL 1	21 30	36	0	24	0	NA	0.30 140 3	12	4	0.72 537 3	0	0	0.34 615 4	1	0	6.48 E-07	- 14. 249 8	6.48 E-07	- 14. 249 8	1	1.02 E- 05	1.02 E- 05
ARH GEF6	23 31	66	0	51	0	NA	0.34 701 5	15	11	0.79 566 5	0	0	0.22 641 5	1	0	1.56 E-06	- 13. 370 1	1.56 E-06	- 13. 370 1	1	2.32 E- 05	2.32 E- 05
SYN GAP 1	40 32	61	0	40	0	NA	0	21	9	0.39 582 1	0	0	0.33 333 3	1	0	3.81 E-06	- 12. 477 5	3.81 E-06	- 12. 477 5	1	5.36 E- 05	5.36 E- 05
ARH GAP 12	25 41	52	0	39	0	NA	0.26 865 7	13	11	0.76 705 8	0	0	0.25	1	0	4.43 E-06	- 12. 327	4.43 E-06	- 12. 327	1	5.81 E- 05	5.81 E- 05
RAS A2	25 50	48	0	34	0	NA	0.15 451 6	14	4	0.66 992	0	0	0.28 125	1	0	4.63 E-06	- 12. 283 9	4.63 E-06	- 12. 283 9	1	5.81 E- 05	5.81 E- 05
GAB 1	21 75	40	0	25	0	NA	0.08 059 7	15	0	0.46 055 4	0	0	0.38 888 9	1	0	4.79 E-06	- 12. 249 8	4.79 E-06	- 12. 249 8	1	5.81 E- 05	5.81 E- 05
ARH GAP 21	58 74	98	0	78	0	NA	0.13 879 2	20	10	0.59 950 2	0	0	0.20 339	1	0	5.18 E-06	- 12. 169 8	5.18 E-06	- 12. 169 8	1	6.02 E- 05	6.02 E- 05
SKAP 1	10 80	23	0	13	0	NA	0.15 614 2	10	7	0.57 089 6	0	0	0.46 153 8	1	0	7.19 E-06	- 11. 843 3	7.19 E-06	- 11. 843 3	1	8.00 E- 05	8.00 E- 05
ITSN 1	51 66	10 2	0	71	0	NA	0	31	8	0.25 401 8	0	0	0.30 769 2	1	0	7.51 E-06	- 11. 799 3	7.51 E-06	- 11. 799 3	1	8.02 E- 05	8.02 E- 05

SNT G1	15 54	82	0	64	0	NA	0.18 625 2	18	7	0.53 242 4	0	0	0.22 727 3	1	0	1.11 406 E-05	11. 406 5	1.11 406 E-05	11. 406 5	1	0.00 011 4	0.00 011 4
CNK SR1	21 42	30	0	20	0	NA	0.01 169 8	10	4	0.71 393	0	0	0.33 333 3	1	0	1.56 065 E-05	11. 065 2	1.56 065 E-05	11. 065 2	1	0.00 015 5	0.00 015 5
PHLP P1	51 54	48	0	35	0	NA	0.12 348 4	13	8	0.62 888 3	0	0	0.26 666 7	1	0	2.42 628 E-05	10. 628 8	2.42 628 E-05	10. 628 8	1	0.00 023 1	0.00 023 1
ARH GAP 27	16 47	29	0	18	0	NA	0	11	4	0.56 309 4	0	0	0.37 5	1	0	3.01 411 E-05	10. 411 8	3.01 411 E-05	10. 411 8	1	0.00 027 7	0.00 027 7
OSB PL1A	28 53	48	0	36	0	NA	0	12	2	0.66 625 2	0	0	0.25	1	0	4.04 117 E-05	10. 117 6	4.04 117 E-05	10. 117 6	1	0.00 035 9	0.00 035 9
GRB 14	16 23	40	0	28	0	NA	0	12	3	0.61 094 5	0	0	0.29 166 7	1	0	4.17 085 E-05	10. 085 4	4.17 085 E-05	10. 085 4	1	0.00 035 9	0.00 035 9
SOS1	40 02	86	1	77	4	NA	0.55 803 7	9	5	0.79 975 1	0	0.04 651 2	0.10 144 9	5.60 E-05	- 9.7 902	0.02 5353	3.6 748 5	1.42 E-06	13. 465 1	0.00 747 6	0.08 461 7	0.00 063 3
PLEK HA3	90 3	17	0	10	0	NA	0	7	4	0.69 104 5	0	0	0.41 666 7	1	0	7.98 364 E-05	9.4 364 2	7.98 364 E-05	9.4 364 2	1	0.00 066 6	0.00 066 6
RAP H1	37 53	43	0	32	0	NA	0	11	3	0.64 111 3	0	0	0.25	1	0	0.00 0123	9.0 030 3	0.00 012 3	9.0 030 3	1	0.00 099 5	0.00 099 5

CNK SR2	31 05	77	0	64	0	NA	0.35 159 3	13	5	0.66 451 8	0	0	0.17 647 1	1	0	0.00 0244	- 8.3 173 4	0.00 024 4	- 8.3 173 4	1	0.00 191 8	0.00 191 8
PLEK HM3	22 86	24	0	16	0	NA	0	8	7	0.61 380 6	0	0	0.33 333 3	1	0	0.00 0272	- 8.2 093	0.00 027 2	- 8.2 093	1	0.00 207 6	0.00 207 6
VAV 3	25 44	71	0	59	0	NA	0.23 685 9	12	6	0.73 300 2	0	0	0.17 307 7	1	0	0.00 0284	- 8.1 0284	0.00 028 4	- 8.1 0284	1	0.00 211 211	0.00 211 211
PSD	30 75	57	0	45	0	NA	0.02 992 8	12	5	0.56 059 7	0	0	0.21 875	1	0	0.00 0302	- 8.1 054 5	0.00 030 2	- 8.1 054 5	1	0.00 217 9	0.00 217 9
FARP 2	31 65	47	0	32	0	NA	0	15	4	0.33 059 7	0	0	0.31 25	1	0	0.00 0382	- 7.8 697 6	0.00 038 2	- 7.8 697 6	1	0.00 268 5	0.00 268 5
PLEK HA6	31 47	61	0	50	0	NA	0.17 429	11	2	0.73 880 6	0	0	0.17 777 8	1	0	0.00 0509	- 7.5 828 9	0.00 050 9	- 7.5 828 9	1	0.00 348 5	0.00 348 5
TRIO BP	70 98	11 4	0	95	0	NA	0.00 647 7	19	10	0.49 899 2	0	0	0.15 789 5	1	0	0.00 0576	- 7.4 590 2	0.00 057 6	- 7.4 590 2	1	0.00 384 6	0.00 384 6
OSB PL6	28 17	62	0	50	0	NA	0.14 629 2	12	3	0.59 771 5	0	0	0.18 918 9	1	0	0.00 0769	- 7.1 707 8	0.00 076 9	- 7.1 707 8	1	0.00 500 6	0.00 500 6
ACA P2	23 37	44	0	35	0	NA	0.21 715 4	9	6	0.63 764 5	0	0	0.21 428 6	1	0	0.00 0917	- 6.9 938 7	0.00 091 7	- 6.9 938 7	1	0.00 583 3	0.00 583 3
PREX 2	48 21	13 0	0	109	0	NA	0.07 168 2	21	9	0.38 104 9	0	0	0.16	1	0	0.00 1053	- 6.8	0.00 105 3	- 6.8	1	0.00 653 9	0.00 653 9

																	559		559			
																	-		-			
PLEK HA7	33 66	45	0	34	0	NA	0	11	3	0.45 074 6	0	0	0.25	1	0	0.00 1179	6.7 431 6	0.00 117	6.7 431 6	1	0.00 698 6	0.00 698 6
SWA P70	17 58	23	0	17	0	NA				0.12 238 8			0.25	1	0	0.00 1179	6.7 431 6	0.00 117	6.7 431 6	1	0.00 698 6	0.00 698 6
CDC 42BP A	51 60	93	0	78	0	NA				0.07 099 6			0.25	1	0	0.00 1204	- 6.7 224	0.00 120 4	- 6.7 224	1	0.00 698 6	0.00 698 6
ARH GEF1 2	46 35	66	0	54	0	NA				0.14 189 1			0.2	1	0	0.00 1346	- 6.6 106	0.00 134	- 6.6 106	1	0.00 764 6	0.00 764 6
ARH GAP 10	23 61	45	0	34	0	NA				0.03 634 8			0.23 809 5	1	0	0.00 1494	- 6.5 061	0.00 149	- 6.5 061	1	0.00 831 1	0.00 831 1
RTK N2	18 30	40	0	32	0	NA				0.77 415 6			0.19 354 8	1	0	0.00 161	- 6.4 313	0.00 161	- 6.4 313	1	0.00 85	0.00 85
TEC	18 96	41	0	32	0	NA				0.33 033 6			0.19 354 8	1	0	0.00 161	- 6.4 313	0.00 161	- 6.4 313	1	0.00 85	0.00 85
OSB PL3	26 64	50	0	36	0	NA				0.26 769 194			0.30 769 2	1	0	0.00 1624	- 6.4 231	0.00 162	- 6.4 231	1	0.00 85	0.00 85
SBF1	56 04	71	0	56	0	NA				0.30 751 4			0.22 727 3	1	0	0.00 1867	- 6.2 832	0.00 186	- 6.2 832	1	0.00 958 8	0.00 958 8

PLEK HH2	44 82	95	0	81	0	NA	0	14	9	0.57 432 8	0	0	0.14 545 5	1	0	0.00 1985	- 6.2 221 8	0.00 198 5	- 6.2 221 8	1	0.00 999 9	0.00 999 9
ARH GEF2 5	17 43	22	0	16	0	NA	0	6	2	0.67 388 1	0	0	0.26 666 7	1	0	0.00 2898	- 5.8 436 8	0.00 289 8	- 5.8 436 8	1	0.01 433 433	0.01 433 433
DNM 2	26 13	37	0	27	0	NA	0	10	5	0.43 262 9	0	0	0.25	1	0	0.00 3737	- 5.5 895 2	0.00 373 7	- 5.5 895 2	1	0.01 786 9	0.01 786 9
AKA P13	84 54	12 0	0	105	0	NA	0.17 443 5	15	6	0.61 795 9	0	0	0.12 162 2	1	0	0.00 3748	- 5.5 865 7	0.00 374 8	- 5.5 865 7	1	0.01 786 9	0.01 786 9
SPTB N1	70 95	11 3	0	95	0	NA	0	18	14	0.25 019 6	0	0	0.17 857 1	1	0	0.00 5657	- 5.1 748 3	0.00 565 7	- 5.1 748 3	1	0.02 649 9	0.02 649 9
DGK D	36 45	69	0	53	0	NA	0	16	7	0.26 194	0	0	0.22 222 2	1	0	0.00 5876	- 5.1 368 1	0.00 587 6	- 5.1 368 1	1	0.02 659 3	0.02 659 3
PLEK HH3	23 82	28	0	22	0	NA	0.50 417 4	6	2	0.65 671 6	0	0	0.22 222 2	1	0	0.00 5876	- 5.1 368 1	0.00 587 6	- 5.1 368 1	1	0.02 659 3	0.02 659 3
VEP H1	25 02	73	0	62	0	NA	0	11	0	0.56 136	0	0	0.14 634 1	1	0	0.00 6886	- 4.9 782 4	0.00 688 6	- 4.9 782 4	1	0.03 064 3	0.03 064 3
PLCL 1	29 94	12 2	0	107	0	NA	0.10 211 3	15	13	0.55 659 2	0	0	0.11 764 7	1	0	0.00 747	- 4.8 968 747	0.00 4.8 968	- 4.8 968	1	0.03 269 9	0.03 269 9

ANLN	3375	51	0	40	0	NA	0.187396	11	5	0.399254	0	0	0.2	1	0	0.008704	-4393	0.008704	-4393	1	0.036313	0.036313
DNM3	2592	54	0	44	0	NA	0.262297	10	5	0.377799	0	0	0.2	1	0	0.008704	-4393	0.008704	-4393	1	0.036313	0.036313
FGD4	2301	31	0	24	0	NA	0	7	3	0.639552	0	0	0.2	1	0	0.008704	-4393	0.008704	-4393	1	0.036313	0.036313
FGD6	4293	71	0	62	0	NA	0.500966	9	6	0.668159	0	0	0.12766	1	0	0.013302	-4393	0.013302	-4393	1	0.054642	0.054642
ARHGEF19	2409	26	0	20	0	NA	0.139596	6	3	0.509595	0	0	0.230769	1	0	0.015331	-4393	0.015331	-4393	1	0.060198	0.060198
PLEKHG1	4158	69	0	51	0	NA	0	18	6	0.188602	0	0	0.230769	1	0	0.015331	-4393	0.015331	-4393	1	0.060198	0.060198
RTKN	1692	23	0	18	0	NA	0	5	1	0.570896	0	0	0.230769	1	0	0.015331	-4393	0.015331	-4393	1	0.060198	0.060198
CYTH3	1200	21	0	13	0	NA	0	8	2	0.244776	0	0	0.4	1	0	0.016126	-4393	0.016126	-4393	1	0.061588	0.061588
PHLDB2	3714	80	2	70	0	NA	0.295155	10	6	0.606408	0	0	0.122449	1	0	0.016147	-4393	0.016147	-4393	1	0.061588	0.061588

FGD 1	28 86	61	0	49	0	NA	0	12	0	0.22 761 2	0	0	0.21 428 6	1	0	0.01 8913	- 3.9 678 9	0.01 891 3	- 3.9 678 9	1	0.07 013 7	0.07 013 7
PLEK HA8	13 68	20	0	15	0	NA	0.02 487 6	5	1	0.68 700 6	0	0	0.21 428 6	1	0	0.01 8913	- 3.9 678 9	0.01 891 3	- 3.9 678 9	1	0.07 013 7	0.07 013 7
RAS AL2	38 10	72	0	64	0	NA	0.23 268 9	8	5	0.71 091 9	0	0	0.11 764 7	1	0	0.01 9385	- 3.9 432 7	0.01 938 5	- 3.9 432 7	1	0.07 09	0.07 09
ARH GAP 20	35 76	79	0	68	0	NA	0.03 026 3	11	4	0.48 507 5	0	0	0.13 157 9	1	0	0.02 0482	- 3.8 882 1	0.02 048 2	- 3.8 882 1	1	0.07 291 6	0.07 291 6
CDH 2	27 21	94	0	82	0	NA	0.23 172	12	10	0.40 497 5	0	0	0.13 157 9	1	0	0.02 0482	- 3.8 882 1	0.02 048 2	- 3.8 882 1	1	0.07 291 6	0.07 291 6
PLCG 2	37 98	82	0	68	0	NA	0	14	6	0.31 343 3	0	0	0.15 384 6	1	0	0.02 202	- 3.8 157 8	0.02 157 202	- 3.8 157 8	1	0.07 736 1	0.07 736 1
SPTB N5	11 02 5	84	0	74	0	NA	0.01 094 5	10	2	0.47 014 9	0	0	0.12 820 5	1	0	0.02 2705	- 3.7 851 6	0.02 270 5	- 3.7 851 6	1	0.07 844 7	0.07 844 7
OPH N1	24 09	51	0	41	0	NA	0	10	3	0.28 891 3	0	0	0.2	1	0	0.02 2917	- 3.7 758 8	0.02 291 7	- 3.7 758 8	1	0.07 844 7	0.07 844 7
PLEK HA2	12 78	17	0	10	0	NA	0	7	2	0.35 634 3	0	0	0.33 333 3	1	0	0.02 3521	- 3.7 498 4	0.02 352 1	- 3.7 498 4	1	0.07 949 6	0.07 949 6

AKT3	14 40	40	0	33	0	NA	0.13 854	6	5	0.72 537 3	0	0	0.13 793 1	1	0	0.03 1701	- 3.4 514	0.03 170 1	- 3.4 514	1	0.10 231 7	0.10 231 7
ARH GEF5	47 94	38	0	28	0	NA	0	10	6	0.17 993 4	0	0	0.28 571 4	1	0	0.03 2024	- 3.4 412 8	0.03 202 4	- 3.4 412 8	1	0.10 231 7	0.10 231 7
ARH GEF7	23 49	67	0	57	0	NA	0.01 410 1	10	4	0.25 213 2	0	0	0.17 647 1	1	0	0.03 219	- 3.4 361 1	0.03 219	- 3.4 361 1	1	0.10 231 7	0.10 231 7
CDC 42BP B	51 36	58	0	46	0	NA	0	12	5	0.28 768 7	0	0	0.17 647 1	1	0	0.03 219	- 3.4 361 1	0.03 219	- 3.4 361 1	1	0.10 231 7	0.10 231 7
APPL 2	19 95	43	0	37	0	NA	0.39 054 7	6	5	0.71 393	0	0	0.12 903 2	1	0	0.03 9339	- 3.2 355 4	0.03 933 9	- 3.2 355 4	1	0.12 357 1	0.12 357 1
DGK K	33 54	50	0	37	0	NA	0	13	2	0.16 324 6	0	0	0.25	1	0	0.04 1527	- 3.1 814 1	0.04 152 7	- 3.1 814 1	1	0.12 892 7	0.12 892 7
ARH GEF1	27 84	51	0	41	0	NA	0.15 164 2	9	1	0.37 779 9	0	0	0.15 789 5	1	0	0.04 3125	- 3.1 436 5	0.04 312 5	- 3.1 436 5	1	0.13 084 5	0.13 084 5
ASA P2	30 21	46	0	38	0	NA	0	8	2	0.40 705 6	0	0	0.15 789 5	1	0	0.04 3125	- 3.1 436 5	0.04 312 5	- 3.1 436 5	1	0.13 084 5	0.13 084 5
ABR	25 80	47	0	35	0	NA	0	12	2	0.18 140 1	0	0	0.22 222 2	1	0	0.05 1933	- 2.9 577 9	0.05 193 3	- 2.9 577 9	1	0.15 58	0.15 58

CAD PS2	38 91	73	0	61	0	NA	0	12	7	0.28 768 7	0	0	0.14 285 7	1	0	0.05 5665	- 2.8 883 9	0.05 566	- 2.8 883 9	1	0.16 514 1	0.16 514 1
CAD PS	39 45	10 8	0	98	0	NA	0.06 193 5	10	7	0.46 112 5	0	0	0.1	1	0	0.05 7568	- 2.8 547 9	0.05 756 8	- 2.8 547 9	1	0.16 707 3	0.16 707 3
KALR N	89 61	17 7	0	158	0	NA	0	19	14	0.28 109 5	0	0	0.1	1	0	0.05 7568	- 2.8 547 9	0.05 756 8	- 2.8 547 9	1	0.16 707 3	0.16 707 3
PRK D1	27 39	87	0	78	0	NA	0.19 509	9	6	0.40 878 9	0	0	0.11 111 1	1	0	0.06 2605	- 2.7 709 1	0.06 260 5	- 2.7 709 1	1	0.17 563 6	0.17 563 6
ARH GAP 24	22 47	28	0	22	0	NA	0	6	5	0.36 247 3	0	0	0.2	1	0	0.06 315	- 2.7 622 4	0.06 315	- 2.7 622 4	1	0.17 563 6	0.17 563 6
ARH GAP 9	22 53	38	0	32	0	NA	0.16 417 9	6	6	0.25 621 9	0	0	0.2	1	0	0.06 315	- 2.7 622 4	0.06 315	- 2.7 622 4	1	0.17 563 6	0.17 563 6
SH2 B3	17 28	12	0	9	0	NA	0.38 413 2	3	0	0.81 275 4	0	0	0.2	1	0	0.06 315	- 2.7 622 4	0.06 315	- 2.7 622 4	1	0.17 563 6	0.17 563 6
ARH GEF9	15 45	39	0	34	0	NA	0.34 130 7	5	2	0.58 315 6	0	0	0.13 043 5	1	0	0.06 973	- 2.6 631 3	0.06 973	- 2.6 631 3	1	0.18 997 9	0.18 997 9
SH2 B1	20 52	37	0	32	0	NA	0	5	4	0.62 349 5	0	0	0.13 043 5	1	0	0.06 973	- 2.6 631 3	0.06 973	- 2.6 631 3	1	0.18 997 9	0.18 997 9

PLEK HG6	23 73	23	0	19	0	NA	0	4	2	0.48 507 5	0	0	0.18 181 8	1	0	0.07 5091	- 2.5 890 6	0.07 509 1	- 2.5 890 6	1	0.20 251 8	0.20 251 8
ARH GAP 15	14 28	41	0	36	0	NA	0.11 924 2	5	3	0.59 950 2	0	0	0.12	1	0	0.08 5218	- 2.4 625 4	0.08 521 8	- 2.4 625 4	1	0.22 753 3	0.22 753 3
ADR BK2	20 67	41	0	35	0	NA	0	6	2	0.28 482 6	0	0	0.16 666 7	1	0	0.08 7675	- 2.4 341 1	0.08 767 5	- 2.4 341 1	1	0.22 950 3	0.22 950 3
DAP P1	84 3	16	0	13	0	NA	0	3	0	0.75 970 1	0	0	0.16 666 7	1	0	0.08 7675	- 2.4 341 1	0.08 767 5	- 2.4 341 1	1	0.22 950 3	0.22 950 3
SBF2	55 50	86	0	77	0	NA	0.35 176 4	9	7	0.48 507 5	0	0	0.09 523 8	1	0	0.09 8095	- 2.3 218 1	0.09 809 5	- 2.3 218 1	1	0.25 428 6	0.25 428 6
SKAP 2	10 80	17	0	14	0	NA	0.68 003 7	3	3	0.75 970 1	0	0	0.15 384 6	1	0	0.10 0829	- 2.2 943 3	0.10 082 9	- 2.2 943 3	1	0.25 885 8	0.25 885 8
SOS2	39 99	62	1	56	1	NA	0.32 089 6	6	6	0.69 222 9	0	0	0.09 302 3	1	0	0.10 4759	- 2.2 561	0.10 475 9	- 2.2 561	1	0.26 638 6	0.26 638 6
FARP 1	31 38	45	0	38	0	NA	0	7	6	0.31 343 3	0	0	0.14 285 7	1	0	0.11 4481	- 2.1 673 4	0.11 448 1	- 2.1 673 4	1	0.28 566 8	0.28 566 8
IRS1	37 29	69	0	61	0	NA	0	8	2	0.20 420 6	0	0	0.14 285 7	1	0	0.11 4481	- 2.1 673 4	0.11 448 1	- 2.1 673 4	1	0.28 566 8	0.28 566 8

PLCD 4	22 89	29	0	25	0	NA	0	4	2	0.53 411 5	0	0	0.13 333 3	1	0	0.12 8568	- 2.0 513	0.12 856 8	- 2.0 513	1	0.31 207	0.31 207
PLEK HF2	75 0	21	0	18	0	NA	0.39 771 7	3	1	0.69 962 7	0	0	0.13 333 3	1	0	0.12 8568	- 2.0 513	0.12 856 8	- 2.0 513	1	0.31 207	0.31 207
PSD2	23 16	51	0	45	0	NA	0.05 879 3	6	2	0.29 436 2	0	0	0.13 333 3	1	0	0.12 8568	- 2.0 513	0.12 856 8	- 2.0 513	1	0.31 207	0.31 207
ACA P1	22 23	30	0	27	0	NA	0	3	3	0.53 656 7	0	0	0.12 5	1	0	0.14 3029	- 1.9 447 1	0.14 302 9	- 1.9 447 1	1	0.34 097 1	0.34 097 1
PLCL 2	30 06	57	0	50	0	NA	0	7	5	0.28 073 9	0	0	0.12 5	1	0	0.14 3029	- 1.9 447 1	0.14 302 9	- 1.9 447 1	1	0.34 097 1	0.34 097 1
AGA P1	25 74	43	0	37	0	NA	0	6	1	0.10 364 8	0	0	0.25	1	0	0.15 7338	- 1.8 493 6	0.15 733 8	- 1.8 493 6	1	0.36 012 5	0.36 012 5
FLJ1 0357	25 02	22	0	18	0	NA	0	4	0	0.19 900 5	0	0	0.25	1	0	0.15 7338	- 1.8 493 6	0.15 733 8	- 1.8 493 6	1	0.36 012 5	0.36 012 5
IPCE F1	13 17	13	0	10	0	NA	0	3	0	0.27 910 4	0	0	0.25	1	0	0.15 7338	- 1.8 493 6	0.15 733 8	- 1.8 493 6	1	0.36 012 5	0.36 012 5
PLD1	32 25	64	0	53	0	NA	0	11	7	0.06 377 2	0	0	0.25	1	0	0.15 7338	- 1.8 493 6	0.15 733 8	- 1.8 493 6	1	0.36 012 5	0.36 012 5
APB B1P	20 01	56	0	50	0	NA	0	6	3	0.30 764 023	0	0	0.11 764 7	1	0	0.15 7807	- 1.8 463 8	0.15 780 7	- 1.8 463 8	1	0.36 012 5	0.36 012 5

PLEK HH1	40 95	45	0	39	0	NA	0.13 497	6	3	0.40 998 1	0	0	0.11 111 1	1	0	0.17 2852	- 1.7 553 2	0.17 285 2	- 1.7 553 2	1	0.38 782 7	0.38 782 7
RAL GPS2	17 52	32	0	28	0	NA	0.04 011 2	4	2	0.54 944	0	0	0.11 111 1	1	0	0.17 2852	- 1.7 553 2	0.17 285 2	- 1.7 553 2	1	0.38 782 7	0.38 782 7
FAM 109B	78 0	9	0	8	0	NA	0	1	0	0.52 798 5	0	0	0.2	1	0	0.19 264	- 1.6 469 3	0.19 264	- 1.6 469 3	1	0.41 148	0.41 148
KIF1 A	50 73	80	0	71	0	NA	0	9	4	0.06 005 7	0	0	0.2	1	0	0.19 264	- 1.6 469 3	0.19 264	- 1.6 469 3	1	0.41 148	0.41 148
PHL DA2	45 9	7	0	6	0	NA	0.26 865 7	1	0	0.74 253 7	0	0	0.2	1	0	0.19 264	- 1.6 469 3	0.19 264	- 1.6 469 3	1	0.41 148	0.41 148
PLEK HM1	31 71	44	0	39	0	NA	0	5	1	0.11 318 4	0	0	0.2	1	0	0.19 264	- 1.6 469 3	0.19 264	- 1.6 469 3	1	0.41 148	0.41 148
PLEK HM2	33 69	29	0	25	0	NA	0	4	1	0.18 470 1	0	0	0.2	1	0	0.19 264	- 1.6 469 3	0.19 264	- 1.6 469 3	1	0.41 148	0.41 148
RAS A3	25 05	36	0	32	0	NA	0	4	0	0.14 179 1	0	0	0.2	1	0	0.19 264	- 1.6 469 3	0.19 264	- 1.6 469 3	1	0.41 148	0.41 148
NGE F	21 33	52	0	47	0	NA	0.07 091 2	5	5	0.39 516 7	0	0	0.09 523 8	1	0	0.21 9115	- 1.5 181 6	0.21 911 5	- 1.5 181 6	1	0.46 431 6	0.46 431 6

PLEK HG4	35 76	31	0	25	0	NA	0.08 927 1	6	2	0.19 083 2	0	0	0.16 666 7	1	0	0.22 6464	- 851 7	0.22 646 4	- 851 7	1	0.47 611	0.47 611
CIT	61 68	95	0	85	0	NA	0	10	7	0.23 396 9	0	0	0.09 090 9	1	0	0.23 4776	- 1.4 491 2	0.23 477 6	- 1.4 491 2	1	0.48 972 7	0.48 972 7
TIA M1	47 76	11 4	0	102	0	NA	0	12	9	0.20 358 2	0	0	0.08 695 7	1	0	0.25 0494	- 1.3 843 2	0.25 049 4	- 1.3 843 2	1	0.51 198 9	0.51 198 9
ARH GEF1 1	46 89	69	0	56	0	NA	0	13	5	0.10 746 3	0	0	0.14 285 7	1	0	0.25 8871	- 1.3 514 3	0.25 887 1	- 1.3 514 3	1	0.51 198 9	0.51 198 9
ARH GEF2	28 77	41	0	35	0	NA	0	6	3	0.16 631 1	0	0	0.14 285 7	1	0	0.25 8871	- 1.3 514 3	0.25 887 1	- 1.3 514 3	1	0.51 198 9	0.51 198 9
DNM 1	25 95	30	0	27	0	NA	0	3	1	0.24 477 6	0	0	0.14 285 7	1	0	0.25 8871	- 1.3 514 3	0.25 887 1	- 1.3 514 3	1	0.51 198 9	0.51 198 9
DOC K10	65 55	11 7	0	102	0	NA	0.02 240 7	15	8	0.05 597	0	0	0.14 285 7	1	0	0.25 8871	- 1.3 514 3	0.25 887 1	- 1.3 514 3	1	0.51 198 9	0.51 198 9
OSB PL7	25 29	35	0	28	0	NA	0	7	0	0.19 900 5	0	0	0.14 285 7	1	0	0.25 8871	- 1.3 514 3	0.25 887 1	- 1.3 514 3	1	0.51 198 9	0.51 198 9
PLEK HA1	12 15	20	0	16	0	NA	0	4	4	0.37 064 7	0	0	0.14 285 7	1	0	0.25 8871	- 1.3 514 3	0.25 887 1	- 1.3 514 3	1	0.51 198 9	0.51 198 9

GRB 10	17 85	39	0	34	0	NA	0	4	3	0.20 615 7	0	0	0.12 5	1	0	0.28 992	- 1.2 381 5	0.28 992	- 1.2 381 5	1	0.56 502 7	0.56 502 7
IQSE C2	40 11	47	0	40	0	NA	0	7	1	0.16 889 2	0	0	0.12 5	1	0	0.28 992	- 1.2 381 5	0.28 992	- 1.2 381 5	1	0.56 502 7	0.56 502 7
FER MT2	20 43	25	0	23	0	NA	0.30 348 3	2	0	0.35 634 3	0	0	0.11 111 1	1	0	0.31 9668	- 1.1 404 7	0.31 966 8	- 1.1 404 7	1	0.60 106 7	0.60 106 7
FGD 3	21 78	46	0	40	0	NA	0	6	1	0.20 309 2	0	0	0.11 111 1	1	0	0.31 9668	- 1.1 404 7	0.31 966 8	- 1.1 404 7	1	0.60 106 7	0.60 106 7
ITK	18 63	53	0	46	0	NA	0	7	4	0.17 611 9	0	0	0.11 111 1	1	0	0.31 9668	- 1.1 404 7	0.31 966 8	- 1.1 404 7	1	0.60 106 7	0.60 106 7
PLEK	10 53	32	0	28	0	NA	0.14 676 6	4	3	0.27 910 4	0	0	0.11 111 1	1	0	0.31 9668	- 1.1 404 7	0.31 966 8	- 1.1 404 7	1	0.60 106 7	0.60 106 7
STAP 1	88 8	21	0	17	0	NA	0	3	3	0.42 786 1	0	0	0.11 111 1	1	0	0.31 9668	- 1.1 404 7	0.31 966 8	- 1.1 404 7	1	0.60 106 7	0.60 106 7
DOK 1	14 46	17	0	16	0	NA	0.24 990 4	1	0	0.57 089 6	0	0	0.1	1	0	0.34 8171	- 1.0 550 6	0.34 817 1	- 1.0 550 6	1	0.63 239 1	0.63 239 1
DOK 5	92 1	27	0	23	0	NA	0.30 348 3	4	1	0.35 634 3	0	0	0.1	1	0	0.34 8171	- 1.0 550 6	0.34 817 1	- 1.0 550 6	1	0.63 239 1	0.63 239 1

MPR IP	30 78	39	0	34	0	NA	0	5	2	0.25 621 9	0	0	0.1	1	0	0.34 8171	- 1.0 550 6	0.34 817 1	- 1.0 550 6	1	0.63 239 1	0.63 239 1
OSB PL10	22 95	31	0	27	0	NA	0	4	4	0.31 343 3	0	0	0.1	1	0	0.34 8171	- 1.0 550 6	0.34 817 1	- 1.0 550 6	1	0.63 239 1	0.63 239 1
PLCH 1	49 68	10 4	0	93	0	NA	0	11	5	0.09 373 1	0	0	0.1	1	0	0.34 8171	- 1.0 550 6	0.34 817 1	- 1.0 550 6	1	0.63 239 1	0.63 239 1
AFA P1L1	23 07	37	0	35	0	NA	0	2	1	0.28 891 3	0	0	0.09 090 9	1	0	0.37 5478	- 0.9 795 5	0.37 547 8	- 0.9 795 5	1	0.67 283 7	0.67 283 7
ELM O1	21 84	67	0	58	0	NA	0	8	4	0.17 039 8	0	0	0.09 090 9	1	0	0.37 5478	- 0.9 795 5	0.37 547 8	- 0.9 795 5	1	0.67 283 7	0.67 283 7
DOC K9	62 07	68	0	61	0	NA	0	7	4	0.17 109 6	0	0	0.08 333 3	1	0	0.40 1642	- 0.9 121 9	0.40 164 2	- 0.9 121 9	1	0.71 018 9	0.71 018 9
OSB PL9	22 41	26	0	23	0	NA	0	3	2	0.45 238 1	0	0	0.08 333 3	1	0	0.40 1642	- 0.9 121 9	0.40 164 2	- 0.9 121 9	1	0.71 018 9	0.71 018 9
RAS GRF 2	37 14	90	0	82	0	NA	0.10 751 3	8	2	0.15 739 5	0	0	0.07 142 9	1	0	0.45 0728	- 0.7 968 9	0.45 072 8	- 0.7 968 9	1	0.79 173 9	0.79 173 9
BMX	20 28	45	0	43	0	NA	0.01 221 2	2	1	0.36 247 3	0	0	0.06 25	1	0	0.49 5787	- 0.7 016 1	0.49 578 7	- 0.7 016 1	1	0.86 519 6	0.86 519 6

ACA P3	22 80	20	0	17	0	NA	0	3	1	0	0	0	0	1	0	1	0	1	0	1	1	1
ADA P1	11 25	6	0	5	0	NA	0	1	0	0.01 306	0	0	0	1	0	1	0	1	0	1	1	1
ADA P2	11 46	15	0	13	0	NA	0	2	2	0.14 179	0	0	0	1	0	1	0	1	0	1	1	1
ADR BK1	20 70	26	0	22	0	NA	0	4	0	0	0	0	0	1	0	1	0	1	0	1	1	1
AFA P1	21 93	23	0	21	0	NA	0	2	1	0	0	0	0	1	0	1	0	1	0	1	1	1
AFA P1L2	24 57	26	0	24	0	NA	0	2	1	0.18 470	0	0	0	1	0	1	0	1	0	1	1	1
AGA P11	25 02	22	0	22	0	NA	0	0	0	0.19 900	0	0	0	1	0	1	0	1	0	1	1	1
AGA P2	25 11	45	0	42	0	NA	0	3	1	0	0	0	0	1	0	1	0	1	0	1	1	1
AGA P3	27 36	44	0	38	0	NA	0	6	1	0	0	0	0	1	0	1	0	1	0	1	1	1
AGA P4	19 92	11	0	10	0	NA	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
AGA P5	19 92	14	0	14	0	NA	0.05 487	0	0	0.09 888	0	0	0	1	0	1	0	1	0	1	1	1
AGA P6	20 61	31	0	26	0	NA	0	5	0	0.01 306	0	0	0	1	0	1	0	1	0	1	1	1
AGA P9	19 77	1	0	1	0	NA	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
AKT2	14 46	22	0	20	0	NA	0	2	1	0	0	0	0	1	0	1	0	1	0	1	1	1
ARA P1	43 53	60	0	48	0	NA	0	12	1	0.03 689	0	0	0	1	0	1	0	1	0	1	1	1

ARH GAP 22	20 97	37	0	33	0	NA	0	4	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GAP 23	54 51	14	0	12	0	NA	0	2	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GAP 25	19 20	61	0	51	0	NA	0	10	3	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GAP 26	24 45	40	0	24	0	NA	0.09 076 2	16	2	0.01 306	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GAP 42	26 25	24	0	22	0	NA	0.18 473 2	2	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GEF1 6	12 66	19	0	19	0	NA	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GEF1 8	30 48	40	0	37	0	NA	0	3	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GEF2 6	26 16	35	0	32	0	NA	0.06 919 9	3	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GEF3 3	15 81	32	0	24	0	NA	0	8	5	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GEF3 9	10 08	1	1	1	0	NA	0.60 995	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GEF4 7	20 73	32	0	30	0	NA	0	2	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ARH GEF4 0	45 60	21	0	15	0	NA	0	6	2	0.03 451 5	0	0	0	0	1	0	1	0	1	0	1	1	1
ASA P1	33 90	69	0	61	0	NA	0	8	7	0	0	0	0	0	1	0	1	0	1	0	1	1	1

ASA P3	27 12	25	0	25	0	NA	0	0	0	0.33 795 3	0	0	0	1	0	1	0	1	0	1	1	1
BCR	38 16	23	0	23	0	NA	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
BTK	19 80	60	0	53	0	NA	0.16 004 1	7	2	0	0	0	0	1	0	1	0	1	0	1	1	1
CDC 42BP G	46 56	38	0	34	0	NA	0	4	3	0	0	0	0	1	0	1	0	1	0	1	1	1
COL4 A3B P	18 75	17	0	16	0	NA	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
CYT H1	11 97	19	0	17	0	NA	0	2	2	0	0	0	0	1	0	1	0	1	0	1	1	1
CYT H2	12 00	23	0	22	0	NA	0.38 413 2	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
CYT H4	11 85	21	0	15	0	NA	0	6	5	0	0	0	0	1	0	1	0	1	0	1	1	1
DEF6	18 96	16	0	15	0	NA	0.16 417 9	1	0	0.19 900 5	0	0	0	1	0	1	0	1	0	1	1	1
DGK H	36 63	52	0	50	0	NA	0	2	2	0	0	0	0	1	0	1	0	1	0	1	1	1
DOK 2	12 39	23	0	21	0	NA	0	2	1	0.18 470 1	0	0	0	1	0	1	0	1	0	1	1	1
DOK 3	14 91	23	0	22	0	NA	0.07 619 8	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
DOK 4	98 1	12	0	11	0	NA	0	1	0	0.39 925 4	0	0	0	1	0	1	0	1	0	1	1	1

DOK 6	99 6	32	0	31	0	NA	0.35 882 2	1	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
DOK 7	15 15	16	0	12	0	NA	0	4	0	0.05 597	0	0	0	0	1	0	1	0	1	0	1	1	1
ELM O2	21 63	33	0	26	0	NA	0	7	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
ELM O3	23 22	21	0	20	0	NA	0	1	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
EXO C8	21 78	16	0	16	0	NA	0.42 639 7	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
FAM 109A	78 9	5	0	5	0	NA	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
FAM 129B	22 41	29	0	28	0	NA	0	1	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
FAM 129C	20 01	23	0	23	0	NA	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
FER MT1	20 34	22	0	16	0	NA	0	6	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
FER MT3	19 98	27	0	25	0	NA	0	2	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
FGD 2	19 68	36	0	32	0	NA	0	4	1	0.01 919	0	0	0	0	1	0	1	0	1	0	1	1	1
FGD 5	43 89	75	0	69	0	NA	0	6	2	0	0	0	0	0	1	0	1	0	1	0	1	1	1
GAB 2	20 31	37	0	36	0	NA	0	1	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
GAB 3	17 61	47	0	38	0	NA	0	9	2	0	0	0	0	0	1	0	1	0	1	0	1	1	1
GAB 4	17 25	38	0	30	0	NA	0	8	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
GRB 7	15 99	26	0	25	0	NA	0	1	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
IQSE C1	28 92	35	0	30	0	NA	0	5	2	0	0	0	0	0	1	0	1	0	1	0	1	1	1

IRS2	40 17	17	0	12	0	NA	0	5	1	0	0	0	0	1	0	1	0	1	0	1	1	1
KIF1 B	53 13	97	0	90	0	NA	0	7	4	0	0	0	0	1	0	1	0	1	0	1	1	1
MCF 2L	33 90	43	0	38	0	NA	0	5	2	0	0	0	0	1	0	1	0	1	0	1	1	1
NET 1	17 91	23	0	22	0	NA	0	1	0	0.41 641 8	0	0	0	1	0	1	0	1	0	1	1	1
OBS CN	25 65 3	35 4	0	305	0	NA	0	49	25	0.00 201 5	0	0	0	1	0	1	0	1	0	1	1	1
OSB P	24 24	26	0	25	0	NA	0.16 982 7	1	1	0	0	0	0	1	0	1	0	1	0	1	1	1
OSB P2	27 51	32	0	28	0	NA	0	4	2	0	0	0	0	1	0	1	0	1	0	1	1	1
OSB PL5	26 40	29	0	26	0	NA	0	3	1	0	0	0	0	1	0	1	0	1	0	1	1	1
OSB PL8	26 70	49	0	46	0	NA	0.30 888 19	3	2	0.09 470 1	0	0	0	1	0	1	0	1	0	1	1	1
PHL DA1	12 06	28	0	27	0	NA	0	1	0	0.18 470 1	0	0	0	1	0	1	0	1	0	1	1	1
PHL DA3	38 4	7	0	7	0	NA	0.02 487 6	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PHL DB1	41 34	39	0	38	0	NA	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PHL DB3	19 23	20	0	19	0	NA	0.04 925 4	1	1	0.44 216 4	0	0	0	1	0	1	0	1	0	1	1	1
PLCD 1	22 71	20	0	19	0	NA	0	1	0	0.16 086 2	0	0	0	1	0	1	0	1	0	1	1	1

PLCD 3	23 67	17	0	15	0	NA	0	2	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLCG 1	38 76	63	0	59	0	NA	0.03 120 8	4	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLCH 2	42 51	49	0	48	0	NA	0	1	1	0.04 371	0	0	0	0	1	0	1	0	1	0	1	1	1
PLD2	28 02	39	0	31	0	NA	0	8	2	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK 2	10 62	10	0	8	0	NA	0	2	1	0.22 761 2	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HA4	23 40	34	0	31	0	NA	0	3	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HB1	73 2	11	0	10	0	NA	0	1	0	0.39 925 4	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HB2	66 6	20	0	17	0	NA	0	3	3	0.10 746 3	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HF1	84 0	8	0	7	0	NA	0	1	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HG2	40 35	64	0	56	0	NA	0	8	5	0.01 919	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HG3	34 92	38	0	36	0	NA	0	2	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HG4 B	38 16	65	0	56	0	NA	0	9	6	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HG5	32 52	30	0	28	0	NA	0	2	1	0.22 761 2	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HG7	11 40	20	0	18	0	NA	0	2	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HJ1	45 0	4	0	4	0	NA	1	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1

PLEK HN1	18 36	23	0	21	0	NA	0	2	1	0.19 900 5	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HO1	12 30	18	0	15	0	NA	0	3	2	0.01 306	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HO2	14 73	26	0	25	0	NA	0.12 238 8	1	1	0.14 179 1	0	0	0	1	0	1	0	1	0	1	1	1
PLEK HS1	11 01	1	0	1	0	NA	0.12 018 9	0	0	0.31 343 3	0	0	0	1	0	1	0	1	0	1	1	1
PREX 1	49 80	10 8	0	101	0	NA	0	7	2	0	0	0	0	1	0	1	0	1	0	1	1	1
PRK D2	26 37	45	0	40	0	NA	0	5	1	0	0	0	0	1	0	1	0	1	0	1	1	1
PRK D3	26 73	59	0	54	0	NA	0.05 123	5	2	0	0	0	0	1	0	1	0	1	0	1	1	1
PSD4	31 71	41	0	39	0	NA	0	2	0	0	0	0	0	1	0	1	0	1	0	1	1	1
RAL GPS1	15 90	31	0	28	0	NA	0.05 123	3	1	0	0	0	0	1	0	1	0	1	0	1	1	1
RAS A4	24 12	2	0	2	0	NA	0.21 24	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
RAS AL1	24 15	66	0	58	0	NA	0	8	6	0.01 919	0	0	0	1	0	1	0	1	0	1	1	1
RAS GRF 1	38 22	63	0	59	0	NA	0	4	2	0	0	0	0	1	0	1	0	1	0	1	1	1
SH2 B2	20 16	8	0	8	0	NA	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
SH3 BP2	16 86	22	0	17	0	NA	0	5	0	0	0	0	0	1	0	1	0	1	0	1	1	1
SNT A1	15 18	14	0	13	0	NA	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1	1
SNT B1	16 17	20	0	19	0	NA	0	1	1	0	0	0	0	1	0	1	0	1	0	1	1	1

SNT B2	16 23	13	0	8	0	NA	0	5	1	0	0	0	0	1	0	1	0	1	0	1	1	1
SPAT A13	19 59	27	0	26	0	NA	0	1	1	0	0	0	0	1	0	1	0	1	0	1	1	1
SPTB	69 87	10 9	0	100	0	NA	0	9	6	0	0	0	0	1	0	1	0	1	0	1	1	1
SPTB N2	71 73	98	0	90	0	NA	0	8	5	0	0	0	0	1	0	1	0	1	0	1	1	1
SPTB N4	76 95	76	0	68	0	NA	0	8	0	0	0	0	0	1	0	1	0	1	0	1	1	1
TBC1 D2	27 54	35	0	33	0	NA	0	2	2	0	0	0	0	1	0	1	0	1	0	1	1	1
TBC1 D2B	27 45	25	0	25	0	NA	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	1
TIA M2	51 06	79	0	73	0	NA	0	6	3	0	0	0	0	1	0	1	0	1	0	1	1	1
VAV 1	25 38	55	0	50	0	NA	0	5	4	0	0	0	0	1	0	1	0	1	0	1	1	1
VAV 2	25 20	47	0	45	0	NA	0	2	1	0	0	0	0	1	0	1	0	1	0	1	1	1

2.4 Discussion

Although PH domain proteins have been studied for more than 20 years, lots of information about these proteins are still illusive. In this chapter, our genomics analysis identified a list of most frequently mutated PHGs and most up-regulated PHGs. Among these genes, AKT1 was the one with most attention and best studied in the past decades. Other than AKT1, we identified a list of interesting genes such as CNKSR2, DOCK, KIF1A, and CADPS. Of these genes, Tiam1 was one of the most interesting genes. Analysis of 13 types of most common cancer dataset revealed that Tiam1 was one of the top10 most frequently mutated genes. Interestingly, when it considered PH domain only, Tiam1 was also one of the most frequently mutated genes. However, it did not show significant difference of mutation frequency among these 13 cancer types. The expression level of Tiam1 only significantly increased in neuroendocrine prostate cancer patients, but not any other types of cancers. This makes Tiam1 a very interesting drug target in this type of cancer. Through structural multiple sequence alignment, several conserved residues, such as arginine and tryptophan, involved in PIPs binding and recognition were identified. These conserved residues provide precious information to identify the binding pocket of PIPs especially in proteins without PH-PIPs complex crystal structures.

Then we collected all the PH domain proteins with crystal structures and tried to identify the residues which were responsible for the PIPs binding specificity. However, no clear cluster was obtained due to the limit of the crystal structure numbers. So we collected all PH-PIPs binding information from PubMed and saved it as a database. Using these information, we built a CNN based machine learning model to predict their PIPs binding ability. The visualization of the protein clusters and PIPs binding specificities were displayed in the webserver. To our

knowledge, for the first time, PH-PIPs binding data were displayed and published in a webserver.
This may bring significant convenience to the community to utilize the data.

Chapter 3: In silico discovery of small molecule inhibitors targeting Tiam1

3.1 Introduction

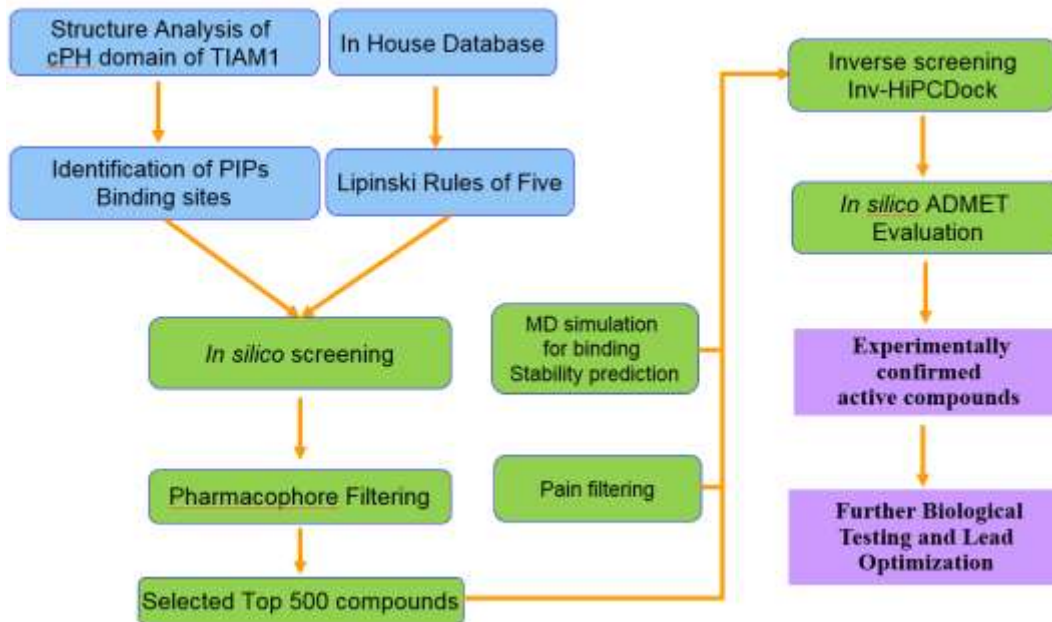
As the most frequently mutated gene in all types of human cancers [41], Ras has received unprecedented attention of developing inhibitors in the past three decades[42]. Despite numerous efforts and strategies, such as disruption of localization of the protein and synthetic lethality, have been intensively attempted, it is still not even close to effectively inhibit aberrant Ras pathways [43]. Recently, direct targeting Ras has been intensively studied and some groups have identified compounds bind to Ras proteins directly [44, 45]. Nevertheless, animal experiments of these compounds are still unavailable at this moment. As a consequence, efforts switched to target the downstream pathways of Ras including the Raf-MEK-ERK pathway and PI3K-AKT pathway, and it has achieved some success. Recently, inhibitors targeting Ral, another Ras-like GTPase which is a downstream pathways of Ras, were developed [46].

Herein, we focus on Tiam1, another downstream pathway of Ras. Tiam1 was characterized as an effector of Ras in 2002 [47], which is responsible for the activation of another Ras-like GTPase, Rac1 [48, 49]. It was reported that Rac1 activation is required for Ras transformation [50]. Rac1 is known to be involved in a lot of normal cell physiological processes including actin dynamics, cell trafficking, cell growth and cell motility [51, 52]. Recently, aberrant activation of Rac1 has been associated with cancer cell migration [53]. The activation of Rac1 requires GEF to catalyze the reactions of GDP release and allows GTP binding. Tiam1 is one of these GEFs responsible for Rac1 activation. Interestingly, Tiam1 has also been found overexpressed in multiple cancers, contributing to cancer cell migration [54, 55]. All these evidence suggests that Tiam1 is a promising target for cancer treatment, especially cancer metastasis. However, this pathway has not

been effectively targeted yet. To date, there is not any small-molecule compound reported to directly bind to Tiam1 and the known Rac1 inhibitors can only produce limited inhibition activity with IC_{50} around $50\mu\text{M}$ [56]. Another strategy is to inhibit Rac1-Tiam1 protein-protein interaction, which shows better efficacy in inhibiting Rac-GTPase level in cancer cells with best IC_{50} of $2.5\mu\text{M}$ [57]. Herein, we propose a novel approach to inhibit Tiam1 using small-molecule inhibitors targeting cPH domain of the Tiam1 protein. It has been shown that Phosphatidylinositol phosphate (PIP) activates Tiam1 through binding to cPH domain [58]. Loss of this binding prevents Rac1 activation *in vivo* [59]. Different from other PH domains, cPH domain of Tiam1 is not responsible for the membrane binding function, but critical for Rac1 activation, given the evidence that mutation of cPH domain abolishes the activation of Rac1 while the membrane association is maintained [59]. Another characteristic of this PH domain is that two loops exist between $\beta 1/ \beta 2$ and $\beta 3/ \beta 4$ strands, which are not observed in other PH domains. A big but flexible pocket was formed between these two loops in the cPH domain which provide the opportunity for binding of compounds.

We hypothesized that PIP competitive inhibitors may exhibit pharmacology through inhibition of Rac1. However, the binding site of PIP in cPH domain is unknown. In this study, we conducted a comprehensive analysis of all PH-PIP bound structures available in PDB to predict the putative binding site of PIP in this PH domain. Through our in-house integrated platform, herein we reported two series of compounds, to our knowledge, which are the first inhibitors directly bind to Tiam1 with strong affinity. These compounds have shown selective inhibition of prostate cell proliferation, invasion, and migration.

Figure 3.1 Workflow of in silico screen process.



3.2 Methods and materials

3.2.1 Sequence alignment and sequence logo generation

All PH-PIPs structure complexes available in PDB (**Table 5.1**) were collected to perform multiple sequence alignment based on their secondary structures using STRAP [35]. The output of the alignment was then used to generate the signatures of conserved residues involved in PIP binding using Weblogo web server[36].

3.2.2 Ensemble docking

In order to explore the structural flexibility of Tiam1 cPH, a 6-ns molecular dynamics simulation was performed on cPH domain of human Tiam1 and snapshots were saved every 10ps. The snapshots were clustered based on single linkage algorithm with a cutoff of 0.1nm using Gromacs 5.0.6. As a result, 10 clusters were generated and one representative structure in each cluster of snapshots was chosen for the ensemble docking study. The variation of these selected structures was analyzed through comparison of their mutual RMSD values. The snapshots showed the good diversity and reasonably represent the conformational variation of the cPH domain of Tiam1.

3.2.3 Chemical dataset and virtual screening

A collection of 10 million compounds was curated from various sources such as Maybridge, Chembridge, and PubChem. The chemical structures were processed and washed using MOE software[60]. In this process, hydrogens were added and the protonation state of ionizable groups were calculated. Then compound structures were passed to energy minimization using the default setting in the MOE software. GOLD was utilized to perform *in silico* screening using our curated library described above through our high performing computing cluster based on the identified

pocket mentioned above. In the molecular docking studies, residues within 6Å of the PIP and small-molecule putative binding pocket were set as flexible. The binding conformations were ranked according to their Gold Scores. A protein structural pharmacophore model was generated using GRID v22c[61] as indicated in our previous publication [31]. Grid calculation was performed within a box containing the docking site with 1Å beyond each dimension. In this process, the GRID directive Move was defined as 1 (MOVE=1) in order to allow the flexibility of the side chains. The molecular interaction fields were calculated in order to define the interaction between the protein receptor and three types of probes including hydrogen bond donors, hydrophobic probes, and hydrogen bond receptors. The derived pharmacophores defined by these binding features were used to evaluate the 5,000 hits with the best scores in the screening. The compounds fit in the pharmacophore would be selected to perform cluster analysis using MACCS fingerprints on the basis of the Tanimoto coefficient. The compound with highest docking score in each cluster was selected and the docking pose was individually selected according to the molecular visualization.

3.2.4 Pharmacophore Modeling

A ligand-based pharmacophore was generated using the MOE program from the active molecules (K_D values) tested based on the enzymatic analysis.

3.2.5 Surface plasmon resonance (SPR) assays.

Binding affinity of Tiam1 cPH domain with compounds were tested using Biacore 2000. Data analysis was performed using BIAevaluation v4.1 and Biacore 2000 control software. Detailed information was described in reference [30].

3.2.6 Culture of prostate cancer and normal prostate cell lines.

Human prostate cancer cell lines (PC3 and DU145) and normal prostate cell lines (RWPE-1) were obtained from the American Type Culture Collection (ATCC). Cells were cultured in DMEM medium with 10% fetal bovine serum. Compounds were dissolved and stored in DMSO.

3.2.7 Rac1 activation assay.

Rac1 activity in cells were measured using G-LISA activation assays kit (Cytoskeleton, Inc, Denver, Co). Prostate cancer cells were cultured until almost confluence, followed by starving for 48 hours. After treatment with compounds for 4 hours, cells were harvested and tested Rac1 activity according to the protocol.

3.2.8 Wound healing assay.

PC-3 and DU145 cells were cultured in six-well plates. Once the cells formed a monolayer, a wound was made from the middle of each well in the plate. Cells were then incubated with different concentration of compounds or DMSO for 18 hours. The wound closure speed was measured by calculating the width of the gap before and after treatment of the compounds.

3.2.9 Lamellipodia formation assay.

Cells were plated in the tissue plates at a concentration of 3000 cells/well and were cultured for 12 hours. After starving for 48 hours, cells were incubated with TPH3 for 12 hours. Then the coverslips were removed and cells were fixed with paraformaldehyde (3.7% dissolved in PBS) for 10 minutes. Then cells were stained with phalloidin- rhodamine in PBS. After mounting using

ProLong Gold, cells were visualized using 40X microscope and pictures were analyzed using imaging software.

3.2.10 Colony formation assay.

Prostate cancer cells were plated in 6-well plates and cultured for 24 hours. Then cells were treated with compounds at concentrations of 10 or 20 μ M. Media was replaced twice a week and compounds were added. After 12 days, colonies were stained with crystal violet [62]. Then the number of colonies were counted using the ColCount software.

3.2.11 Matrigel invasion assay.

Prostate cancer cells were incubated with a compound or DMSO for 4 hours. Then cells were cultured on an upper chamber on which pre-coated with Matrigel. The bottom well were filled with complete media using as chemoattractants. After 24 hours, invading cells were stained using crystal violet and counted at 10 random fields.

Table 3.1. Structures used in structural multiple sequence alignment.

PDB ID	Protein name
1FAO	Cytohesin-3
1FHW	GRP1
1FOE	Tiam1
1MAI	Phospholipase C
1UPR	PEPP1
1UNQ	AKT1
1U27	Cytohesin-2
1W1D	PDK1

3.3 Results

3.3.1 Structural analysis of cPH domain of Tiam1 protein.

Structure of the cPH domain of the Tiam1 protein was retrieved from PDB (PDB ID: 1FOE). Similar to other PH domains, cPH domain of Tiam1 contains seven beta sheets and a c-terminal alpha helix, which are typical components of PH domains. However, a close look of this crystal structure reveals unique characteristics. The loops between $\beta 1$ and $\beta 2$ (loop $\beta 1/\beta 2$), $\beta 3$ and $\beta 4$ strands (loop $\beta 3/\beta 4$) are much longer than other PH domains with crystal structures available. According to the mutagenesis analysis performed by other groups, this big loop is involved in the binding with PI3P and PI5P. It is also reported that mutations of lysines to Glutamines on this loop disrupt the binding of PI3P and PI5P to the Tiam1 cPH domain and significantly impair Rac1 activation in cell lines [59]. The existence of this big loop indicates high structural flexibility, which significantly increases the difficulty to identify the active site of PIP and small molecules. In order to take into account of structural flexibility of this domain, a 6-ns molecular dynamics simulation was performed on the apo structure of Tiam1 cPH domain using GROMACS. According to our experience, 6-ns MD simulation has the best performance of generating structure ensembles for docking studies. Root mean square fluctuations of C-alpha Atoms of residues reveal that residues from 50-70 have the highest fluctuations (**Figure 3.2**). Not surprisingly, these residues comprise the huge loop of cPH domain of Tiam1, which contributes to the binding of PI3P and PI5P. In order to obtain most of the possible conformations of this flexible domain, snapshots were taken every 10ps during the simulation and in total 600 snapshots were generated. To get the most diverse structures for following ensemble docking study, cluster analysis was then performed using GROMACS cluster package. This analysis produced 10 clusters of structures and a representative frame of each cluster was selected for our following ensemble docking study. The

superimposed structures of these ten snapshots were shown in **Figure 3.3**. In order to make sure these snapshots are diverse and can represent the snapshots, RMSD between every two structures were calculated. It was shown that these values range from 1-4, which indicated these clusters represent a good diversity of the c-terminal PH domain of the Tiam1 protein (**Figure 3.4**). The variation of RMSD value in each cluster is reasonable which renders the good quality of our clustering.

Figure 3.2 Root mean square fluctuations of C-alpha Atoms of residues in Tiam1 protein. It shows that residues from 50-70 have the highest fluctuations, which is consistent of the long loop in ther C terminal.

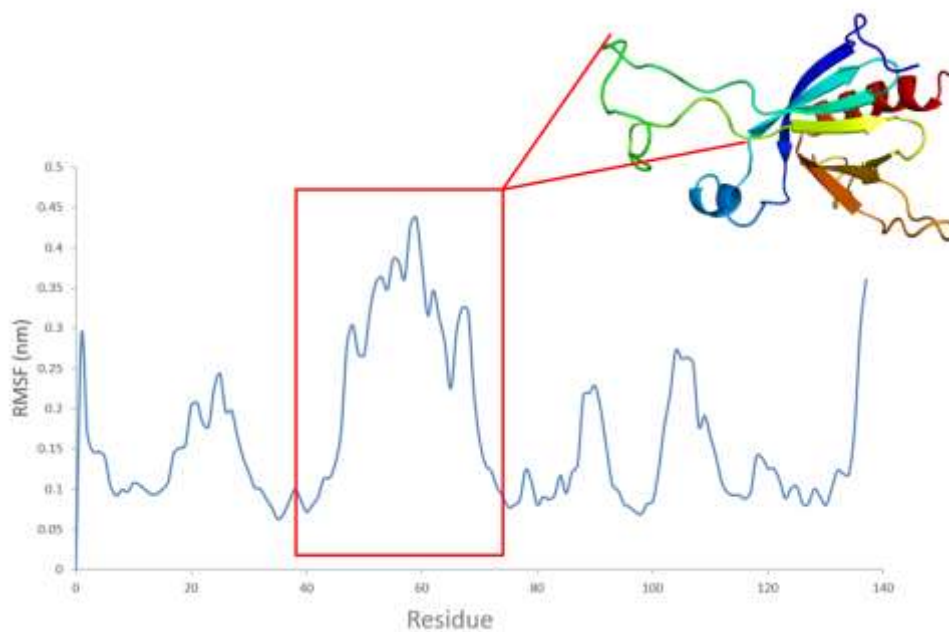


Figure 3.3. Selected snapshots from MD simulation for ensemble docking study.

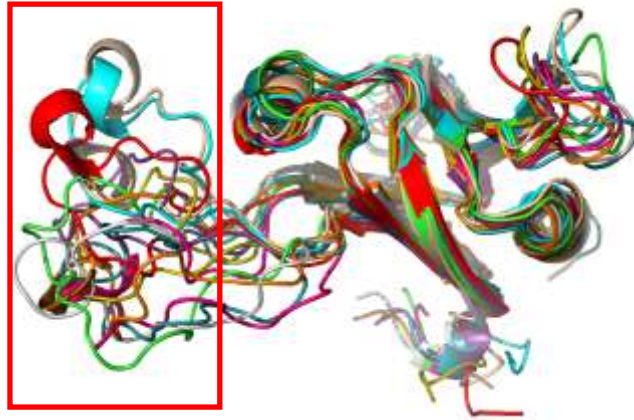


Figure 3.4. RMSDs between each pair of structures.

	1	2	3	4	5	6	7	8	9	10
1	NA	2.314	2.735	2.901	2.549	2.628	2.878	2.565	3.177	3.041
2		NA	1.847	2.367	2.755	2.195	2.417	2.637	3.466	3.907
3			NA	2.049	2.362	2.387	2.558	2.428	3.749	3.956
4				NA	2.442	2.218	2.232	2.138	3.539	3.952
5					NA	2.182	2.519	1.93	3.014	2.702
6						NA	1.64	2.238	2.926	3.494
7							NA	1.844	2.502	3.406
8								NA	2.286	2.489
9									NA	2.124
10										NA

3.3.2 Characterization of PIP binding site in cPH domain of Tiam1.

Although cPH domain does not involve in the process of Tiam1 protein translocation and structurally slightly different from other typical PH domains, PI3P and PI5P have been shown to bind to this domain and play critical roles in the regulation of Tiam1 activity. Also, loss of PIP binding was shown to impair the Rac1 activation in vivo [58, 59]. Although crystal structures of Tiam1 cPH domain have been determined, there is no PIP or drug-like compound bound Tiam1 protein complex solved at this point. Determination of the PIP and PIPs binding site is a critical step in order to rationally develop inhibitors to impair Tiam1 activity. According to the mutation analysis, loop $\beta 1/\beta 2$ and loop $\beta 3/\beta 4$ are involved in the interaction with PI3P and PI5P. However, the PIP is relative small compound while the loops are too large to pinpoint the PIP accurate binding site. Therefore, we collected all PH-PIP complexes in PDB and performed multiple sequence alignment based on their secondary structures using software STRAP. In order to identify the conserved residue for PIP binding, the sequence logos of the PH-PIP complex were generated based on alignment results using Weblogo3 (**Figure 3.5**). Interestingly, we found several invariant residues located on the two loops including Lysines, Arginines, and Tryptophans that were involved in the direct interaction between PIPs and PH domain. These residues provide positively charged interface, which is favorable for the binding of a phosphate group. Since cPH of Tiam1 selectively bind to PI3P and PI5P, we then selected all PIPs-PH complexes containing 3-phosphoinositide and 5-phosphoinositide for another structural alignment. We found three residues, Lys-1286, Tyr1304, Arg-1330 (In the nomenclature of 1FOE) are strictly conserved and these residues contribute to direct interaction with PIPs (**Figure 2B**). Of the nine PH-PIPs domain complexes, eight proteins have a lysine and an arginine in the consistent positions. Actually, we visualized the structure of the protein (1UPR) which does not have lysine in the alignment, there

is a lysine in the PIPs binding pocket and forms direct interaction with one of the phosphate groups. As a result, our docking site was defined as the center of Lys1286, Tyr1304 and Arg1330 with a radius of 10Å, which covered the putative binding sites of all ten selected structures. Electrostatic potential surface maps of cPH domain reveal positive charge around the putative binding pocket (**Figure 3.6**). In order to further characterize residues critical for ligand binding, the binding site of Tiam1 cPH domain was investigated with GRID. For GRID calculation, a grid box was constructed to enclose the target with 1Å beyond each dimension; molecular interaction fields (MIFs) were calculated with three types of probes: hydrophobic residues, nitrogen atoms (hydrogen bond acceptor), and water molecules (hydrogen bond donor). Local energy minima were derived for these three MIFs so that the corresponding residues could be identified to analyze the interactions between the protein and the small-molecule ligands. The PIP compound comprises of a long hydrophobic acyl tail and a flexible inositol phosphate group. Taking into account that the flexibility and hydrophobicity of the long acyl tail may decrease the accuracy of docking and simulation, we only use the PIP head group, rather than full-length PIP, in our current docking and simulation studies. Ensemble docking was performed against this putative binding site using maximal searching efficiency and all conformations were kept for visualization. Then MD simulations were performed on selected binding poses to refine the protein-ligand complex structures. All the PIP-PH complexes bind stably during the simulations. Poses selection are based on these criteria: ability of PIP to bind to lipid tail and attach to the cell membrane; interaction with the conserved basic residues; docking score. The best binding poses with high GOLD docking scores are shown in **Figure 3.7**. Interestingly, in our study, PI3P and PI5P have very similar docking score. The Gold Scores of PI3P and PI5P are 60.62 and 62.38 respectively. Moreover, both of these compounds share very similar docking poses. It seems the 5-phosphate group tends

to flip to the 3-phosphate position in order to acquire lower energy and more stable binding mode. In this way, both of PI3P and PI5P form hydrogen bonds with Tyr1304 and Arg1330. However, for both PI3P and PI5P, each phosphate can only form one hydrogen bond in our best-scored conformations. In this study, we also included PI(4,5)P, a compound showed very weak binding in experimental results, as a negative control of our docking studies. Interestingly, in our docking process, PI(4,5)P was not able to dock into the pocket with reasonable conformations based on our criteria. One possibility is that the 4-phosphate group position was not able to form polar interactions with the receptor due to the large space in that area. All these observations were highly in agreement with the experimental results that binding affinity of PI3P and PI5P ($\sim 20\mu\text{M}$) are much weaker compared to PIPs with more than one phosphate group ($10\text{nM}\sim 590\text{nM}$), reflecting the accuracy of our docking study.

Figure 3.5 Structural sequence alignment of Tiam1 protein crystal structures.

```

1FOE.pdb  D..L.....SM
1UNQ.pdb  SMSD.....VA
1U27.pdb
1FHW.pdb

1FOE.pdb  GDLLHTSVIWLNPPASLGKW..KEPELAAFV.KTAVVLVYK.....
1UNQ.pdb  I..VKEGWLHKRGEYI.....K..TWRPRYFLK.NDGTFIGYKERPQ.
1U27.pdb  P..DREGWLLKLGGGR.....VK..TWKRRWFILT....DNCLYFFEYTTD
1FHW.pdb  P..DREGWLLKLGGRV.....K..TWKRRWFILT....DNCLYFFEYTTD
          β1                β2                β3

1FOE.pdb  ...DGSKQKKKLVGSHRLSIYEEWDPFFRHMIPTEA.LQVRALPSADA
1UNQ.pdb  ...DVDQ.....REAPLNNFSVA.Q.CQLMKTE....
1U27.pdb  ...K.....EPRGIIPLE..N.LSIREVD....
1FHW.pdb  ...K.....EPRGIIPLE..N.LSIREVE....
          β4                β5

1FOE.pdb  EA.....N..AVCEIVH....VKSESEGRPE.....
1UNQ.pdb  ..R.....P.R.P..NTFIIRCLQWTT.....V.....
1U27.pdb  ..D.....PRK.P..NCFELY.....IPNNKGQLIKAC
1FHW.pdb  ..D.....PRK.P..NCFELY.....NPSHKGQVIKAC
          β6

1FOE.pdb  .....RVFHLCCSSPESRKDFLKSVHSILRDK
1UNQ.pdb  .....IERTFHVETPEEREEWTTAIQTVADGLKKQEEEE
1U27.pdb  KTEADGRVVEGNHMVYRISAPTQEEKDEWIKSIQAAVSVD
1FHW.pdb  KTEADGRVVEGNHVVYRISAPSPEEKEEWMKSIKASISRDPFYD
          β7

```

Figure 3.6 Electrostatic potential surface maps of Tiam1 cPH domain reveal positive charge around the putative binding pocket.

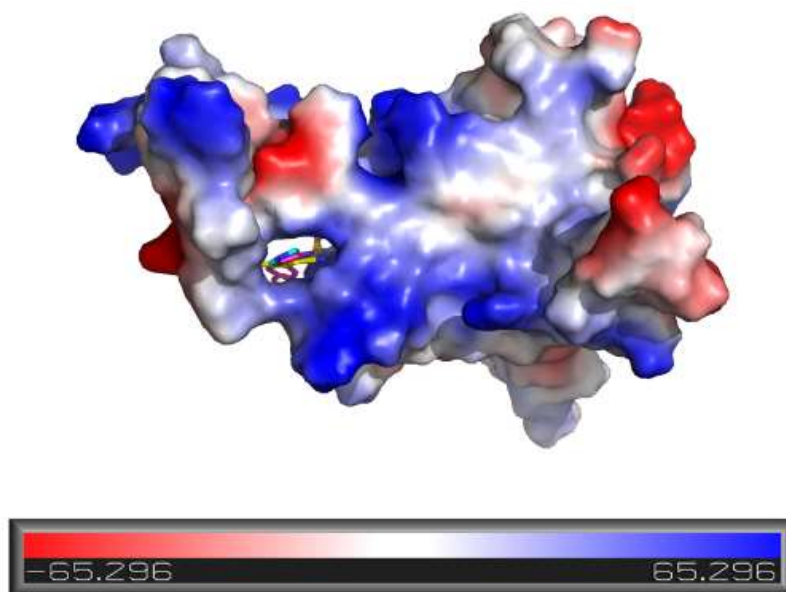
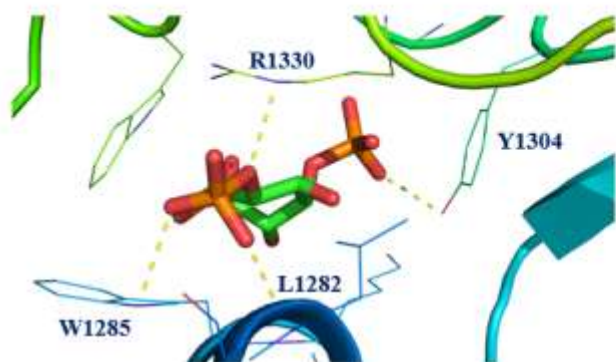
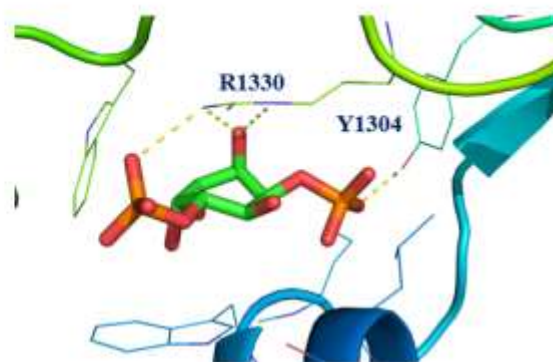


Figure 3.7. Predicted PI3P and PI5P binding poses with best docking scores.



Putative binding mode
of PI(3)P to cPH TIAM1

Gold score: 60.62



Putative binding mode
of PI(5)P to cPH TIAM1

Gold score: 62.38

3.3.3 *In silico* discovery of inhibitors of Tiam1 binding to cPH domain.

Aiming to identify small-molecule compounds fitting into the binding pocket of PI3P and PI5P, our in-house chemical database containing ten million commercially available compounds were screened using the GOLD software. Moreover, a protein structure pharmacophore was generated as a filter of the virtual screening hits (**Figure 3.8**). Residue R1330, K1284, K1286, K1287, K1305, were selected as residues in favor of interacting with hydrogen bond acceptors; Residue D1306 is identified as a hydrogen donor; Residues W1285, W1326, Y1304, F1331 form hydrophobic probes which prefer to interact with hydrophobic moieties. Considering the possible false positive result caused by pain compounds, the docking results were then filtered using our in-house platform for pain compound prediction. The top 5000 hits were then further performed cluster analysis on the basis of their chemical diversity. 203 clusters of compounds were generated and the hits with highest docking scores in each cluster was chosen and performed another docking experiment with flexible residues in the binding pocket and maximum searching efficiency. The 100 hits with highest docking scores were manually visualized to analyze their interaction in the docking results. Finally, we selected 22 top-ranked compounds to test their binding affinity and inhibition activity of Rac1 *in vitro*. ADMET properties were calculated using MOE software as described in the Methods and materials section. Seven out of the selected 22 compounds were found to bind to Tiam1 cPH domain ($KD < 50 \mu M$) using SPR binding assays. Compounds TPH3 ($KD = 0.73 \pm 0.1 \mu M$, $IC_{50} = 5.9 \pm 0.1 \mu M$) (**Figure 3.10**) (**Figure 3.11**), TPH3 ($KD = 2.7 \pm 0.2 \mu M$, $IC_{50} = 2.38 \pm 0.98 \mu M$) were two identified hits with strong binding affinity to Tiam1 cPH domain and potent inhibition of Rac1 (**Figure 3.9**). Subsequently, these two compounds were treated with PC-3 prostate cancer cell line for 4 hours at concentrations ranging from 1 μM to 20 μM . Then Rac1 activities of these cells were detected using G-LISA Kit BK128. TPH3 significantly decreased the

Rac1-GTP binding in a dose-dependent manner, indicating inhibition of the cellular activity of Rac1 (**Figure 3.12**). At the concentration of 10 μM , TPH3 exhibited as much as 40% inhibition of Rac1 activation in cells and the IC_{50} value of this compound was $2.38 \pm 0.98 \mu\text{M}$. Interestingly, TPH3 showed relative stronger binding affinity while weaker inhibition of Rac1 in cells.

Figure 3.8. A structure-based pharmacophore was generated using GRID method as a filter of the virtual screening hits.



Figure 3.9. Putative binding site of TPH3 in Tiam1 cPH domain.

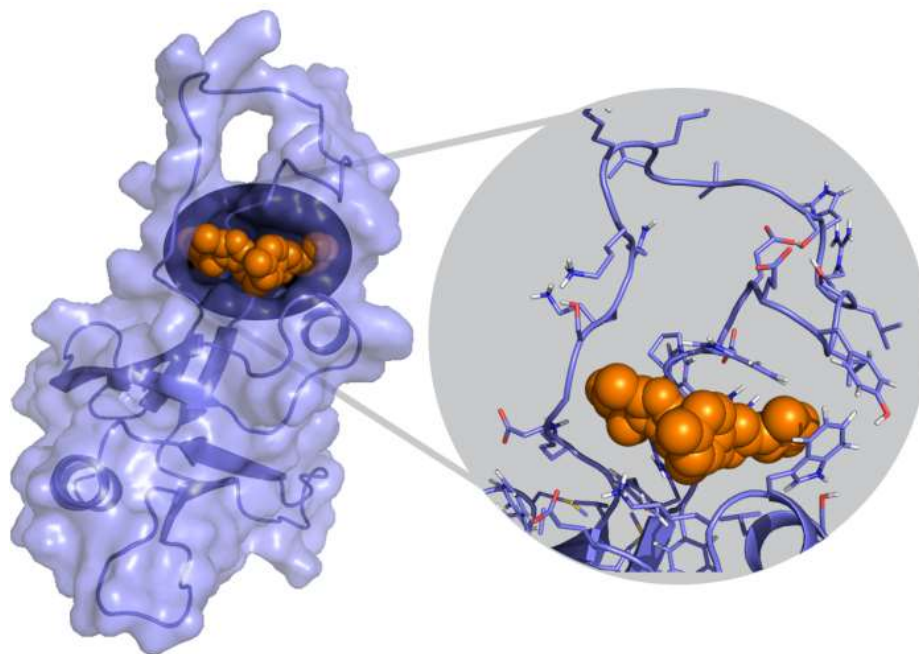
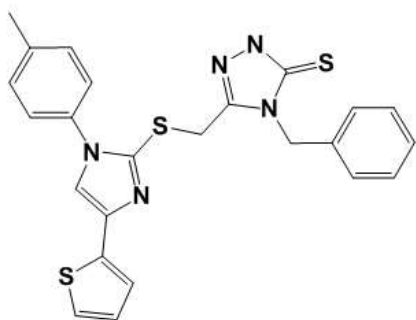
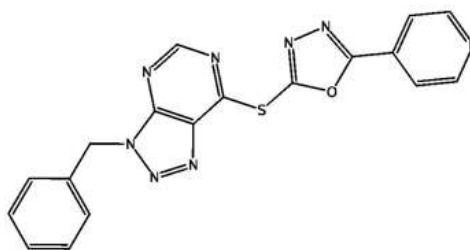


Figure 3.10 Chemical structures of the hits.



(TPH3)



(TPH7)

Figure 3.11 TPH3 binds to the Tiam1 cPH domain.

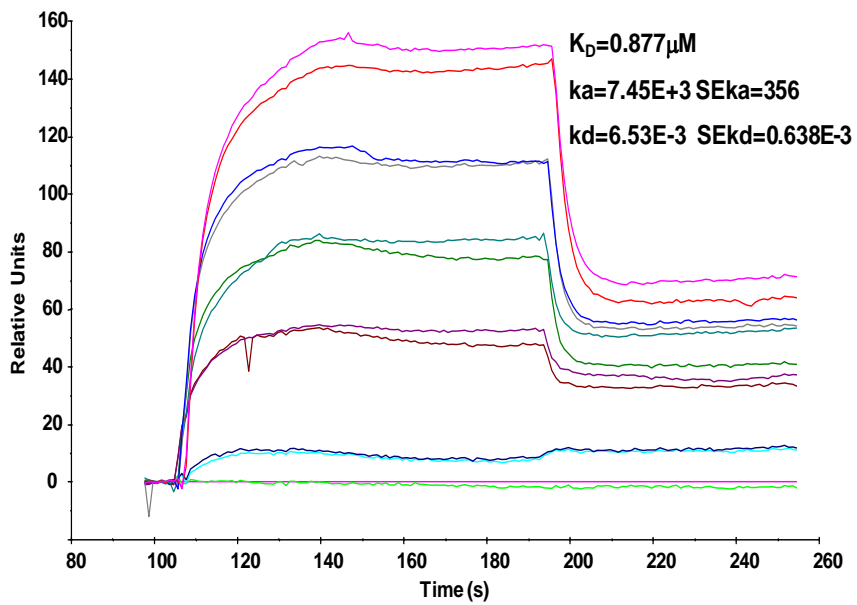
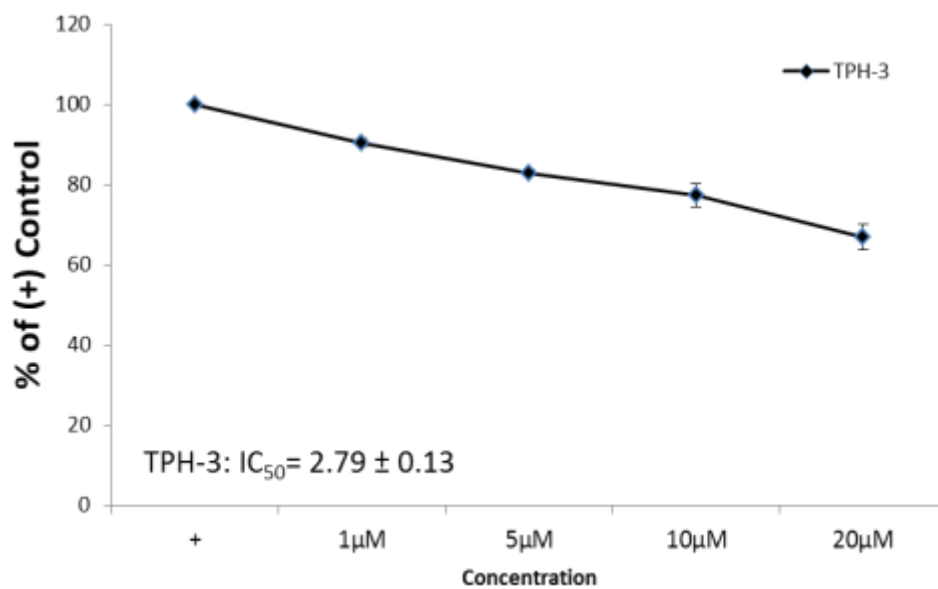


Figure 3.12 TPH3 inhibits Rac1 activity in cells.



3.3.4 TPH3 inhibits prostate cancer cell proliferation and migration.

Rac1 has been known to be a critical controller of the cell motility through regulating actin cytoskeleton in cells. [63]. Lamellipodia is one of most important protein induced by Rac1 [64]. With the treatment of 20 μ M TPH3 for 18 hours, the wound was unable to close in PC3 cells. **(Figure 3.13)** In order to clarify whether the impaired cell motility of PC3 cells were due to lamellipodia dysfunction, lamellipodia formation and matrigel cell invasion studies were performed. Interestingly, the lamellipodia formation was significantly reduced in the treatment group compared with the control group. **(Figure. 3.14)**. Thus, the reduction of prostate cancer cell motility after the treatment of TPH3 may be due to the downregulation of lamellipodia and actin disruption. Then we tested the effect of TPH3 in invasion assays using prostate cancer cells. Treatment with 20 μ M TPH3 reduced the number of invading cells in the lower chambers **(Figure. 3.15)**. Finally, TPH3 decreased the colony forming abilities of PC-3, and DU145 at the concentration of 20 μ M **(Figure. 3.16)**. In a nutshell, these results demonstrated that TPH3 reduced the cells capacities to migrate and invade in prostate cancer cell line models.

Figure 3.13. TPH3 inhibits wound healing in prostate cancer cell line.

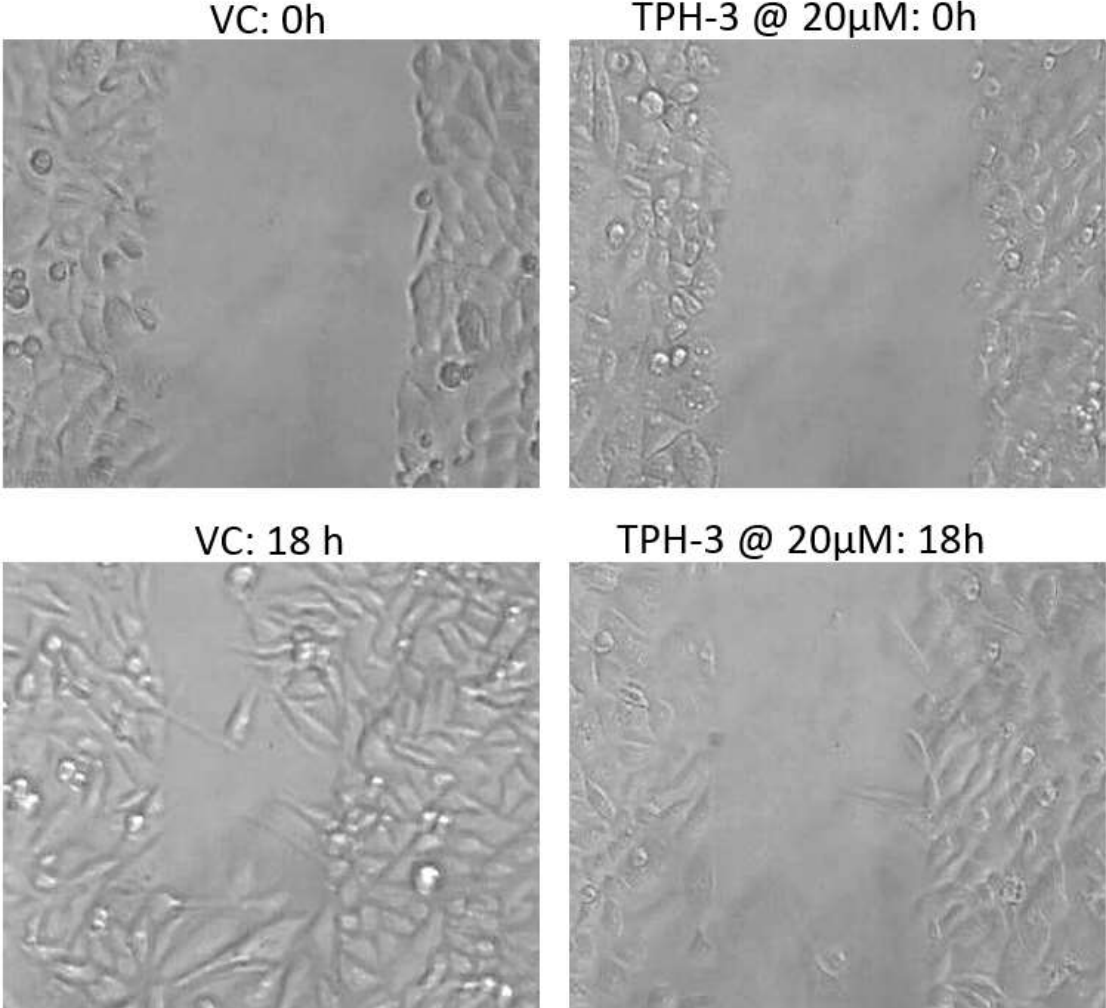


Figure 3.14 TPH3 inhibited lamellipodia formation in prostate cancer cells.

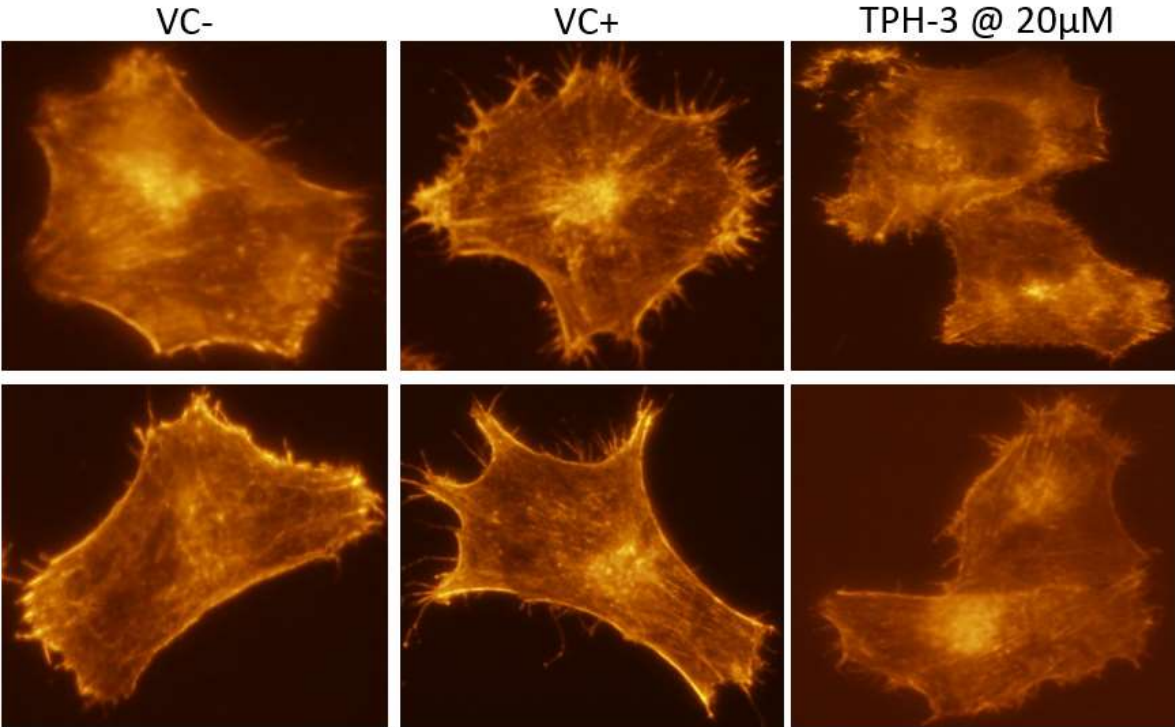


Figure 3.15 TPH3 inhibits prostate cancer cell invasion in Matrigel invasion assay.

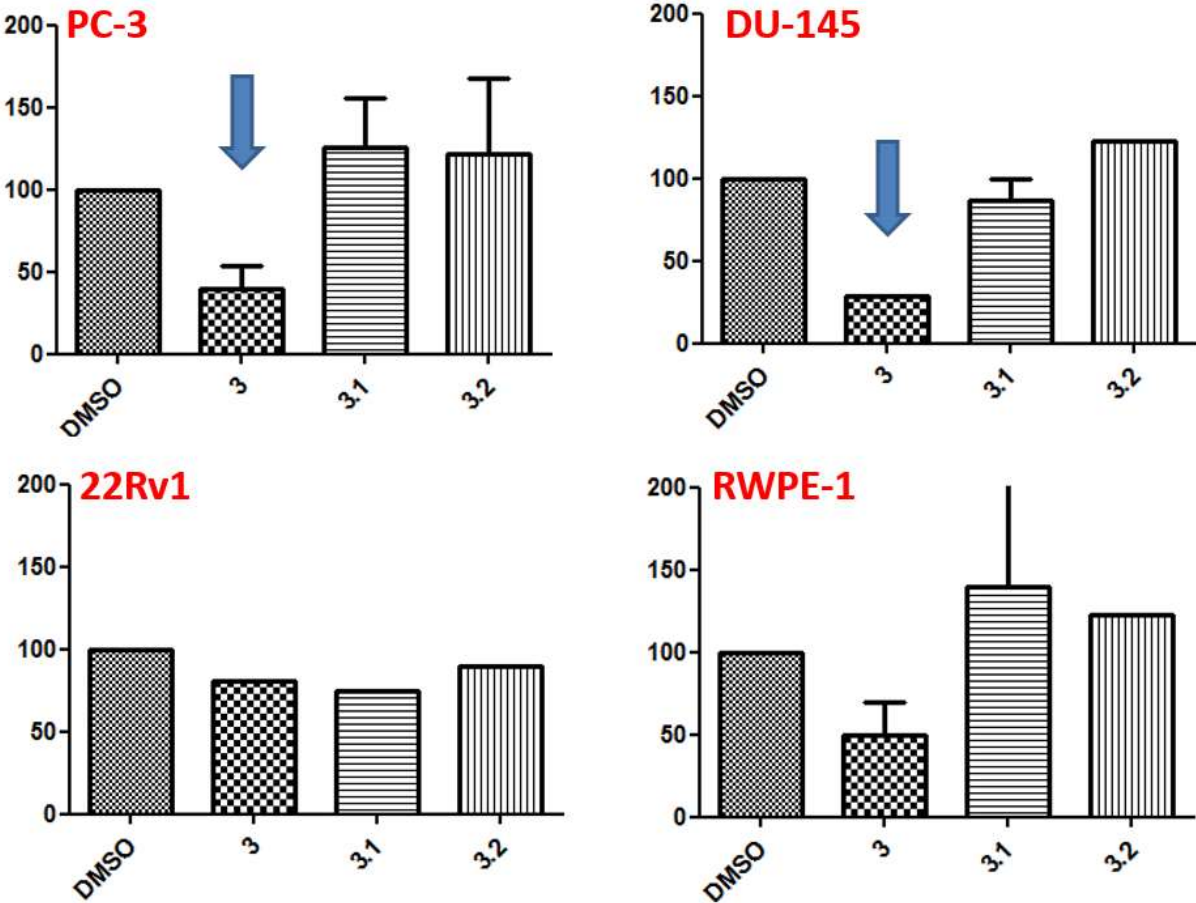
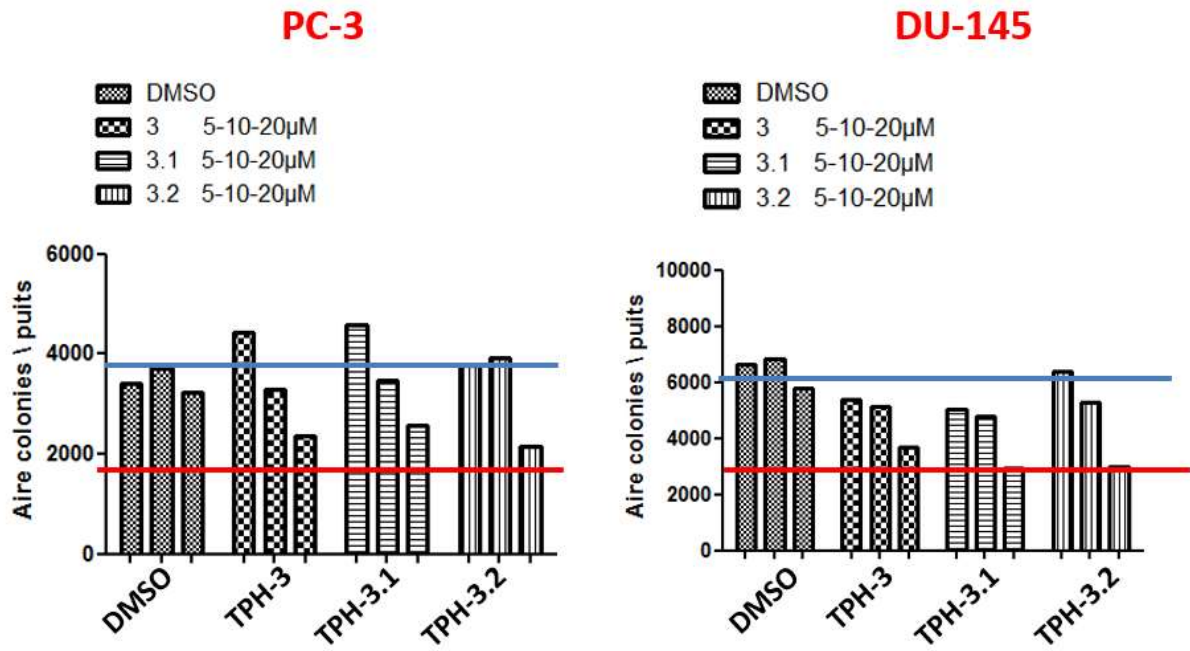


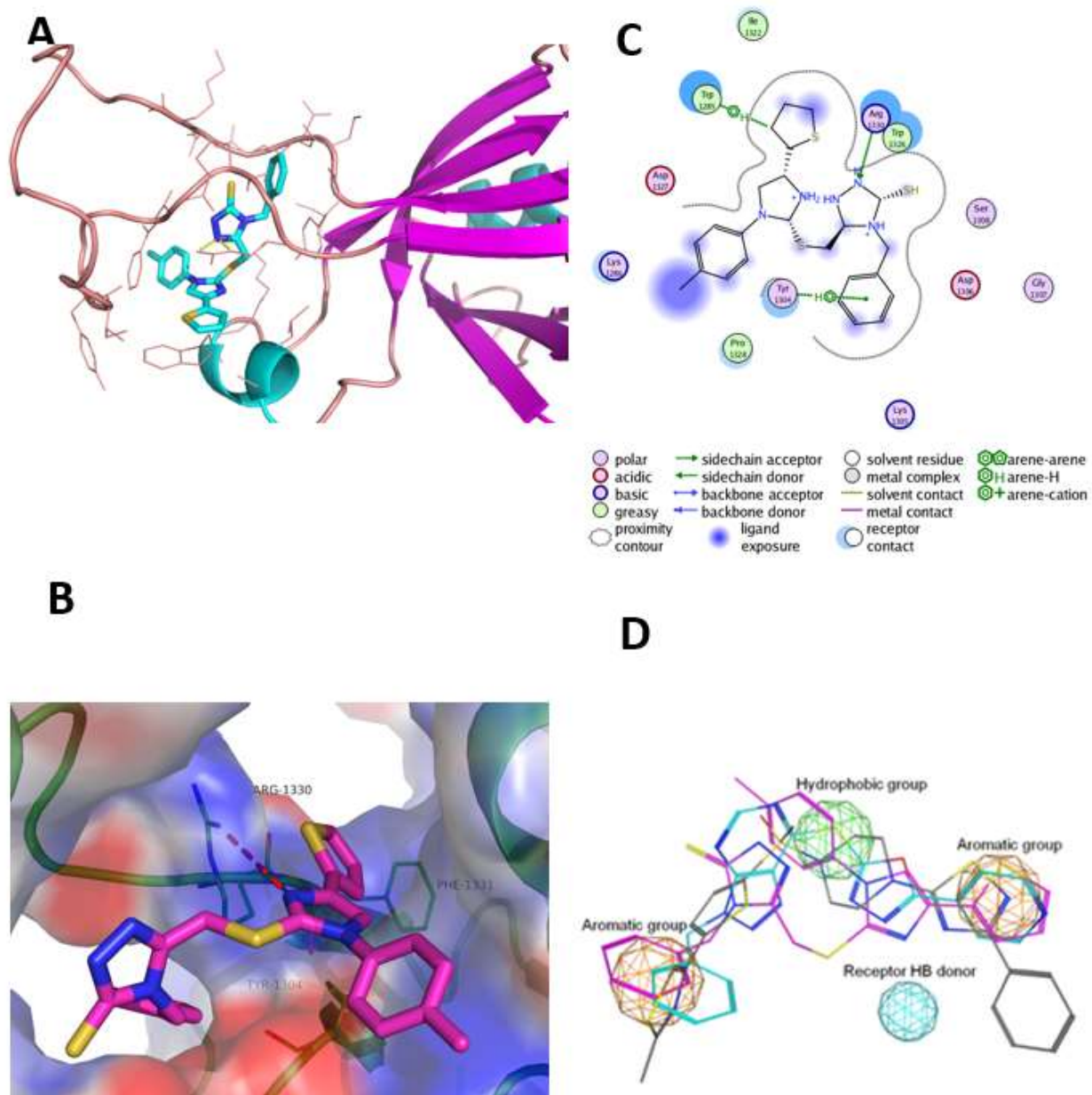
Figure 3.16 TPH3 inhibits prostate cancer cell proliferation in colony assay.



3.3.6. Complex structure refinement and Prediction of molecular interaction between TPH3 and cPH Tiam1.

At this moment, we are not able to determine the cPH-PIP or cPH-TPH3 complex structures, probably due to the high flexibility of the long loops in this domain. In order to obtain more detailed insight into the structural basis of binding of our experimentally verified inhibitors with Tiam1, we subsequently performed more careful docking studies using GOLD software followed by structural refinement using MD simulation to obtain the most stable binding modes. Not surprising, the docking score of TPH3 (GOLD Score = 84.48) is significantly higher than TPH3 (GOLD Score=73.16), and it is consistent with our experimental data in which TPH3 had better binding affinity than TPH7. And such consistency also strongly supports the binding modes of our inhibitors. **Figure 3.17A** shows that both TPH3 and TPH7 fit in a similar binding pocket. Trp1285 forms interaction with a hydrogen in thiophene of TPH3 and Tyr1304 forms H-arene interaction with the benzene group of TPH3 (**Figure 3.17B**). TPH7 has similar interaction with these two residues (**Figure 3.17C**). Especially, both of these two compounds form hydrogen bonds with Arg1330, a critical residue involved in binding of PIP with Tiam1 cPH domain. Interestingly, TPH3 forms two hydrogen bonds with the receptor while TPH7 can form only one. The better binding to this Arg1330 residue may also explain the better binding affinity of TPH3 comparing to TPH7. Based on these active compounds, we generate the pharmacophore of the inhibitors of Tiam1 cPH domain. All the inhibitors strongly bind to Tiam1 cPH domain contain two aromatic groups and a hydrophobic group which forms interactions with residues around (**Figure 3.17C**). They also receive hydrogen from Arg1330 to form hydrogen bonds. Together, our model indicates how TPH3 and TPH7 interact with Tiam1 cPH domain and critical components of the active inhibitors (**Figure 3.17D**).

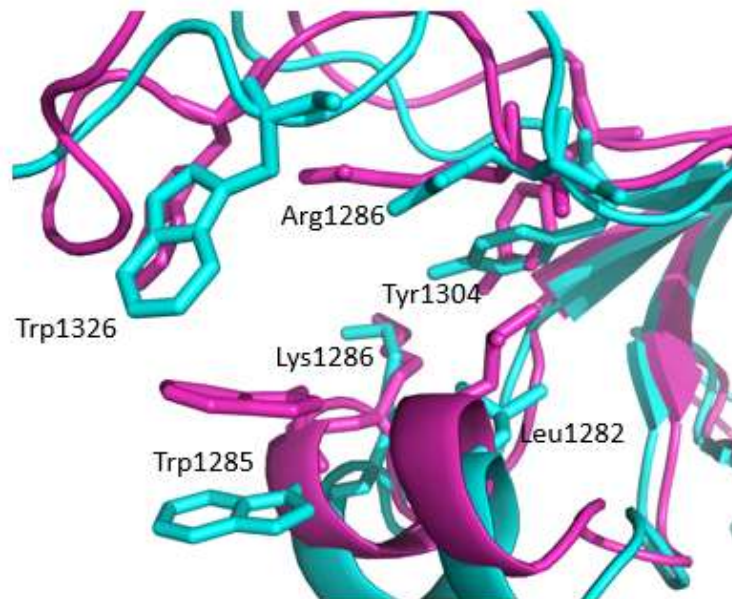
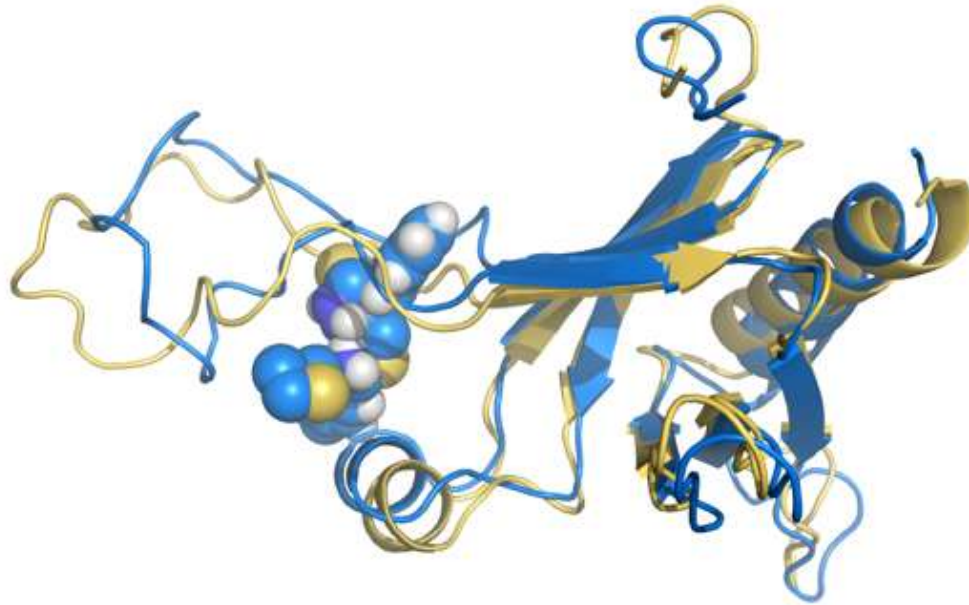
Figure 3.17 Complex structure refinement and Prediction of molecular interaction between TPH3 and cPH Tiam1.

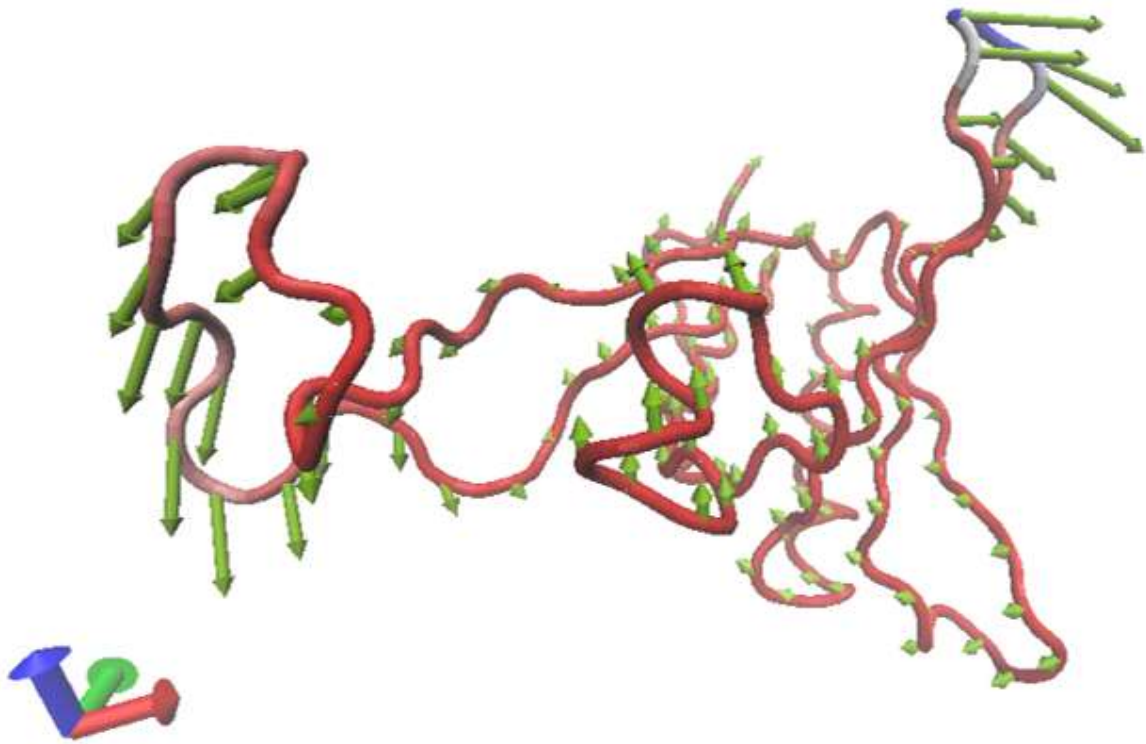


3.3.7. Bound ligands induce Tiam1 cPH domain conformational changes and stabilize protein complexes.

In order to investigate the structural changes of cPH induced by our inhibitor, we performed 100ns MD simulation on the cPH-TPH3 complex using the selected docking pose as the initial structure. The protein-ligand structure bound stably during the simulation. We were surprised to find that the alpha helix between $\beta 1/ \beta 2$ strands moves upward, which is not observed either in the MD simulation of apo and PIP bound Tiam1 cPH domain (**Figure 3.18**). As a result, Trp1285, a residue in the alpha helix, moves upward significantly and forms interaction with a hydrogen in thiophene. This movement forms a hydrophobic core which favorably interacts with aromatic groups which exist both in TPH3 and TPH7. At the same time, the loop $\beta 3/\beta 4$ strands move inward and form a “closed” conformation. As a consequence, the position of Trp1326 significantly changes and forms interaction of the terminal aromatic group of TPH3, which is not seen in TPH7. This may also partly explain why TPH3 has better binding affinity than TPH3. Also, Arg1330 is able to move inward and forms critical hydrogen bonds with a nitrogen on triazolidine. With the binding of our ligands, the orientation of Tyr1304 shifts and forms interaction with the benzene group of the compound. Rearrangement of these side chains created a pocket with a favorable interaction between the ligand and the protein, which properly explain the binding mechanism of our inhibitors. To obtain the mechanical mechanism how our protein-ligand complexes move, we built an Anisotropic Network Model (ANM) for normal mode analysis on our Tiam1-TPH3/9 complex models. The predicted movement of the structure complex is in agreement with our molecular dynamics simulation results. Also, principle component analysis (PCA) was also performed on our MD simulation snapshots (**Figure 3.18**). .

Figure 3.18 Bound ligands induce Tiam1 cPH domain conformational changes and stabilize protein complexes.





3.4 Discussion

Great progress has been made in delineating the relationship between Rac1, Tiam1, and prostate cancer metastasis. Over-activation of Rac1 was identified to increase cancer cell motility [53]. Therefore, Rac1 has received a lot of attention as a target for cancer therapy. Initially, Gao et al. reported a small-molecule compound, NSC23766, which bound to Rac1-GEF interactions. This compound inhibited Rac1 activation induced by TrioN in activity with an IC_{50} of around $50\mu M$. Based on the same model, Ruffoni et al subsequently de novo designed a diverse small-molecule lead compound that bound to Rac1-Tiam1 protein-protein interaction. The most active compound showed Rac1 inhibition with an IC_{50} of $2.5\mu M$ in smooth muscle cells. All these reported compounds were focused on Rac1-GEF protein-protein interactions. In sharp contrast, we use another strategy to target cPH domain of Tiam1 in order to achieve better potency and efficacy. Another important reason that our interest shifts to developing compounds targeting Tiam1 is that upregulated Tiam1 has been also observed in prostate cancer patients and it is highly associated with tumor metastasis. Also, Tiam1 was reported to be as an independent overall survival marker of prostate cancer patients. Therefore, our compounds may not only contribute as a probe to obtain a better understanding of Tiam1 signaling in cancer cells, but also potentially benefit patients with aberrant Tiam1 expression levels.

Besides providing a particularly appealing target for cancer metastasis, Tiam1 is also a very interesting target as a protein containing two PH domains. The N-terminal PH domain serves as a typical function related to protein translocation to the cell membrane, which is similar to other PH domains. However, its cPH domain is more related to Rac1 activation and it prefers to bind to the PI(5)P instead of PIPs with two or more phosphate groups. The atypical function of cPH domain

may be related to its different structure compared to other PH domains. Two long loops are observed in this PH domain between $\beta 1/\beta 2$ and $\beta 3/\beta 4$ sheets. A big cavity is formed between these two big loops and the PIP was known to bind to this cavity. Despite high flexibility caused by these loops, we integrated multiple sequence alignment and ensemble docking to identify the putative PIP binding site in this cPH domain. Then based on the binding pocket, we identified two series of compounds, TPH3 and TPH7, which bound to the receptor in nanomolar or low micromolar K_D values. Both of these compounds bind to a very similar position in the pocket and shared similar pharmacophore characteristics. Both TPH3 and TPH7 have two hydrophobic moieties and form hydrogen bonds with Arg1330 of the receptor. Upon binding to the receptor, TPH3 and TPH7 induce conformational changes within the cPH domain. The alpha helix located in loop $\beta 1/\beta 2$ moves upward and the loop $\beta 3/\beta 4$ moves inward to form a “closed conformation”, which enhance the binding of these two compounds.

Although our compounds showed strong binding affinity to the Tiam1 cPH domain and efficient inhibition of Rac1 activity, a number of limitations and questions remain. Firstly, the mechanism of how PIP is involved in Rac1 activation is unknown. A hypothesis is that the PIP binding changes the orientation of the protein around the cell membrane and promotes the Rac1 activation. And the conformational changes induced by our inhibitors may inhibit the structural change in PIP binding state and further inactivate Tiam1 and Rac1. Further experiments need to be performed to verify these hypotheses. Another question is that TPH3 has better binding affinity but lower inhibition effect compared to TPH7. One possible reason is the permeability of TPH3 to the cell membrane is not as good as TPH7. Improvement of the cell membrane permeability will be important to get a more potent compound in TPH3 series.

In summary, through our integrated computational platform, we successfully identified the first Tiam1 inhibitor that binds to its cPH domain and exhibits inhibition of cancer cell progression and migration. The compounds efficiently inhibit Rac1 activation induced by Tiam1. Although we have identified useful compounds for the purpose of exploring Tiam1 signaling pathways, it is still far from the application in the clinic. From the translational perspective, the selectivity of our compounds against other PH domains needs to further test; while from the structural perspective, optimization of these two leads will focus on increasing the cell membrane permeability of TPH3 and increasing binding affinity of TPH7.

Chapter 4 Summary and future directions

4.1 Summary of PH domain protein features

In chapter 2, we have performed comprehensive genomics analysis of PHGs across 13 most common cancer types in TCGA dataset. Most frequently mutated PHGs were identified: AKT1, AKT2, Tiam1, OBSCN, KALRN, TRIO, PREX2, and PLCL1. AKT1, FAM129B, SPTBN1 and Tiam1 are most frequently upregulated genes among all 313 PHGs. PHGs were overrepresented in different pathways including endocytosis, Ras signaling pathway, and B/T cell receptor signaling pathways. Across all the 13 types of cancers, PH domain proteins have highest mutation rates in colon cancer, uterine cancer and lung adenocarcinoma. Interestingly, the variety of the mutation numbers in these samples were much higher comparing to other cancer types. Of the 313 PHGs, Tiam1 was found to be one of the top10 most frequently mutated genes. But the mutation frequency did not show significant difference in different cancer types. Then we found that Tiam1 was significantly increased in neuroendocrine prostate cancer, but not other molecular type of prostate cancers. Structure multiple sequence alignment revealed conserved residues like Arginine and Tryptophan in $\beta 1/\beta 2$ loop and $\beta 3/\beta 4$ loop were involved in PIPs binding. Within PH domain, the frequency of Tiam1 mutation was also among the top 10 most mutated, which makes Tiam1 an interesting target in prostate cancer cell therapy. Also, we collected all available PH domain proteins and their PIPs binding selectivity data to build a webservice for public availability.

4.2 Summary of small molecule inhibitors targeting Tiam1

In chapter 3, we identified two potent and selective Tiam1 cPH domain small molecule inhibitors using an integrative in silico screening method. Hits were picked from an ensemble docking

experiment using 10 diverse snapshots obtained from molecular dynamics. Our molecular modeling results showed that the compounds bound to a pocket between loop $\beta 1/\beta 2$ and loop $\beta 3/\beta 4$. These two compounds showed strong binding affinity to Tiam1 cPH domain and inhibition of Rac1 activity in prostate cancer cells. TPH3 also showed inhibition of cancer cell proliferation in prostate cancer cell lines. Also, TPH3 inhibited prostate cancer cell migration using a wound healing assay. Finally, through ANM and PCA analysis of snapshots obtained from molecular dynamics experiments, we found that the Tiam1 intrinsic motility at least partially contributes to the binding of the compounds.

4.3 Future directions in small molecule inhibitors targeting Tiam1

Although our compounds showed promising binding affinity to the Tiam1 and significant inhibition of Rac1 activity in cells, more potent compounds are still in need. Continuous study of the effect of these two compounds Future direction of computational design of small molecule inhibitors may pay more attention on the Arginine and Tryptophan within the binding pocket. However, the chemical properties such as ability to penetrate the cells should be considered in the process of in silico screen and lead optimization. Polar compounds may help with the binding affinity improvement but may be difficult to enter the cells. Another concern is the selectivity to other PH domain proteins. As is discussed above, there were around 300 PH domain proteins in human proteome and they have very similar secondary structures although with low sequence identity. However, the docking programs now available in the market have not shown good correlation between the docking scores and actual activity. As a result, new scoring functions which can improve the docking accuracy of PH domain proteins are unmet needs.

4.3 Future directions in drug discoveries targeting PH domain proteins.

As discussed above, accurate scoring of docking against PH domains remains a great challenge in small molecule inhibitor development in this protein family. So more reliable scoring functions need to be developed. PH domain proteins may go to different cellular sub localizations such as cytoplasmic membrane, endosome and Golgi apparatus. However, it is still not clear that which proteins go to which membrane by binding to which PIPs. So annotation of these 313 proteins would be an important step in order to develop small inhibitors targeting these proteins. Next is to achieve the selectivity of compounds. One method is to use machine learning techniques to build a model to predict PIPs binding specificity based on published data.

Bibliography

1. Haslam RJ, Koide HB, Hemmings BA. Pleckstrin domain homology. **Nature**. 1993;363:309-310.
2. Mayer BJ, Ren R, Clark KL et al. A putative modular domain present in diverse signaling proteins. **Cell**. 1993;73:629-630.
3. Michell RH. Inositol derivatives: evolution and functions. **Nature reviews Molecular cell biology**. 2008;9:151-161.
4. Di Paolo G, De Camilli P. Phosphoinositides in cell regulation and membrane dynamics. **Nature**. 2006;443:651-657.
5. Maffucci T, Falasca M. Analysis, Regulation, and Roles of Endosomal Phosphoinositides. **Method Enzymol**. 2014;535:75-91.
6. Lemmon MA, Ferguson KM, O'Brien R et al. Specific and high-affinity binding of inositol phosphates to an isolated pleckstrin homology domain. **Proceedings of the National Academy of Sciences of the United States of America**. 1995;92:10472-10476.
7. Garcia P, Gupta R, Shah S et al. The pleckstrin homology domain of phospholipase C-delta 1 binds with high affinity to phosphatidylinositol 4,5-bisphosphate in bilayer membranes. **Biochemistry**. 1995;34:16228-16234.
8. Stauffer TP, Ahn S, Meyer T. Receptor-induced transient reduction in plasma membrane PtdIns(4,5)P2 concentration monitored in living cells. **Current biology : CB**. 1998;8:343-346.
9. Paterson HF, Savopoulos JW, Perisic O et al. Phospholipase C delta 1 requires a pleckstrin homology domain for interaction with the plasma membrane. **The Biochemical journal**. 1995;312 (Pt 3):661-666.
10. Franke TF, Kaplan DR, Cantley LC et al. Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. **Science**. 1997;275:665-668.

11. Fruman DA, Rameh LE, Cantley LC. Phosphoinositide binding domains: embracing 3-phosphate. **Cell**. 1999;97:817-820.
12. Klippel A, Kavanaugh WM, Pot D et al. A specific product of phosphatidylinositol 3-kinase directly activates the protein kinase Akt through its pleckstrin homology domain. **Molecular and cellular biology**. 1997;17:338-344.
13. Dowler S, Currie RA, Campbell DG et al. Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. **The Biochemical journal**. 2000;351:19-31.
14. Rameh LE, Arvidsson A, Carraway KL, 3rd et al. A comparative analysis of the phosphoinositide binding specificity of pleckstrin homology domains. **The Journal of biological chemistry**. 1997;272:22059-22066.
15. Levine TP, Munro S. Targeting of Golgi-specific pleckstrin homology domains involves both PtdIns 4-kinase-dependent and -independent components. **Current biology : CB**. 2002;12:695-704.
16. Lemmon MA, Ferguson KM, Schlessinger J. PH domains: diverse sequences with a common fold recruit signaling molecules to the cell surface. **Cell**. 1996;85:621-624.
17. Falke JJ. Membrane Recruitment as a Cancer Mechanism: A Case Study of Akt PH Domain. **Cellscience**. 2007;4:25-30.
18. Cantley LC. The phosphoinositide 3-kinase pathway. **Science**. 2002;296:1655-1657.
19. Engelman JA, Luo J, Cantley LC. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. **Nature reviews Genetics**. 2006;7:606-619.
20. Bader AG, Kang S, Zhao L et al. Oncogenic PI3K deregulates transcription and translation. **Nature reviews Cancer**. 2005;5:921-929.
21. Der CJ, Van Dyke T. Stopping ras in its tracks. **Cell**. 2007;129:855-857.
22. Carpten JD, Faber AL, Horn C et al. A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. **Nature**. 2007;448:439-444.

23. Yi KH, Lauring J. Recurrent AKT mutations in human cancers: functional consequences and effects on drug sensitivity. **Oncotarget**. 2016;7:4241-4251.
24. Bayascas JR, Wullschleger S, Sakamoto K et al. Mutation of the PDK1 PH domain inhibits protein kinase B/Akt, leading to small size and insulin resistance. **Molecular and cellular biology**. 2008;28:3258-3272.
25. Powis G, Aksoy IA, Melder DC et al. D-3-deoxy-3-substituted myo-inositol analogues as inhibitors of cell growth. **Cancer chemotherapy and pharmacology**. 1991;29:95-104.
26. Kozikowski AP, Sun H, Brognard J et al. Novel PI analogues selectively block activation of the pro-survival serine/threonine kinase Akt. **Journal of the American Chemical Society**. 2003;125:1144-1145.
27. Meuillet E, Stratton S, Prasad Cherukuri D et al. Chemoprevention of prostate cancer with selenium: an update on current clinical trials and preclinical findings. **Journal of cellular biochemistry**. 2004;91:443-458.
28. Kumar CC, Madison V. AKT crystal structure and AKT-specific inhibitors. **Oncogene**. 2005;24:7493-7501.
29. Mahadevan D, Powis G, Mash EA et al. Discovery of a novel class of AKT pleckstrin homology domain inhibitors. **Mol Cancer Ther**. 2008;7:2621-2632.
30. Moses SA, Ali MA, Zuohe S et al. In vitro and in vivo activity of novel small-molecule inhibitors targeting the pleckstrin homology domain of protein kinase B/AKT. **Cancer Res**. 2009;69:5073-5081.
31. Chen L, Du-Cuny L, Moses S et al. Novel inhibitors induce large conformational changes of GAB1 pleckstrin homology domain and kill breast cancer cells. **PLoS computational biology**. 2015;11:e1004021.
32. Finn RD, Attwood TK, Babbitt PC et al. InterPro in 2017-beyond protein family and domain annotations. **Nucleic acids research**. 2017;45:D190-D199.

33. UniProt Consortium T. UniProt: the universal protein knowledgebase. **Nucleic acids research**. 2018;46:2699.
34. Van den Eynden J, Fierro AC, Verbeke LP et al. SomInaClust: detection of cancer genes based on somatic mutation patterns of inactivation and clustering. **BMC bioinformatics**. 2015;16:125.
35. Gille C, Frommel C. STRAP: editor for STRuctural Alignments of Proteins. **Bioinformatics**. 2001;17:377-378.
36. Crooks GE, Hon G, Chandonia JM et al. WebLogo: a sequence logo generator. **Genome research**. 2004;14:1188-1190.
37. Kandath C, McLellan MD, Vandin F et al. Mutational landscape and significance across 12 major cancer types. **Nature**. 2013;502:333-339.
38. Gauthier NP, Reznik E, Gao J et al. MutationAligner: a resource of recurrent mutation hotspots in protein domains in cancer. **Nucleic acids research**. 2016;44:D986-991.
39. Gao J, Aksoy BA, Dogrusoz U et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. **Science signaling**. 2013;6:p11.
40. Vonkova I, Saliba AE, Deghou S et al. Lipid Cooperativity as a General Membrane-Recruitment Principle for PH Domains. **Cell reports**. 2015;12:1519-1530.
41. Vogelstein B, Papadopoulos N, Velculescu VE et al. Cancer genome landscapes. **Science**. 2013;339:1546-1558.
42. Cox AD, Fesik SW, Kimmelman AC et al. Drugging the undruggable RAS: Mission possible? **Nature reviews Drug discovery**. 2014;13:828-851.
43. Tan Z, Zhang S. Past, Present, and Future of Targeting Ras for Cancer Therapies. **Mini reviews in medicinal chemistry**. 2015.
44. Ostrem JM, Peters U, Sos ML et al. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. **Nature**. 2013;503:548-551.

45. Hunter JC, Gurbani D, Ficarro SB et al. In situ selectivity profiling and crystal structure of SML-8-73-1, an active site inhibitor of oncogenic K-Ras G12C. **Proceedings of the National Academy of Sciences of the United States of America**. 2014;111:8895-8900.
46. Yan C, Liu D, Li L et al. Discovery and characterization of small molecules that target the GTPase Ral. **Nature**. 2014;515:443-447.
47. Lambert JM, Lambert QT, Reuther GW et al. Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. **Nature cell biology**. 2002;4:621-625.
48. Michiels F, Habets GG, Stam JC et al. A role for Rac in Tiam1-induced membrane ruffling and invasion. **Nature**. 1995;375:338-340.
49. Kraemer A, Goodwin M, Verma S et al. Rac is a dominant regulator of cadherin-directed actin assembly that is activated by adhesive ligation independently of Tiam1. **American journal of physiology Cell physiology**. 2007;292:C1061-1069.
50. Khosravi-Far R, Solski PA, Clark GJ et al. Activation of Rac1, RhoA, and mitogen-activated protein kinases is required for Ras transformation. **Molecular and cellular biology**. 1995;15:6443-6453.
51. Waterman-Storer CM, Worthylake RA, Liu BP et al. Microtubule growth activates Rac1 to promote lamellipodial protrusion in fibroblasts. **Nature cell biology**. 1999;1:45-50.
52. Bosco EE, Mulloy JC, Zheng Y. Rac1 GTPase: a "Rac" of all trades. **Cellular and molecular life sciences : CMLS**. 2009;66:370-374.
53. Yang WH, Lan HY, Huang CH et al. RAC1 activation mediates Twist1-induced cancer cell migration. **Nature cell biology**. 2012;14:366-374.
54. Li J, Liang S, Jin H et al. Tiam1, negatively regulated by miR-22, miR-183 and miR-31, is involved in migration, invasion and viability of ovarian cancer cells. **Oncology reports**. 2012;27:1835-1842.

55. Hofbauer SW, Krenn PW, Ganghammer S et al. Tiam1/Rac1 signals contribute to the proliferation and chemoresistance, but not motility, of chronic lymphocytic leukemia cells. **Blood**. 2014;123:2181-2188.
56. Gao Y, Dickerson JB, Guo F et al. Rational design and characterization of a Rac GTPase-specific small molecule inhibitor. **Proceedings of the National Academy of Sciences of the United States of America**. 2004;101:7618-7623.
57. Ruffoni A, Ferri N, Bernini SK et al. 2-Amino-3-(phenylsulfanyl)norbornane-2-carboxylate: an appealing scaffold for the design of Rac1-Tiam1 protein-protein interaction inhibitors. **Journal of medicinal chemistry**. 2014;57:2953-2962.
58. Viaud J, Lagarrigue F, Ramel D et al. Phosphatidylinositol 5-phosphate regulates invasion through binding and activation of Tiam1. **Nature communications**. 2014;5:4080.
59. Baumeister MA, Martinu L, Rossman KL et al. Loss of phosphatidylinositol 3-phosphate binding by the C-terminal Tiam-1 pleckstrin homology domain prevents in vivo Rac1 activation without affecting membrane targeting. **The Journal of biological chemistry**. 2003;278:11457-11464.
60. Vilar S, Cozza G, Moro S. Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. **Current topics in medicinal chemistry**. 2008;8:1555-1572.
61. Grid
62. Stea B, Falsey R, Kislin K et al. Time and dose-dependent radiosensitization of the glioblastoma multiforme U251 cells by the EGF receptor tyrosine kinase inhibitor ZD1839 ('Iressa'). **Cancer Lett**. 2003;202:43-51.
63. Habets GG, Scholtes EH, Zuydgeest D et al. Identification of an invasion-inducing gene, Tiam-1, that encodes a protein with homology to GDP-GTP exchangers for Rho-like proteins. **Cell**. 1994;77:537-549.
64. Hall A. Rho GTPases and the actin cytoskeleton. **Science**. 1998;279:509-514.

Vita

Zhi Tan was born in Zhongshan, China, the son of Rongfang Tan and Xiaoqiu Li. After completing his work at Zhongshan Memorial Middle School, Zhongshan, China in 2007, he entered Wuhan University in Wuhan, China. He received the degree of Bachelor of Medicine with a major in Clinical Medicine from Wuhan University in June, 2012. In September of 2012, he entered The University of Texas Graduate School of Biomedical Sciences at Houston.

Publications

Tan, Z., L. Chen and S. Zhang. "Comprehensive Modeling and Discovery of Mebendazole as a Novel Traf2- and Nck-Interacting Kinase Inhibitor." *Sci Rep* 6, (2016): 33534.

Tan, Z., R. Chaudhai and S. Zhang. "Polypharmacology in Drug Development: A Minireview of Current Technologies." *ChemMedChem* 11, no. 12 (2016): 1211-8.

Reddy, A. S*., **Z. Tan*** and S. Zhang. "Curation and Analysis of Multitargeting Agents for Polypharmacological Modeling." *J Chem Inf Model* 54, no. 9 (2014): 2536-43. (*Co-first author)

Robichaux, J. P., Y. Y. Elamin, **Z. Tan**, B. W. Carter, S. Zhang, S. Liu, S. Li, T. Chen, A. Poteete, A. Estrada-Bernal, A. T. Le, A. Truini, M. B. Nilsson, H. Sun, E. Roarty, S. B. Goldberg, J. R. Brahmer, M. Altan, C. Lu, V. Papadimitrakopoulou, K. Politi, R. C. Doebele, K. K. Wong and J. V. Heymach. "Mechanisms and Clinical Activity of an Egfr and Her2 Exon 20-Selective Kinase Inhibitor in Non-Small Cell Lung Cancer." *Nat Med* 24, no. 5 (2018): 638-646.

Permanent address:

No.28, Xiangyuntong, Zhongshan 3 Road,

Room C701

Zhongshan, 528400 China