bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dark kinase annotation, mining and visualization using the Protein Kinase Ontology

Saber Soleymani[3], Nathan Gravel[2], Liang-Chin Huang[2], Wayland Yeung[2], Elika Bozorgi[3], Nathaniel G. Bendzunas[1], Krzysztof J. Kochut[3]* and Natarajan Kannan[1,2]*

[1]Department of Biochemistry & Molecular Biology, University of Georgia, Athens, GA 30602, USA

[2]Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA

[3]Department of Computer Science, University of Georgia, Athens, GA 30602, USA

*Corresponding authors: nkannan@uga.edu & kkochut@uga.edu

ABSTRACT

The Protein Kinase Ontology (ProKinO) is an integrated knowledge graph that conceptualizes the complex relationships connecting protein kinase sequence, structure, function, and disease in a human and machine-readable format. Here we extend the scope of ProKinO as a discovery tool by including new classes and relationships capturing information on kinase ligand binding sites, expression patterns, and functional features, and demonstrate its application in uncovering new knowledge regarding understudied members of the protein kinase family. Specifically, through graph mining and aggregate SPARQL queries, we identify the p21-activated protein kinase 5 (PAK5) as one of the most frequently mutated dark kinase in human cancers with abnormal expression in multiple cancers, including an unappreciated role in acute myeloid leukemia. We identify recurrent oncogenic mutations in the PAK5 activation loop predicted to alter substrate binding and phosphorylation and identify common ligand/drug binding residues in PAK family kinases, highlighting the potential application of ProKinO in drug discovery. The updated ontology browser and a web component, ProtVista, which allows interactive mining of kinase sequence annotations in 3D structures and Alphafold models, provide a valuable resource for the signaling community. The updated ProKinO database is accessible at http://prokino.uga.edu/browser/.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

The protein kinase gene family with nearly 535 human members (collectively called the human kinome) is one of the biomedically important gene families with direct associations with many human diseases such as cancer, diabetes, Alzheimer's, Parkinson's, and inflammatory disorders. They make up one-third of target discovery research in the pharmaceutical industry, with over 50 FDA-approved drugs developed since 2001 (1,2). However, despite decades of research on the protein kinase family, our current knowledge of the kinome is skewed towards a subset of well-studied kinases with nearly one third of the kinome largely understudied. These understudied kinases, collectively referred to as the "dark" kinome by the Knowledge Management Center (KMC) (3) within the Illuminating the Druggable Genome (IDG) consortium, constitute both active kinases and inactive pseudokinases, which lack one or more of the active site residues, but perform important scaffolding and regulatory roles in signaling pathways (4-7) and are druggable (8). Incomplete knowledge of the structure, function, and regulation of these understudied kinases and pseudokinases presents a major bottleneck for drug discovery efforts. While multiple initiatives are beginning to generate essential tools and resources to characterize dark kinases, integrative mining of these datasets is necessary to develop new testable hypotheses on dark kinase functions.

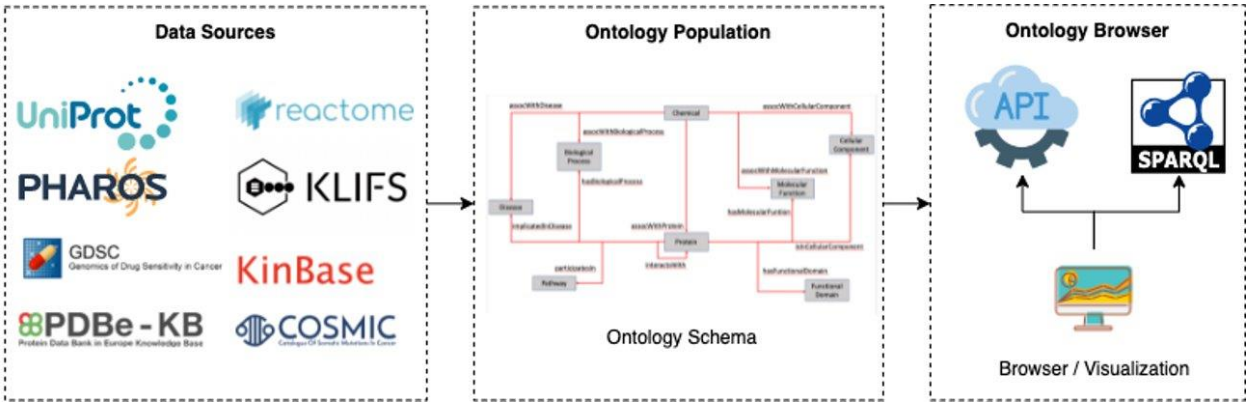

Figure 1. The ProKinO architecture and work-flow. Left panel shows a subset of curated data sources used in ontology population. The middle panel shows a schematic of the ontology schema with classes (boxes) and relationships (lines) connecting the classes. The right panel shows applications for ontology browsing and navigation.

Integrative mining of protein kinases data, however, is a challenge because of the diverse and disparate nature of protein kinase data sources and formats. Information on the structural and

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

functional aspects of dark kinases, for example, is scattered in the literature posing unique challenges for researchers interested in formulating routine queries such as “disease mutations mapping to conserved structural and functional regions of the kinome” or “post-translational modifications (PTMs) in the activation loop of dark kinases.” Formulating such aggregate queries requires researchers to go through the often time-consuming and error-prone process of collating information from various data sources through customized computer programs, which results in duplication of efforts across laboratories, and does not scale well with the growing complexity and diversity of protein kinase data. For these reasons, the IDG consortium has developed a unified resource, Pharos, for collating diverse forms of information on druggable proteins, including protein kinases (3,9,10). A focused dark kinase knowledgebase has also been developed to make experimental data available on dark kinases to the broader research community (11,12). While these unified resources provide a wider range of valuable information on druggable proteins, they offer limited data analytics capabilities in terms of mining sequence and structural data, and do not conceptualize the detailed structural and functional knowledge of protein kinases in ways protein kinase researchers use and understand. Thus, to accelerate the biochemical characterization of understudied dark kinases, a semantically meaningful and mineable representation of the kinase knowledge base is needed (**Figure 1**).

To semantically represent protein kinase data in ways protein kinase researchers use and understand, we previously reported the development of a focused protein kinase ontology, ProKinO (13-15), which integrates and conceptualizes diverse forms of protein kinase data in computer- and human-readable format (**Figure 2**). The ontology is instantiated with curated data from internal and external sources and enables aggregate queries linking diverse forms of data in one place. ProKinO enables the generation of new knowledge regarding kinases and pathways altered in various cancer types, and new testable hypotheses regarding the structural and functional impact of disease mutations (13,15-20,21 ,22-32). For example, through iterative ProKinO queries and follow-up experimental studies, we identified oncogenic mutations associated with abnormal protein kinase activation and drug sensitivity (13,16,19,21,33-35). We have also employed federated queries linking ProKinO with other widely used ontologies and resources such as the Protein Ontology (PRO), neXtProt, Reactome, and the Mouse Genome Informatics (MGI) to prioritize understudied dark kinases for functional studies and generate testable hypotheses regarding post-translational modification and cancer mutations (36).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

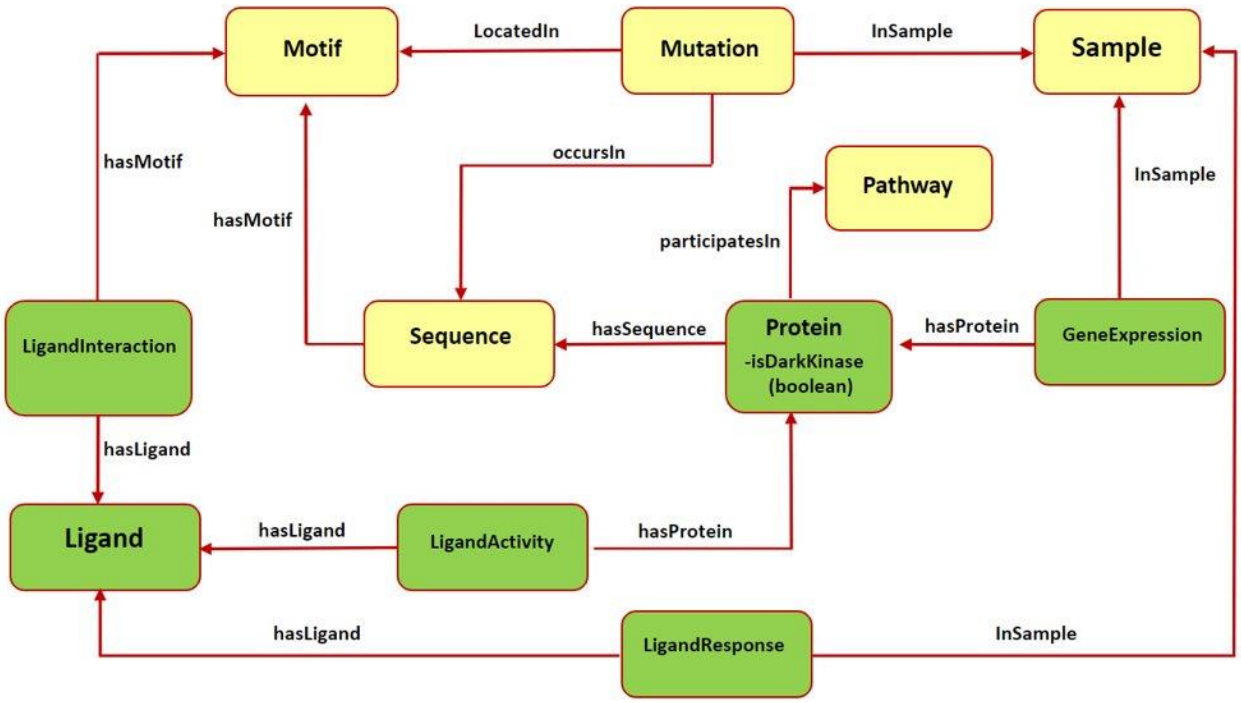

Figure 2. Subset of the updated ProKinO schema with new classes and relationships. The full schema can be accessed at http://prokino.uga.edu/. New classes are colored in green and pre-existing classes are colored in yellow. Red arrows indicate new relationships introduced to connect the new classes.

While our preliminary studies have demonstrated the utility of ProKinO in hypothesis generation and knowledge discovery, to fully realize the impact of ProKinO in drug discovery and dark kinome mining, the ontology, and the associated analytics tools need to be further developed to expand its scope and usability. For example, mutations at specific functional regions of the protein kinase domain, such as the gatekeeper and activation segments, are known to impact drug binding efficacies (37,38). Likewise, kinase mRNA expression profiles strongly correlate with drug response (39-42). Thus, integrative mining of disease mutations with drug sensitivity profiles and expression patterns can provide new hypotheses/data for the development and administration of combinatorial drugs where multiple mutated kinases in distinct pathways can be targeted for drug repurposing (43,44), as demonstrated by the repurposing of Gleevec for targeting c-kit kinase in Gastrointestinal tumors (45).

Furthermore, the recent generation of structural models of various dark kinases using AlphaFold (46) provides a new framework for generating new hypotheses by interactive mining and visualization of sequence annotations in the context of 3D models. However, the lack of interactive visualization tools to overlay sequence and functional annotations in 3D structural

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

models presents a bottleneck in the effective use of AlphaFold models for function prediction. To address this and other challenges described above related to dark kinase mining and annotation, we have expanded ProKinO by including kinase expression data, as well as a variety of data related to ligand-motif interaction, and ligand response prediction (47). We demonstrate the application of these new tools in knowledge discovery by identifying mutational hotspots in the understudied p21 activated protein kinase 5. We provide several example SPARQL queries for ontology mining and hypotheses generation. We have also significantly revamped the ProKinO browser and included new visualization tools for interactive mining of sequence annotations in the context of experimentally determined 3D structures and AlphaFold models. The updated ontology and browser provide a valuable resource for mining, visualizing and annotating the dark kinome and pseudokinome.

MATERIALS AND METHODS

Data Sources

The ProKinO ontology includes data obtained from our own sources and from various external sources. For several years, the external sources included KinBase, UniProt, COSMIC, Reactome, and PDB. We described and published the process of designing and building the ontology, retrieving the relevant data, and populating it with a vast amount of kinase-related data in (15,16). Here, we describe the recent enhancements and additions to ProKinO, focusing on using the evolutionary and functional context of well-studied kinases to annotate and generate testable hypotheses on understudied dark kinasesIn a separate, significant project, we have identified and classified nearly 30,000 pseudokinases spanning over 1,300 organisms (48). The schematic representation of the classification of kinases into groups, families and subfamilies was already in place (49,50). Consequently, the addition of the pseudokinases and their classification was relatively simple. However, it significantly enhanced ProKinO as a comprehensive knowledge graph representing kinase-related data.

The definition and nomenclature of several kinome-wide conserved motifs were standardized based on several previously published studies which describe the kinase structural features such as subdomains (49), regulatory spine/shell (31), and catalytic spine (51). A subset of redundant or family-specific motifs were removed to prevent confusion.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ligand interactions. The Kinase-Ligand Interaction Fingerprints and Structures (KLIFS) (5) is a kinase-ligand interaction database. The KLIFS stores detailed drug-protein kinase interaction information derived from diverse (>2900) structures of catalytic domains of human and mouse protein kinases deposited in the Protein Data Bank to provide insights into the structural determinants of kinase-ligand binding and selectivity at the motif and residue level. In addition, KLIFS provides an Application Programming Interface (API) for programmatic access of data related to chemicals and structural chemogenomics (5) (52). However, it lacks information regarding kinase pathways or diseases which prevents the user from investigating the effect of drug-mutant protein binding on downstream pathways or diseases. KLIFS annotations, which report PDB residue positions, were converted to UniProt residue numbering using PDBrenum (53), then converted to prototypic Protein Kinase A (PKA) numbering using Multiply Aligned Profiles for Global Alignment of Protein Sequences (MAPGAPS) (54). Entries that could not be mapped or did not map to the kinase domain were filtered out.

Ligand responses. We included the data relevant to drug sensitivity in kinases in this step. In particular, we retrieved the fitted dose and response data of kinase-relevant ligands from GDSC (55). Kinase-relevant ligands are defined based on our previous study (56), which collected 143 small-molecule protein kinase inhibitors from GDSC based on four drug-target databases: DrugBank (57), Therapeutic Target Database (58), Pharos (3), and LINCS Data Portal (59). GDSC provides the half-maximal inhibitory concentration values (IC_{50}) of these 143 ligands in 988 cancer cell lines.

Ligand activities. Ligand activities were retrieved from Pharos, a flagship resource (3) of the National Institutes of Health (NIH) Illuminating the Druggable Genome (IDG) program that includes data on small molecules, including the approved drug data and bioassay data. Based on the protein classification (60), the drug targets in Pharos include kinases, ion channels, and G-protein coupled receptors (GPCRs), and others. In this phase of the project, we decided to include the data relevant to ligand binding in kinases. Pharos integrates drug-target relationships from several resources, such as ChEMBL (61) and DrugCentral (62).

Expression data. An important part of our recent additions was kinase expression data. Genomic expression data (protein, RNA), as well as the transcription factors and epigenomic associations are among many facets of the data included in Pharos. Furthermore, GDSC repository contains gene expression data (Affymetrix Human Genome U219 Array), as well. Additionally, COSMIC's Cell Lines Project includes a significant amount of gene expression data, including kinase expression.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dark kinases. Dark kinases were labelled based on the information from Dark Kinase Knowledgebase (11).

Protein kinase knowledge graph: schema and data organization

The ProKinO ontology consists of classes, sub-classes, class types, relationships, relationship types, and constraints of protein kinase and related data (**Figure 2**). The hierarchy connects all classes to the root, which is *ProKinOEntity*. Moreover, the schema defines types and constraints for the relationships. With such explicit and constrained schema, composing queries is more intuitive than conventional relational databases. In particular, to enable integrative mining of dark kinase expression data in the context of kinase sequence and structural features, we have introduced three new classes in ProKinO, the *Ligand* class (including its name, source, and chemical structure) and the following three related classes: (1) *LigandInteraction*, placed between the *Ligand* and (already existing) *Motif* classes to capture kinase-ligand binding and selectivity at the motif and residue level, (2) *LigandActivity, placed between the Ligand and* (already existing) *Protein classes to represent* kinases targeted by ligands (and drugs), and (3) *LigandResponse,* located between the *Ligand* and (already existing) *Sample* classes and representing ligand (and drug) sensitivity in kinases. To capture kinase expression, we added the *GeneExpression* relationship linking the Protein and Sample classes. The outline of the recently added classes and their relationships in ProKinO is illustrated as a UML class diagram, shown in **Figure 2**.

ProKinO Population

The ProKinO knowledge graph is automatically populated from several external and our local data sources at regular intervals, as originally described (15), ProKinO schema and the associated knowledge graph population software are routinely updated to incorporate additional sources of data such as pseudokinase and “dark” kinase classification and incorporating information on ligand interactions, ligand responses, ligand activities, kinase expression and associated object and datatype properties. We have been using the Protégé ontology editor for the schema creation and its subsequent modifications. The organization of the schema after these modifications is available at https://prokino.uga.edu/about.

The population software has been coded in Java and uses the Jena Framework. The population process is performed in several steps to add instances, their properties, and a combination of reading the prepared data from CSV, RDF, XML, and other file formats and accessing many remote data sources using their provided API (for example, Reactome’s REST API). Entity

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

interconnections across data retrieved from different data sources are accomplished using UniProt identifiers, kinase names, and other accession identifiers. We modified the population software to create instances and properties for the newly added classes and relationships.

More specifically, using the KLIFS API, we retrieved the relevant kinases, ligands, and residue-level interaction data. The data was retrieved and then processed by custom Perl scripts. ProKinO ontology schema was modified, and ligands were included as new data, while interaction data (motifs) were either reconciled with the motifs already present in ProKinO or added as new, if not already there.

Similarly, the ligand response data was retrieved from GDSC and then processed by custom Perl scripts to create suitable CSV files. Additional ligands were included as new data, while the response data and the relevant samples were either reconciled with the samples already present in ProKinO or added as new, if not already there.

In order to populate the data on ligand activities, we retrieved from Pharos kinase-relevant ligands, as well as their binding data on targeted kinases, for example, IC50 values. This data was retrieved and then processed by custom Perl scripts to produce the necessary CSV files. Additional ligands, not included in the KLIFS dataset, were included as new data. All kinases targeted by ligands were already present in ProKinO, so they were reused in this step.

Data on kinase expression was first retrieved from Pharos, COSMIC, and GDSC. As before, the relevant kinases were already present in the ProKinO knowledge graph. The expression data was stored as individuals in the *Expression* class. Some of the relevant data about samples were already present in ProKinO, as we already had a significant amount of sample data from COSMIC. Additional samples were included as new data.

We reviewed and updated all the motifs already present in ProKinO. Furthermore, we updated the motif naming in cases where there were differences with the standard motif names.

Finally, we assembled an up-to-date list of dark kinases (11) and added a Boolean datatype property, *isDarkKinase*, to identify them among all other kinases in the ProKinO knowledge graph.

RESULTS

The expanded ontology and its knowledge graph provide a wealth of data unifying the information available on both well studied (light) kinases and understudied (dark) kinases that serve as a unified resource for mining the kinome. The current version of ProKinO (version 64),

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

includes 842 classes, 31 object and 67 data properties, and over 7.3M individuals (knowledge graph nodes). ProKinO contains information on 153 dark kinases. 137 dark kinases have information on structural motifs, 148 have disease mutations mapped to the kinase domain, 45 dark kinases have pathway information, and 26 are associated with specific reactions, as defined in Reactome.

Users can navigate the ontology using the ontology browser by searching for a specific kinase of interest or by performing aggregate SPARQL queries linking multiple forms of data. Nearly 34 pre-written queries linking different data types can be executed using the ProKinO browser (http://prokino.uga.edu/queries). A user can also download the ontology or browse data based on organisms, functional domains, diseases or kinase domain evolutionary hierarchy. Below, we focus on the application of complex SPARQL queries and the ProtVista visualization tools for the illumination of understudied dark kinases.

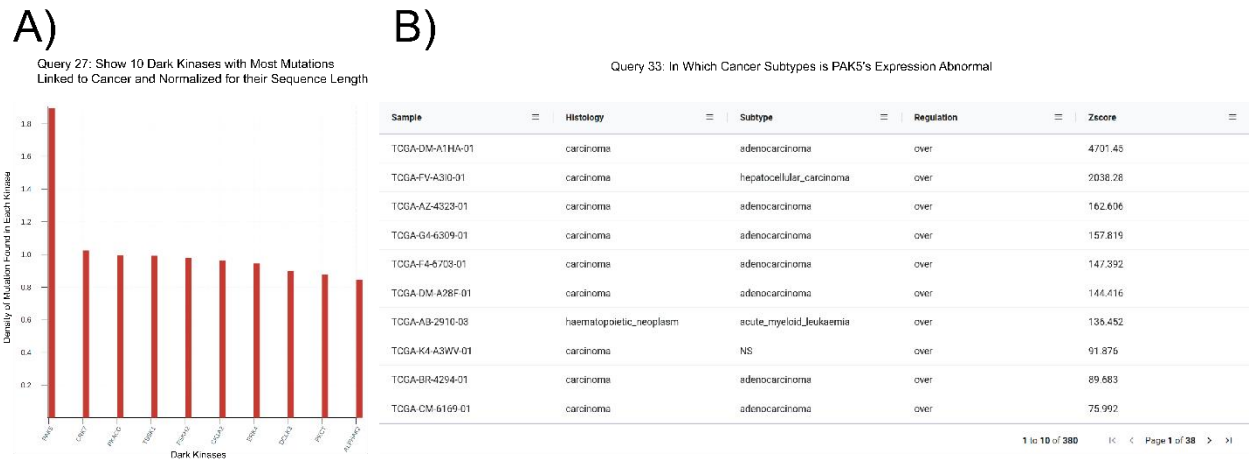

Figure 3. SPARQL query results **A)** Output of Query 27 requesting top 10 dark kinases with most mutations in cancer. **B)** Output of Query 33 listing samples with abnormal PAK5 expression. The query also lists histology, cancer subtypes, regulation and Z-score.

Mutation and expression of understudied PAK5 in human cancers: One possible way to prioritize dark kinases for functional studies is to ask the question, “which dark kinases are most mutated in human diseases, such as cancers?”. Typically answering this question would require collating and post-processing data from multiple resources such as COSMIC, Pharos, and the dark kinase knowledgebase. However, with the updated Protein Kinase Ontology, these questions can be quickly answered using SPARQL. In particular, having the *isDarkKinase* property within the *Protein* class and the RDF triples connecting *Mutation*, *Sample* and *Sequence* classes, one can formulate aggregate queries requesting all dark kinases mutated in

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cancer samples. To avoid biases introduced by the length of protein/gene sequences (longer proteins tend to have more mutations), the query can be modified to normalize mutation counts by sequence length. Executing this modified query (query 27, available at http://prokino.uga.edu/queries) displays the rank-ordered list of dark kinases based on mutational density. The top ten dark kinases with the highest mutational dentistry are shown in **Figure 3A**. Notably, the p21 activated kinase 5 (PAK5) is at the top of the list with a mutational density of 1.88, followed by CRK7 (0.995), TSSK1 (0.978), PKACG (0.977), PSKH2 (0.948), CK1A2 (0.946), ERK4 (0.9318), DCLK3 (0.876), PKCT (0.866) and ALPHAK2 (0.832). Having identified PAK5 as the most frequently mutated dark kinase in cancers, one can further query the ontology to explore the role of this kinase in various cancers. With the addition of the new *GeneExpression* class in ProKinO and the RDF triples connecting gene expression to the *Sample* and *Protein* classes (*GeneExpression:InSample: Sample*; *GeneExpression:hasProtein: Protein*), one can formulate queries requesting for PAK5 expression in different samples. Rank ordering the samples based on PAK5 expression (Query 33) reveals cancer types such as adenocarcinoma (Zscore: 4701.5) and hepatocellular carcinoma (Zscore: 2038.2) that have previously been associated with abnormal PAK5 expression (63-66). However, the role of PAK5 in other cancer types such as acute myeloid leukemia (Zscore: 136.4) is relatively understudied (67). The identification of new cancer sub-types with dark kinase expression and regulation further exemplifies the use of ProKinO in knowledge discovery.

Mutational hotspots in the activation loop of PAK5: Because ProKinO encodes a wealth of information on the structural and regulatory properties of kinases, it can be used to generate mechanistic predictions on cancer mutation impact. We demonstrate this for the PAK kinases by asking the question “where are PAK5 mutations located in the protein kinase domain?” Using the RDF triples connecting the *Mutation*, *Motif* and *Sequence* classes (*Mutation: LocatedIn: Motif; Mutation:InSequence: Sequence*), one can formulate a query (Query 28) listing mutations in different structural regions/motifs of the PAK5 kinase domain. Examination of the query results reveals that the C-terminal substrate binding lobe (C-lobe) is more frequently mutated (318 mutations) relative to the N-terminal ATP binding lobe (N-lobe: 170 mutations) (**Figure 4A**). Within the C-lobe, nearly 78 mutations map to the activation loop, which is known to play a critical role in substrate recognition and activation in a diverse array of kinases (68-70). Despite the prevalence of activation loop mutations in PAK5, there is currently no information on how these mutations impact PAK5’s structure and function. Nonetheless, based on the evolutionary relationships captured in ProKinO (based on the alignment of human kinases to the prototypic

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protein kinase A), one can formulate queries mapping mutations to specific aligned positions in the shared protein kinase domain. A query listing WT type and mutant type residues in the activation loop of PAK5 and the equivalent aligned residue positions in PKA (query 29) provides additional context for activation loop mutations in PAK5. For example, two distinct mutations map to residue S602 in the activation loop of PAK5 that structurally corresponds to a phosphorylatable residue, T197, in PKA (71). Having this contextual information provides a testable hypothesis that S602 mutations in PAK5 impact kinase phosphorylation and regulation. Likewise, WT residue P604^{PAK5} is mutated in four distinct cancer samples and this position is equivalent to PKA residue P202, which configures the activation loop for substrate recognition (72). Thus, mutation of this critical residue is expected to impact substrate binding and activation loop phosphorylation in PAK5. Additional insights into these mutations can also be obtained my visualizing these residues in the context of the PAK5 AlphaFold models using the ProtVista viewer described below.

A) Query 28: Count Unique Cancer-Linked Mutations in Different Structural Locations of the PAK5 Kinase.

Motif	Cancer mutations
C-lobe	319
N-lobe	171
activation loop	78
subdomain XI	67
subdomain VIII	64
subdomain I	62
subdomain III	44
alphaC	43
subdomain VIa	38
alphaE	36

B) Query 29: List WT and Mutant Type (Missense) Residues in the Activation Loop of PAK5.

Wild Type	Position	Mutant Type	PKA Position	Disease	Primary Site	Subsite
G	588	C	186	carcinoma	large_intestine	NS
G	588	S	186	malignant_melanoma	skin	NS
F	589	V	187	malignant_melanoma	NS	NS
F	589	S	187	carcinoma	liver	NS
E	596	Q	193	carcinoma	breast	NS
E	596	G	193	carcinoma	breast	NS
E	596	K	193	malignant_melanoma	skin	NS
S	602	L	197	malignant_melanoma	skin	NS
S	602	L	197	carcinoma	skin	head_neck
V	604	I	199	malignant_melanoma	skin	NS
V	604	F	199	carcinoma	kidney	NS
V	604	F	199	carcinoma	lung	NS
V	604	I	199	malignant_melanoma	skin	scalp
P	607	L	202	malignant_melanoma	skin	head_neck
P	607	L	202	malignant_melanoma	skin	NS
P	607	S	202	malignant_melanoma	skin	NS
P	607	L	202	carcinoma	skin	NS
P	607	S	202	carcinoma	skin	NS

Figure 4. **A)** Output of Query 28 listing the number of unique cancer-linked mutations at various structural locations of PAK5 kinase. **B)** Output of Query 29 listing unique point mutations in the activation loop of PAK5 kinase. The query also lists the equivalent PKA position, disease type, primary site of the tissue sample, and subtype of the tissue sample. Entries containing only one mutation per position were filtered from the original query.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Insights into PAK5 ligand binding sites: With the conceptualization of new information related to kinase ligands, their mode of action and interaction with specific motifs in the kinase domain, new aggregate queries linking mutated kinases to drug sensitivity profiles, mode of action, and ligand binding sites can be performed using the updated ProKinO. For example, queries such as “list proteins and drugs or ligands interacting with the protein's gatekeeper residue (GK.45)” (query 31) and “list ligands targeting the Epidermal Growth Factor Receptor (EGFR) kinase and their mode of action” (query 34) can be rapidly performed using the updated ProKinO ontology. We demonstrate the application of these new additions in the context of PAK5 by asking the question “what are the drugs targeting PAK family (PAK1-6) kinases?” Query 30 answers this question using the RDF triples connecting the *Ligand*, *Motif* and *Protein* classes (list triples) (**Figure 5**). Examination of the query results indicates multiple drugs targeting PAK family kinases, including STAUROSPORINE and N2-[(1R-2S)-2-AMINOCYCLOHEXYL] that bind to structurally equivalent residues/motifs in the ligand binding pocket of PAK4 and PAK5, respectively. The ligand binding sites, and associated interactions can also be visualized using the Protvista viewer described below. Additional queries linking dark kinases to drug sensitivities, structural motifs, and pathways are listed on the ProKinO website at https://prokino.uga.edu/queries.

Query 30: List Motifs Interacting with Ligands in the PAK Family Kinases (PAK1-6) Along with Equivalent PKA Positions.

Protein	Ligand Name	Motif	Position	PKA Position
PAK4	STAUROSPORINE	I.3	327	50
PAK5	N2-[(1R,2S)-2-AMINOCYCLOHEX...	I.3	455	50
PAK4	STAUROSPORINE	g.l.4	328	51
PAK5	N2-[(1R,2S)-2-AMINOCYCLOHEX...	g.l.4	456	51
PAK4	STAUROSPORINE	g.l.5	329	52
PAK5	N2-[(1R,2S)-2-AMINOCYCLOHEX...	g.l.5	457	52
PAK4	STAUROSPORINE	hinge.47	397	123
PAK5	N2-[(1R,2S)-2-AMINOCYCLOHEX...	hinge.47	525	123
PAK4	STAUROSPORINE	hinge.48	398	124
PAK5	N2-[(1R,2S)-2-AMINOCYCLOHEX...	hinge.48	526	124
PAK4	STAUROSPORINE	linker.51	401	127
PAK5	N2-[(1R,2S)-2-AMINOCYCLOHEX...	linker.51	529	127

Figure 5. Output of Query 30 listing ligands interactions with each PAK family member (PAK1-6). It also includes motif names and positions of full sequence and PKA positioning. The output of Query 30 was rearranged to highlight the homology of PAK4 and PAK5 motif/ligand interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

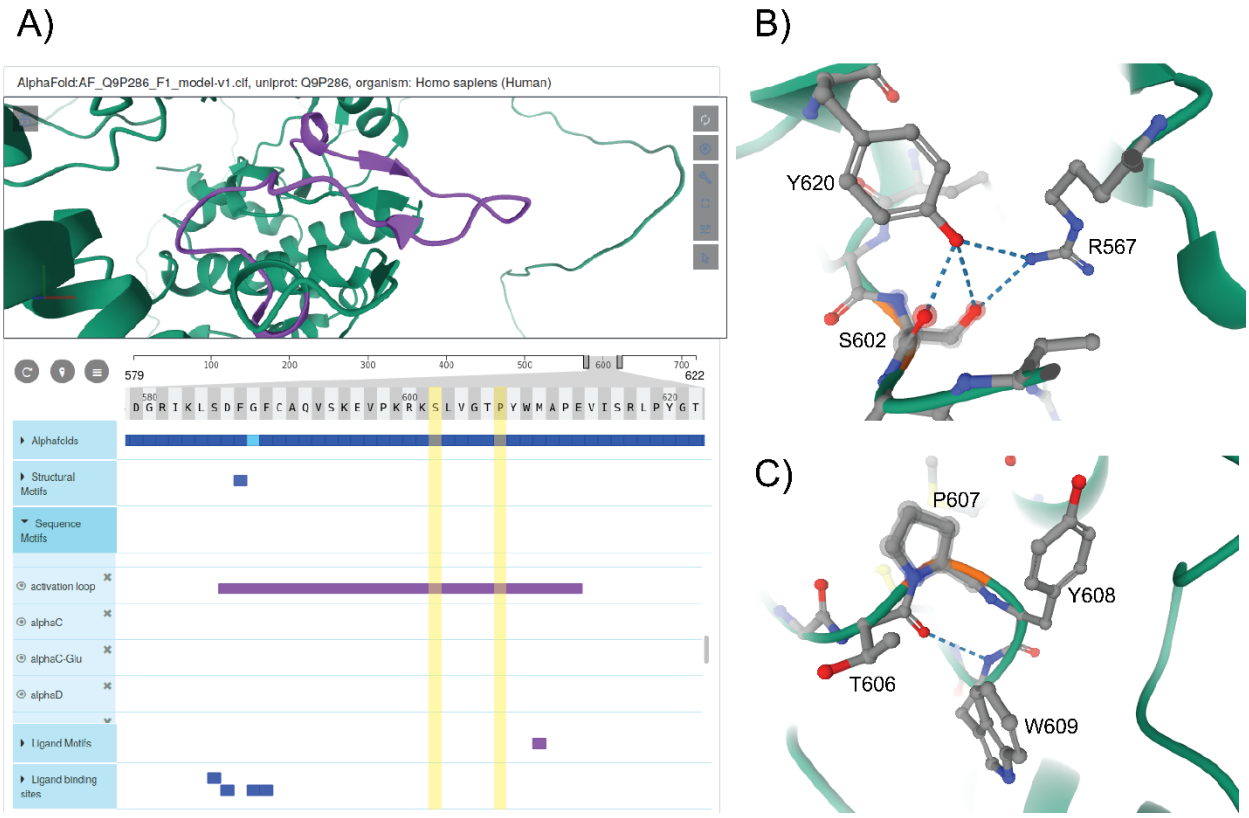

Figure 6. Protvista viewer **A)** AlphaFold model of PAK5 kinase is shown in the structure viewer (top panel) and the sequence viewer with annotations are shown in the bottom panel. **B-C)** Zoomed in view of structural interactions associated with S602 and P607 in PAK5 activation loop.

Visualization tools for dark kinase annotation and mining

ProtVista viewer: To provide structural context for cancer mutations and to enable interactive mining of dark kinase sequence annotations in the context of 3D structures and predicted models from AlphaFold (46,73), we developed and incorporated a modified version of the ProtVista viewer in ProKinO. The viewer can be deployed for any protein kinase of interest by navigating to the *Structure* tab in the protein summary page and selecting either a PDB structure or AlphaFold model of interest. A snapshot of the ProtVista viewer displaying the AlphaFold model of PAK5 kinase is shown in **Figure 6**. The ProtVista viewer uses an enhanced version of the Mol* viewer and the PDB web component (developed by the PDBe team) to provide two-way interactive navigation between 3D structure (**Figure 6A**, top panel) and annotation viewer (**Figure 6A**, bottom panel).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The annotation viewer consists of multiple tracks populated dynamically based on data from ProKinO and external sources such as UniProt. In addition, prediction confidence scores for AlphaFold models are displayed in the annotation viewer along with additional annotations such as conserved sequence motifs, subdomains, and structural motifs involved in kinase regulation. The annotation viewer also shows other annotations from external sources such as ligand binding sites and predicted functional sites. Users can hover over the residues on the 3D structure viewer to view the equivalent information on the annotation viewer and vice versa. For example, selecting the “activation loop” in the annotation viewer highlights the corresponding structural region in the AlphaFold model of PAK5 (**Figure 6A**). Likewise, the selection of residues in the activation loop (S602 and P607) in the structure viewer highlights the annotations associated with these and interacting residues in the sequence viewer. Such interactive mining is expected to accelerate the functional characterization of dark kinases and provide new insights into disease mutations. For example, visualizing the interactions associated with S602 in the activation loop of PAK5 (**Figure 6B**) indicates a hydrogen bonding interaction with R567, which is part of the conserved HRD motif (sequence annotation). Because the HRD-Arg is known to play a role in kinase regulation by stabilizing activation loop conformation (68), it provides additional context for predicting the impact of S602 altering mutations in PAK5. Likewise, examining the structural and sequence context of P604 interacting residues provides new insights into how alteration of this residue might impact substrate binding and kinase regulation. Together, these examples, highlight the value added by the ProtVista viewer in the visualization and annotation of mutations in dark kinases.

CONCLUSION AND FUTURE DIRECTIONS

In this work, we present an updated version of the protein kinase ontology for mining and annotating dark kinases. ProKinO was developed following FAIR (Findable, Accessible, Interoperable, and Reusable) principles and serves as an integrated knowledge graph for relating and conceptualizing diverse forms of disparate data related to protein kinase sequence, structure, function, regulation, and disease (cancer). We present a new ontology browser for navigating these data and demonstrate the application of aggregate SPARQL queries in uncovering new testable hypotheses regarding understudied members. We also provide several pre-written SPARQL queries that be used to rapidly retrieve a wealth of information related to protein kinase mutations, pathways, expression, and ligand binding sites. However, writing new queries requires prior knowledge of the ontology schema and the SPARQL query language,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which most bench biologists may not have. To alleviate this challenge, we are currently building a graphical SPARQL query interface, which will intuitively enable query formulation through the navigation of the knowledge graph schema. We are also exploring the application of ProKinO for machine learning-based knowledge discovery and hypotheses generation.

DATA AVAILABILITY

The protein kinase ontology (ProKinO)'s latest OWL file, along with previous versions, is publicly available at http://prokino.uga.edu/downloads.html. Future versions of the ontology also will be placed at the same address. Also, the ontology's browser is accessible at https://prokino.uga.edu/browser. Users can save results of queries in diagrams or other formats such as CSV.

ACKNOWLEDGEMENTS

We thank members of the NK lab for feedback on the ProKinO browser and usage.

FUNDING

Funding from NIH (U01CA239106) is acknowledged

CONFLICT OF INTREST

Conflict of interest statement. None declared.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. Ferguson, F.M. and Gray, N.S. (2018) Kinase inhibitors: the road ahead. *Nat Rev Drug Discov*, **17**, 353-377.
2. Zhang, J., Yang, P.L. and Gray, N.S. (2009) Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*, **9**, 28-39.
3. Nguyen, D.T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A., Holmes, J., Jensen, L.J., Karlsson, A. *et al.* (2017) Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res*, **45**, D995-D1002.
4. Byrne, D.P., Foulkes, D.M. and Eyers, P.A. (2017) Pseudokinases: update on their functions and evaluation as new drug targets. *Future Med Chem*, **9**, 245-265.
5. Eyers, P.A., Keeshan, K. and Kannan, N. (2017) Tribbles in the 21st Century: The Evolving Roles of Tribbles Pseudokinases in Biology and Disease. *Trends Cell Biol*, **27**, 284-298.
6. Eyers, P.A. and Murphy, J.M. (2013) Dawn of the dead: protein pseudokinases signal new adventures in cell biology. *Biochem Soc Trans*, **41**, 969-974.
7. Murphy, J.M., Mace, P.D. and Eyers, P.A. (2017) Live and let die: insights into pseudoenzyme mechanisms from structure. *Curr Opin Struct Biol*, **47**, 95-104.
8. Daniel M Foulkes1, D.P.B., Wayland Yeung2, Safal Shrestha2, Fiona P Bailey1, Samantha Ferries1,3, Claire E Eyers1,3, Karen Keeshan4, Carrow Wells5, David H Drewry5, William J Zuercher5,6 Natarajan Kannan2 and Patrick A Eyers1. (2018) Covalent EGFR/HER2 kinase inhibitors induce cellular degradation of human Tribbles 2 (TRIB2) pseudokinase. *Science Signaling*.
9. Sheils, T., Mathias, S.L., Siramshetty, V.B., Bocci, G., Bologa, C.G., Yang, J.J., Waller, A., Southall, N., Nguyen, D.T. and Oprea, T.I. (2020) How to Illuminate the Druggable Genome Using Pharos. *Curr Protoc Bioinformatics*, **69**, e92.
10. Sheils, T.K., Mathias, S.L., Kelleher, K.J., Siramshetty, V.B., Nguyen, D.T., Bologa, C.G., Jensen, L.J., Vidović, D., Koleti, A., Schürer, S.C. *et al.* (2021) TCRD and Pharos 2021: mining the human proteome for disease biology. *Nucleic Acids Res*, **49**, D1334-d1346.
11. Berginski, M.E., Moret, N., Liu, C., Goldfarb, D., Sorger, P.K. and Gomez, S.M. (2021) The Dark Kinase Knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Res*, **49**, D529-d535.
12. Moret, N., Liu, C., Gyori, B.M., Bachman, J.A., Steppi, A., Hug, C., Taujale, R., Huang, L.-C., Berginski, M.E. and Gomez, S.M. (2021) A resource for exploring the understudied human kinome for research and therapeutic opportunities. *BioRxiv*, 2020.2004. 2002.022277.
13. McSkimming, D.I., Dastgheib, S., Talevich, E., Narayanan, A., Katiyar, S., Taylor, S.S., Kochut, K. and Kannan, N. (2015) ProKinO: a unified resource for mining the cancer kinome. *Hum Mutat*, **36**, 175-186.
14. Gosal, G., Kannan, N. and Kochut, K. (2011) ProKinO: A framework for protein kinase ontology. *Proceedings of the IEEE International Conference on Bioinformatics & Biomedicine, Atlanta, Georgia*, 550-555.
15. Gosal, G., Kochut, K.J. and Kannan, N. (2011) ProKinO: an ontology for integrative analysis of protein kinases in cancer. *PLoS One*, **6**, e28782.
16. McSkimming, D.I., Dastgheib, S., Baffi, T.R., Byrne, D.P., Ferries, S., Scott, S.T., Newton, A.C., Eyers, C.E., Kochut, K.J., Eyers, P.A. *et al.* (2016) KinView: a visual comparative sequence analysis tool for integrated kinome research. *Mol Biosyst*.
17. Liu, D., Liang, X.C. and Zhang, H. (2016) Culturing Schwann Cells from Neonatal Rats by Improved Enzyme Digestion Combined with Explants-culture Method. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao*, **38**, 388-392.
18. Cicenas, J. and Cicenas, E. (2016) Multi-kinase inhibitors, AURKs and cancer. *Med Oncol*, **33**, 43.
19. Mohanty, S., Oruganty, K., Kwon, A., Byrne, D.P., Ferries, S., Ruan, Z., Hanold, L.E., Katiyar, S., Kennedy, E.J., Eyers, P.A. *et al.* (2016) Hydrophobic Core Variations Provide a Structural Framework for Tyrosine Kinase Evolution and Functional Specialization. *PLoS Genet*, **12**, e1005885.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Vazquez, M., Pons, T., Brunak, S., Valencia, A. and Izarzugaza, J.M. (2016) wKinMut-2: Identification and Interpretation of Pathogenic Variants in Human Protein Kinases. *Hum Mutat*, **37**, 36-42.
21. Ruan, Z. and Kannan, N. (2015) Mechanistic Insights into R776H Mediated Activation of Epidermal Growth Factor Receptor Kinase. *Biochemistry*, **54**, 4216-4225.
22. Taylor, S.S., Shaw, A.S., Kannan, N. and Kornev, A.P. (2015) Integration of signaling in the kinome: Architecture and regulation of the alphaC Helix. *Biochim Biophys Acta*, **1854**, 1567-1574.
23. Nguyen, T., Ruan, Z., Oruganty, K. and Kannan, N. (2015) Co-conserved MAPK features couple D-domain docking groove to distal allosteric sites via the C-terminal flanking tail. *PLoS One*, **10**, e0119636.
24. Bailey, F.P., Byrne, D.P., McSkimming, D., Kannan, N. and Eyers, P.A. (2015) Going for broke: targeting the human cancer pseudokinome1. *Biochem J*, **465**, 195-211.
25. Simonetti, F.L., Tornador, C., Nabau-Moreto, N., Molina-Vila, M.A. and Marino-Buslje, C. (2014) Kin-Driver: a database of driver mutations in protein kinases. *Database (Oxford)*, **2014**, bau104.
26. Hu, J., Ahuja, L.G., Meharena, H.S., Kannan, N., Kornev, A.P., Taylor, S.S. and Shaw, A.S. (2015) Kinase regulation by hydrophobic spine assembly in cancer. *Mol Cell Biol*, **35**, 264-276.
27. McClendon, C.L., Kornev, A.P., Gilson, M.K. and Taylor, S.S. (2014) Dynamic architecture of a protein kinase. *Proc Natl Acad Sci U S A*, **111**, E4623-4631.
28. U, M., Talevich, E., Katiyar, S., Rasheed, K. and Kannan, N. (2014) Prediction and prioritization of rare oncogenic mutations in the cancer Kinome using novel features and multiple classifiers. *PLoS Comput Biol*, **10**, e1003545.
29. Goldberg, J.M., Griggs, A.D., Smith, J.L., Haas, B.J., Wortman, J.R. and Zeng, Q. (2013) Kinannote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. *Bioinformatics*, **29**, 2387-2394.
30. Oruganty, K. and Kannan, N. (2013) Evolutionary variation and adaptation in a conserved protein kinase allosteric network: Implications for inhibitor design. *Biochim Biophys Acta*.
31. Meharena, H.S., Chang, P., Keshwani, M.M., Oruganty, K., Nene, A.K., Kannan, N., Taylor, S.S. and Kornev, A.P. (2013) Deciphering the structural basis of eukaryotic protein kinase regulation. *PLoS Biol*, **11**, e1001680.
32. McSkimming, D.I., Dastgheib, S., Talevich, E., Narayanan, A., Katiyar, S., Taylor, S.S., Kochut, K. and Kannan, N. (2014) ProKinO: A Unified Resource for Mining the Cancer Kinome. *Hum Mutat*.
33. Ruan, Z., Katiyar, S. and Kannan, N. (2017) Computational and Experimental Characterization of Patient Derived Mutations Reveal an Unusual Mode of Regulatory Spine Assembly and Drug Sensitivity in EGFR Kinase. *Biochemistry*, **56**, 22-32.
34. Lubner, J.M., Dodge-Kafka, K.L., Carlson, C.R., Church, G.M., Chou, M.F. and Schwartz, D. (2017) Cushing's syndrome mutant PKA(L)(205R) exhibits altered substrate specificity. *FEBS Lett*, **591**, 459-467.
35. Patani, H., Bunney, T.D., Thiyagarajan, N., Norman, R.A., Ogg, D., Breed, J., Ashford, P., Potterton, A., Edwards, M., Williams, S.V. *et al.* (2016) Landscape of activating cancer mutations in FGFR kinases and their differential responses to inhibitors in clinical use. *Oncotarget*, **7**, 24252-24268.
36. Huang, L.C., Ross, K.E., Baffi, T.R., Drabkin, H., Kochut, K.J., Ruan, Z., D'Eustachio, P., McSkimming, D., Arighi, C., Chen, C. *et al.* (2018) Integrative annotation and knowledge discovery of kinase post-translational modifications and cancer-associated mutations through federated protein ontologies and resources. *Sci Rep*, **8**, 6518.
37. Yun, C.-H., Mengwasser, K.E., Toms, A.V., Woo, M.S., Greulich, H., Wong, K.-K., Meyerson, M. and Eck, M.J. (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences*, **105**, 2070-2075.
38. Gajiwala, K.S., Wu, J.C., Christensen, J., Deshmukh, G.D., Diehl, W., DiNitto, J.P., English, J.M., Greig, M.J., He, Y.-A. and Jacques, S.L. (2009) KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. *Proceedings of the National Academy of Sciences*, **106**, 1542-1547.
39. Kim, L.C., Song, L. and Haura, E.B. (2009) Src kinases as therapeutic targets for cancer. *Nature Reviews Clinical Oncology*, **6**, 587-595.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. Benhar, M., Engelberg, D. and Levitzki, A. (2002) ROS, stress-activated kinases and stress signaling in cancer. *EMBO reports*, **3**, 420-425.
41. Duncan, J.S., Whittle, M.C., Nakamura, K., Abell, A.N., Midland, A.A., Zawistowski, J.S., Johnson, N.L., Granger, D.A., Jordan, N.V., Darr, D.B. *et al.* (2012) Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell*, **149**, 307-321.
42. Niepel, M., Hafner, M., Duan, Q., Wang, Z., Paull, E.O., Chung, M., Lu, X., Stuart, J.M., Golub, T.R., Subramanian, A. *et al.* (2017) Common and cell-type specific responses to anti-cancer drugs revealed by high throughput transcript profiling. *Nat Commun*, **8**, 1186.
43. Erika, G., Federica, Z., Martina, S., Anselmo, P., Luigi, R., Marina, M., Davide, C., Eleonora, Z., Monica, V. and Silverio, T. (2016) Old Tyrosine Kinase Inhibitors and Newcomers in Gastrointestinal Cancer Treatment. *Curr Cancer Drug Targets*, **16**, 175-185.
44. Li, Y.Y. and Jones, S.J. (2012) Drug repositioning for personalized medicine. *Genome Med*, **4**, 27.
45. Joensuu, H., Roberts, P.J., Sarlomo-Rikala, M., Andersson, L.C., Tervahartiala, P., Tuveson, D., Silberman, S., Capdeville, R., Dimitrijevic, S., Druker, B. *et al.* (2001) Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N Engl J Med*, **344**, 1052-1056.
46. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.
47. Huang, L.C., Taujale, R., Gravel, N., Venkat, A., Yeung, W., Byrne, D.P., Eyers, P.A. and Kannan, N. (2021) KinOrtho: a method for mapping human kinase orthologs across the tree of life and illuminating understudied kinases. *BMC Bioinformatics*, **22**, 446.
48. Kwon, A., Scott, S., Taujale, R., Yeung, W., Kochut, K.J., Eyers, P.A. and Kannan, N. (2019) Tracing the origin and evolution of pseudokinases across the tree of life. *Sci Signal*, **12**.
49. Hanks, S.K. and Hunter, T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *Faseb j*, **9**, 576-596.
50. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912-1934.
51. Roskoski, R., Jr. (2016) Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol Res*, **103**, 26-48.
52. Kanev, G.K., de Graaf, C., Westerman, B.A., de Esch, I.J.P. and Kooistra, A.J. (2021) KLIFS: an overhaul after the first 5 years of supporting kinase research. *Nucleic Acids Res*, **49**, D562-d569.
53. Faezov, B. and Dunbrack, R.L., Jr. (2021) PDBrenum: A webserver and program providing Protein Data Bank files renumbered according to their UniProt sequences. *PLoS One*, **16**, e0253411.
54. Neuwald, A.F. (2009) Rapid detection, classification and accurate alignment of up to a million or more related protein sequences. *Bioinformatics*, **25**, 1869-1875.
55. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R. *et al.* (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*, **41**, D955-961.
56. Huang, L.C., Yeung, W., Wang, Y., Cheng, H., Venkat, A., Li, S., Ma, P., Rasheed, K. and Kannan, N. (2020) Quantitative Structure-Mutation-Activity Relationship Tests (QSMART) model for protein kinase inhibitor response prediction. *BMC Bioinformatics*, **21**, 520.
57. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*, **46**, D1074-d1082.
58. Li, Y.H., Yu, C.Y., Li, X.X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G. *et al.* (2018) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res*, **46**, D1121-d1127.
59. Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D.J., Turner, J.P., Vidovic, D., Forlin, M., Kelley, T.T., D'Urso, A. *et al.* (2018) Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res*, **46**, D558-d566.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.25.482021; this version posted March 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

60. Lin, Y., Mehta, S., Küçük-McGinty, H., Turner, J.P., Vidovic, D., Forlin, M., Koleti, A., Nguyen, D.T., Jensen, L.J., Guha, R. *et al.* (2017) Drug target ontology to classify and integrate drug discovery data. *J Biomed Semantics*, **8**, 50.
61. Bühlmann, S. and Reymond, J.L. (2020) ChEMBL-Likeness Score and Database GDBChEMBL. *Front Chem*, **8**, 46.
62. Avram, S., Bologa, C.G., Holmes, J., Bocci, G., Wilson, T.B., Nguyen, D.T., Curpan, R., Halip, L., Bora, A., Yang, J.J. *et al.* (2021) DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Res*, **49**, D1160-d1169.
63. Huo, F.C., Pan, Y.J., Li, T.T., Mou, J. and Pei, D.S. (2019) PAK5 promotes the migration and invasion of cervical cancer cells by phosphorylating SATB1. *Cell Death Differ*, **26**, 994-1006.
64. Han, K., Zhou, Y., Tseng, K.F., Hu, H., Li, K., Wang, Y., Gan, Z., Lin, S., Sun, Y. and Min, D. (2018) PAK5 overexpression is associated with lung metastasis in osteosarcoma. *Oncol Lett*, **15**, 2202-2210.
65. Fang, Z.P., Jiang, B.G., Gu, X.F., Zhao, B., Ge, R.L. and Zhang, F.B. (2014) P21-activated kinase 5 plays essential roles in the proliferation and tumorigenicity of human hepatocellular carcinoma. *Acta Pharmacol Sin*, **35**, 82-88.
66. Zhang, Y.C., Huo, F.C., Wei, L.L., Gong, C.C., Pan, Y.J., Mou, J. and Pei, D.S. (2017) PAK5-mediated phosphorylation and nuclear translocation of NF-κB-p65 promotes breast cancer cell proliferation in vitro and in vivo. *J Exp Clin Cancer Res*, **36**, 146.
67. Quan, L., Cheng, Z., Dai, Y., Jiao, Y., Shi, J. and Fu, L. (2020) Prognostic significance of PAK family kinases in acute myeloid leukemia. *Cancer Gene Ther*, **27**, 30-37.
68. Huse, M. and Kuriyan, J. (2002) The conformational plasticity of protein kinases. *Cell*, **109**, 275-282.
69. Kornev, A.P. and Taylor, S.S. (2015) Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem Sci*, **40**, 628-647.
70. Oruganty, K. and Kannan, N. (2012) Design principles underpinning the regulatory diversity of protein kinases. *Philos Trans R Soc Lond B Biol Sci*, **367**, 2529-2539.
71. Yonemoto, W., Garrod, S.M., Bell, S.M. and Taylor, S.S. (1993) Identification of phosphorylation sites in the recombinant catalytic subunit of cAMP-dependent protein kinase. *J Biol Chem*, **268**, 18626-18632.
72. Knighton, D.R., Zheng, J.H., Ten Eyck, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S. and Sowadski, J.M. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, **253**, 407-414.
73. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A. *et al.* (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, **596**, 590-596.