Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

ProKinO: A Unified Resource for Mining the Cancer Kinome

Daniel Ian McSkimming,[1] Shima Dastgheib,[2] Eric Talevich,[3] Anish Narayanan,[4] Samiksha Katiyar,[4] Susan S. Taylor,[5] Krys Kochut,[2] and Natarajan Kannan[1,4]*

[1] Institute of Bioinformatics, University of Georgia, Athens, Georgia; [2] Department of Computer Science, University of Georgia, Athens, Georgia; [3] Department of Dermatology, UCSF School of Medicine, San Francisco, California; [4] Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia; [5] Department of Chemistry and Biochemistry, Department of Pharmacology, UCSD, La Jolla, California

Communicated by Peter N. Robinson

Received 24 June 2014; accepted revised manuscript 21 October 2014.
Published online 10 November 2014 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22726

ABSTRACT: Protein kinases represent a large and diverse family of evolutionarily related proteins that are abnormally regulated in human cancers. Although genome sequencing studies have revealed thousands of variants in protein kinases, translating "big" genomic data into biological knowledge remains a challenge. Here, we describe an ontological framework for integrating and conceptualizing diverse forms of information related to kinase activation and regulatory mechanisms in a machine readable, human understandable form. We demonstrate the utility of this framework in analyzing the cancer kinome, and in generating testable hypotheses for experimental studies. Through the iterative process of aggregate ontology querying, hypothesis generation and experimental validation, we identify a novel mutational hotspot in the αC-β4 loop of the kinase domain and demonstrate the functional impact of the identified variants in epidermal growth factor receptor (EGFR) constitutive activity and inhibitor sensitivity. We provide a unified resource for the kinase and cancer community, ProKinO, housed at http://vulcan.cs.uga.edu/prokino.

Hum Mutat 36:175–186, 2015. Published 2014 Wiley Periodicals, Inc.**

KEY WORDS: personalized medicine; cancer therapy; database; drug discovery; big data; disease; mutation; resistance; kinase; conformation; regulation

Introduction

Cancer is a family of diseases characterized by the accumulation of variants in a subset of genes that confer a growth and survival advantage to the cell. The 518 protein kinase genes in the human genome (collectively called the kinome [Manning et al., 2002b]) represent one of the largest families of genes that are mutationally activated or repressed in human cancers [Futreal et al., 2004; Lahiry et al., 2010; Brognard and Hunter, 2011]. Many known cancer-associated variants occur in the conserved protein kinase domain, which catalyzes phosphorylation [Knighton et al., 1991; Zheng et al., 1993] and provides regulation of complex signal transduction networks. The prominent roles protein kinases play in cancer initiation and progression have contributed to extensive studies on these proteins and, consequently, a wealth of data resulting from several "omic" efforts. Sequencing of the kinome from nearly 218 cancer types have resulted in over 17,000 non-synonymous variants in the protein kinase domain [Parthiban et al., 2006; Li et al., 2009; Sim et al., 2012; Worth et al., 2011]. Likewise, drug discovery efforts have resulted in several United States Food and Drug Administration approved drugs to target the cancer kinome, including the block buster drug imatinib [Deininger et al., 2005; Simpson et al., 2009], which targets the Abelson tyrosine kinase in chronic myeloid leukemia [Izarzugaza et al., 2012]. More recently, the kinome has been the focus of several proteomic efforts to map the signaling networks altered in cancer and drug-resistant states [Kannan et al., 2007; Manning et al., 2002a; Hashimoto et al., 2012; Dixit and Verkhivker, 2014]. In addition, numerous investigator-initiated structural and comparative genomics studies have revealed the sequence and structural basis for protein kinase evolution [Hanks and Hunter, 1995; Manning et al., 2002b; Kannan et al., 2007; Lim and Pawson, 2010; Oruganty and Kannan, 2012] and regulation [Huse and Kuriyan, 2002; Taylor et al., 2013]. These efforts have resulted in massive amounts of data that can potentially be used to accelerate the functional characterization of the cancer kinome by providing new testable hypotheses for experimental studies. In particular, distinguishing the causative "driver" mutations from the large number of harmless "passenger" mutations requires many hypotheses to be formulated and tested based on integrative analysis of existing data. However, the complex and disparate nature of protein kinase data sets, and the difficulties in integrating and analyzing large, complex datasets have hindered progress.

Information on kinase cancer variants is stored in various sources such as COSMIC [Forbes et al., 2008], KinMutBase [Ortutay et al., 2005], and cancer bioportals [Cerami et al., 2012; Cline et al., 2013]. Likewise, information on the structural and functional aspects of kinases is buried in the literature and scattered across diverse databases. Consequently, to answer simple questions such as "Which kinases have variants in the ATP or drug binding pocket?" or "Which pathways are altered by mutated kinases in cancer?", researchers must go through the time consuming and error prone process of collecting information from disparate sources and data

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Natarajan Kannan, Department of Biochemistry and Molecular Biology, Institute of Bioinformatics, University of Georgia, Athens, GA 30602. E-mail: kannan@bmb.uga.edu

Contract grant sponsors: National Science Foundation (MCB-1149106); American Cancer Society (RSG-10-188-01-TBE); Georgia Research Alliance (1021RR166828); Summer Provost Funds from the University of Georgia.

formats, and post-processing the data through customized scripts and programs. This poses major challenges for bench biologists who do not have the resources or training to write customized scripts for post-processing. Moreover, writing customized software often leads to duplication of efforts across laboratories and does not scale well with the growing complexity and diversity of biological data.

Bio-ontologies, such as the Gene Ontology [Ashburner et al., 2000], have served as a vehicle of knowledge for the biological community for nearly two decades and provide a framework for integrating data in ways computers can read and humans can understand. To address the data integration challenge in the protein kinase field, we previously reported the Protein Kinase Ontology (ProKinO, http://vulcan.cs.uga.edu/prokino) [Gosal et al., 2011a; Gosal et al., 2011b], which provides a controlled vocabulary of terms and relations linking data on protein kinase sequence, structure, function, pathway and disease. Here, we expand the scope of ProKinO by conceptualizing information on conserved sequence and structural motifs that contribute to protein kinase allosteric regulation. We show that conceptualizing existing knowledge on kinase regulatory motifs in a machine readable format provides useful context for predicting variant impact, allows rapid comparisons of protein kinase sequences and structures across the kinome and enables hypothesis generation and reasoning over existing data. Furthermore, through iterative ontology querying, reasoning and experimental studies, we identify a novel mutational hotspot in the kinase domain and demonstrate the functional significance of the predicted mutations on epidermal growth factor receptor (EGFR) activation and drug sensitivity.

Methods

Nomenclature

The protein symbol and variant nomenclature used is in accordance with the Human Genome Variation Society (HGVS, http://www.hgvs.org) and the HUGO Gene Nomenclature Committee (HGNC, http://www.genenames.org), with one exception: we initially identify Protein Kinase A (PKA) with both the HGNC approved name (PRKCA) and the common abbreviation PKA, but subsequently use the common abbreviation alone. When referring to specific residue positions and mutations, we generally use the PKA numbering. However, in cases where the native protein numbering is specified, we indicate the equivalent PKA numbering as superscript.

Sequence Alignment Methods

Protein kinase sequences were aligned using the MAPGAPS program [Neuwald, 2009], as described previously [Talevich and Kannan, 2013]. The prototypic protein kinase A (PRKCA, PKA) sequence was used as the frame of reference for mapping equivalent residue positions in aligned kinase sequences. By considering the PKA equivalent position of a residue, we can identify and analyze interactions concerning residues in structural and sequence motifs across the kinome. This serves as an important starting point for providing structural and functional context for disease variants.

Modification of ProKinO Schema

To conceptualize information on kinase structural motifs and to provide context for cancer variants, we modified the ProKinO schema to add new properties to two classes, namely the *Mutation* class and *Motif* class. Two subclasses of *Motif* were also created, named *Sequence Motif* and *Structural Motif* (Fig. 1). Instances of the *Sequence Motif* class represent important contiguous regions of kinase sequence, such as the twelve Hanks & Hunter subdomains [Hanks and Hunter, 1995], ATP binding pocket, gatekeeper position, C-helix, activation loop, DFG motif, HRD motif, G-loop, R-spine, C-spine, and so on that are commonly used to describe protein kinase structures. Each instance is assigned associated properties, such as the start and end location of the motif in both the native sequence and with respect to PKA. The *Structural Motif* class contains representations of spatial motifs formed by conserved residues that interact in three-dimensional structures. We implemented each instance of the *Structural Motif* class as a collection of *Sequence Motif* instances, linked using the *contains* relation. These classes were then linked to other classes in ProKinO using the relations shown in Figure 1. Specifically, the *Mutation* class was linked to the *Disease* and *Sequence* classes using the *implicatedIn* and *occursIn* relations, respectively. Likewise, the *Mutation* class was linked to *Motif* class using the *locatedIn* relationship. New instances were added to the *Mutation* and *Motif* classes to capture information on kinase structural motifs and subdomains.

Methods Related to Ontology Population and Instantiation

ProKinO is automatically populated from different data sources including Kinbase, UniProt, COSMIC, Reactome, and manually curated kinome alignments. The development steps of the software have been previously published [Gosal et al., 2011a]. We made significant changes to the previous version of ProKinO, particularly to address the conceptualization of samples and motifs. Instances of some classes, such as the *Mutation* and *Motif* classes, store positional information such as residue numbers within the native protein sequence. To represent the kinome sequence alignment in ProKinO, we created two additional properties in these instances: *hasPKAStartLocation* and *hasPKAEndLocation;* these store the aligned start and end residue positions with respect to PKA. We added the *Sample* class and instantiated it using the data file provided by COSMIC, the same file that is used to instantiate the *Mutation* class. Each row in the file represents an instance (individual) of *Sample* in which an instance of the *Mutation* class is observed. The *Mutation* and *Sample* classes are connected by the *inSample* relation. For example, the instance of variant p.D549E^{184PKA} is connected to *Sample-E35170* with the *inSample* relation (Fig. 1). In addition, we replaced the *Subdomain* class used in the previous version of ProKinO with the *Motif* class and introduced its two subclasses: *Sequence Motif* and *Structural Motif.* The latter is connected to the former using the "*contains*" relation and both inherit their parent relations. For example, KE Salt Bridge (an instance of *Structural Motif*) *contains* β3-lysine (an instance of *Sequence Motif*). The instances of these two classes and the relations between them are generated from the multiple sequence alignments of the human kinome.

Semantic Querying and Methods for Post-Processing Query Results

SPARQL is the W3C [Prud and Seaborne, 2006] recommendation for querying datasets structured according to the Resource Description Framework (RDF) [Lassila and Swick, 1998]. It can also be used to query ontologies represented using the Web Ontology Language (OWL), the language we use to represent ProKinO. Data in RDF and OWL ontologies is represented as statements in the form of subject-predicate-object triples (e.g., "EGFR *hasMutation*

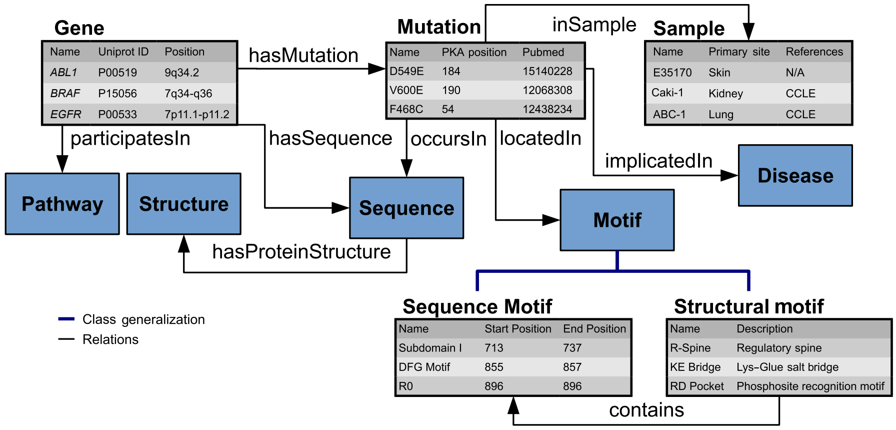

Figure 1. Schematic representation of protein kinase data and knowledge in ProKinO. Boxes denote ontology classes, with arrows showing relations between classes. Several classes (e.g., Gene, Motif, Sample) further show the attributes they store (e.g., Name, Primary Site, References) as well as examples of individual data instances (e.g., BRAF, p.V600E, DFG Motif).

p.T790M”). The core syntax of SPARQL is a set of triple patterns, similar to an RDF triple except that any component in the triple pattern can be a variable. For example, “EGFR *hasMutation* ?mutation” is a SPARQL pattern which queries for all triples describing a variant in EGFR (e.g., “EGFR *hasMutation* p.T790M”). Triple patterns may be combined into conjunctions and disjunctions (optional patterns are possible, as well). SPARQL querying provides us with a method for extracting information from ProKinO and returning the requested data in a comma separated values (csv) file. Charts and graphs were generated from query results using Python v2.7 and the ReportLab graphing library [Sanner, 1999]. Protein structure images were created using PyMol [DeLano, 2002], whereas word cloud images were generated using Wordle [Feinberg, c2001].

Protein Expression and Immunoblotting

Full-length GFP-EGFR (wild type [WT]) was used to generate point variants, p.R748K, p.R776C, p.R776H, p.R831C, p.R831H, p.R831L, p.R832C, p.R832H, p.R836C, and p.R841K, using QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Inc., Santa Clara, CA, USA). Point variants were confirmed by DNA sequencing. Plasmids (1 μg/μl), purified by Maxi-prep kit (Qiagen, Venlo, Limburg, NLD), were used to transiently transfect CHO cells. CHO cells were cultured in Dulbecco’s modified Eagle’s medium containing 10% fetal bovine serum and were plated at a density of 3×10^5 cells in 60 mm plates. Transfection was performed using lipfectamine-2000 (Invitrogen, Waltham, MA, USA) according to manufacturer’s protocol. Transfected cells were allowed to grow for 24 hr followed by serum starvation in Ham’s F12 media for 18 hr.

To detect autophosphorylation of EGFR, cells were stimulated with 100 ng/ml EGF (Sigma-Aldrich, St. Louis, MO, USA) for 5 min. Cells were washed with PBS and lysed in buffer containing 50 mM Tris HCl, pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM EDTA, 1% TritonX-100, and protease inhibitor cocktail (Millipore, Billerica, MA, USA). Cell lysates were centrifuged at 1000 rpm for 5 min and total proteins were resolved on SDS-PAGE and transferred on PVDF membrane for Western blot analysis. Total EGFR level was detected by GFP and autophosphorylation was analyzed using p.Y845, p.Y992, p.Y1045, p.Y1068, and p.Y1173 antibodies (Cell Signaling, Danvers, MA, USA). The effect of WT and mutant EGFR on phosphorylation of the downstream signaling protein STAT3 was monitored using pSTAT3 and the total protein amount was detected with STAT3 antibodies (Cell Signaling, Danvers, MA, USA).

Gefitinib Treatment

To monitor the effect of gefitinib on WT and mutant (p.R776H^{105PKA}) EGFR, CHO cells were cultured, transfected, and starved as described above. Before stimulation with EGF, cells were treated with 0, 0.001, 0.01, 0.1, 1.0, and 10 μM of gefitinib for 1 hr in Ham’s F12 media. After 1 hr, stimulation was performed for 5 min by adding 100 ng/ml EGF in the media already on the plates. Cell lysates were processed as described above. Total and phosphorylated proteins were analyzed as indicated.

Ontology Verification

To ensure the accuracy of the data presented in this article and in ProKinO as a whole, we took various measures to validate that the populated ontology is consistent with its underlying sources. We implemented a manual validation process on a randomly selected subset (1%) of kinases. For each selected kinase, the associated ontology data was collected using the ProKinO browser (http://vulcan.cs.uga.edu/prokino) and cross-checked with the appropriate parent sources. For example, data in the *Mutation*, *Disease*, and *Sample* classes is sourced from COSMIC. From a mutation instance in the ProKinO browser, we followed the link to the originating COSMIC record and validated the specific data. To ensure that no data is missing, we searched COSMIC by gene name and verified that the number of records returned matches the number of variants for that gene in ProKinO. The ontology was validated against other data sources in a similar manner.

Next, we validated our multiple sequence alignment of the human kinome. We visually inspected the alignment to verify that key sequence motifs (e.g., HRD and DGF motifs) and core hydrophobic residues are aligned correctly. For variants mentioned in this article, we performed structural alignments (when crystal structures were

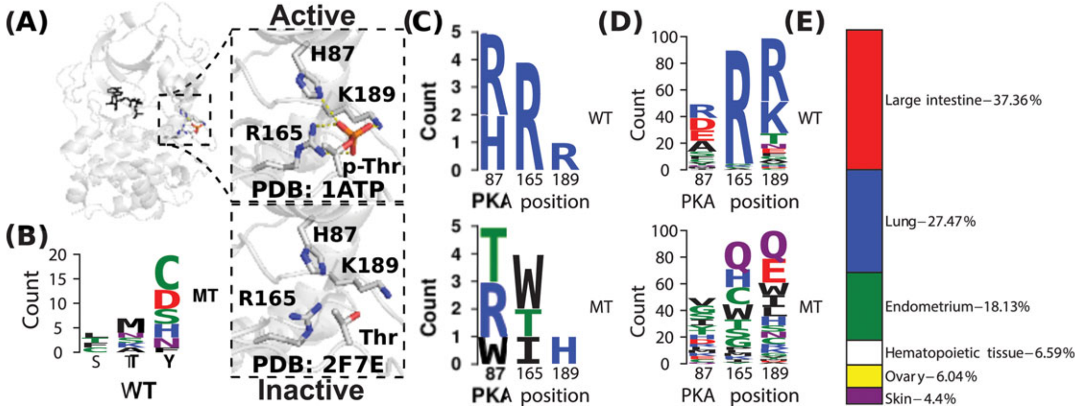

Figure 2. RD Pocket variants. **A**: Phosphorylated threonine interacting with basic residues from the RD pocket (PDB:1ATP). Lack of coordination of RD pocket residues when the activation loop is not phosphorylated (PDB:2F7E). **B**: Variants at activation loop phosphosites. **C**: Variants in the *canonical* RD pocket kinases shown as amino acid logos in which the size of the letter is proportional to the frequency of occurrence of the corresponding amino acid. The top panel shows amino acid counts in wild type (WT) proteins and the bottom panel shows the mutant (MT) forms. **D**: Variants in the *non-canonical* pocket kinases, shown as described in **C**. **E**: Sample primary sites with variants in the canonical RD pocket positions.

available) to verify structurally equivalent positions. Finally, to rigorously ensure the accuracy of ProKinO data on a large scale, we developed a suite of test applications. These applications automatically compare the data in the ontology with the corresponding original data sources for consistency. The absence of disparities between output files produced by the test applications indicates consistency between ProKinO and original data sources.

Results and Discussion

Conceptualizing Kinase Sequence and Structural Motifs in ProKinO Provides a Framework for Knowledge-Based Mining of Cancer Variants

Several conserved structural motifs associated with protein kinase activation and regulation, such as the lysine glutamate (KE) salt bridge, hydrophobic spine, and RD pocket, have been identified through detailed structural studies on PKA [Taylor et al., 2004; Taylor et al., 2013; Kornev and Taylor, 2010] and related members of the protein kinase super-family [Johnson et al., 1996; Sicheri et al., 1997]. Although these motifs are widely used to compare protein kinase structures and explain kinase activation mechanisms, they have not been systematically used to predict variant impact because knowledge on kinase structural motifs is buried in the literature and not represented in a machine-readable format. Inconsistencies in residue numbering between sequence data sources further complicates mapping of variants to crystal structures and comparisons across the kinome. To address these issues, we have developed a consistent numbering scheme (see *Methods*) using the prototypic PKA as the frame of reference. Furthermore, we have introduced new concepts, relations, and instances in ProKinO to represent protein kinase structural knowledge using the same semantics and terminologies used in the literature (Fig. 1; *Methods*). For example, the *Motif* class captures knowledge on the sequence and structural motifs associated with kinase functions, while the *locatedIn* relation between the *Motif* and *Mutation* classes captures the information linking variants to sequence and structural motifs (Fig. 1). Such conceptual representation of knowledge in a machine-readable ontology enables integrative analysis of existing data in ways not possible through other resources. Complex aggregate queries relating cancer variants to kinase structural motifs can be rapidly performed using the ontology, while performing the same queries otherwise will require the user to first retrieve data from various sources such as PDB, COSMIC, and UniProt, and post-process data using customized scripts. Below, we demonstrate the utility of ProKinO in cancer kinome mining and annotation using the knowledge conceptualized on conserved motifs associated with kinase function and regulation. We use the PKA residue numbering throughout while referring to residues and variants in conserved motifs, unless otherwise noted.

Identification of Variants in the RD Pocket and Predicted Impact

The canonical RD pocket is a structural motif formed by basic residues from three regulatory regions of the kinase domain: the C-helix (p.H87), the HRD motif in the catalytic loop (p.R165) and the activation loop (p.R189) (Fig. 2A). The RD pocket concept is widely used in the literature to explain the structural basis of activation loop phosphorylation, a mode of regulation utilized by many kinases [Johnson et al., 1996]. The negatively charged phosphate group of a phosphorylated serine, threonine or tyrosine residue in the activation loop coordinates with the positively charged residues in the RD pocket. This coordination provides a framework for allosteric regulation by positioning key functional elements, such as the C-helix and activation loop, in a catalytically competent conformation [Jeffrey et al., 1995; Russo et al., 1996; Yamaguchi and Hendrickson, 1996].

While most serine/threonine kinases use the canonical RD pocket residues to coordinate with the activation loop phosphate, non-canonical pockets have been described in which the pocket residues emanate from different structural locations. In the JAK2 JH1 domain (PDB: 3E63), for example, the canonical RD pocket residues are not basic even though JAK2 is regulated by activation loop phosphorylation. However, JAK2 conserves three lysines, two in the activation loop (p.K1005^{145PKA}, p.K1009^{148PKA}) and one N-terminal of the F-helix (p.K1030^{169PKA}), that coordinate with the phosphorylated tyrosine residues (p.Y1007^{147PKA}) in the activation loop [Lucet et al., 2006]. For our analysis, we classified RD pockets into canonical and non-canonical based on the nature of amino acids observed

Table 1. A Subset of RD-Pocket Variants Shown with Related Pathway and Reaction Data

Wild type	Position	Mutant type	PKA position	Gene	Pathway	Reaction
E	282	G/K	87	ABL1	Signaling by Robo receptor	Phosphorylation of Rob1 by Abl kinase
R	1,275	L/Q	189	ALK		
T	599	I	189	BRAF	NGF signaling via TRKA from the plasma membrane	(Frs2)Rap1-GTP binds to and activates B-Raf
R	520	P/Q	165	BTK	DAP12 signaling	Phosphorylation and activation of VAV2/VAV3 by SYK
R	544	W	189		Fc epsilon receptor (FCERI) signaling	Phosphorylation of TEC kinases by p-SYK
R	346	C/G/H/S	165	CHEK2	p53-independent G1/S DNA damage checkpoint	Phosphorylation of Cdc25A at Ser-123 by Chk2
K	386	E/N/R	189	CLIK1	–	–
E	758	D/G	87	EGFR	GAB1 signalosome	Gab1 phosphorylation by EGFR kinase
R	836	C/H/S	165			
K	860	E/I	189			
R	769	H/S	189	EPHA3	–	–
R	743	L/Q/W/C	165	EPHB1	–	–
R	745	C/P	165	EPHB2	Axon guidance	Phosphorylation of L1 by EPHB2
R	180	Q/W	165	GSK3B	Axon guidance	Phosphorylation of CRMPs by GSK3beta
R	136	G/S	165	HPK1		
D	894	E/G	87	JAK2	Signaling by Leptin	JAK2 phosphorylates LEPR
A	636	V	87	KIT	Signaling by SCF-KIT	Phosphorylation of JAK2
K	175	M	165	LKB1	PI3K cascade	Activation of cytosolic AMPK by phosphorylation
E	199	K/Q	189			
R	162	C/H	87	MAP2K7	Cellular responses to stress	Phosphorylation of human JNKs by activated MKK4/MKK7
R	242	H	165		FCERI-mediated MAPK activation	Phosphorylation of MEK7 by MEKK1
R	1,221	G	165	MET	Semaphorin interactions	SEMA4D interacts with Plexin-B1:Met
R	691	C/H	165	NTRK2	–	–
N	201	D/I/S	189	NUAK1	–	–
R	841	K/S	189	PDGFRA	Signaling by PDGF	Autophosphorylation of PDGF alpha receptors
C	386	R/S	87	PRKCB	Disinhibition of SNARE formation	Phosphorylation of platelet Sec-1
R	465	H	165		Signaling by GPCR	Gz is a substrate for PKC
C	424	R	87	PRKCQ	–	–
S	241	L/T	87	TGFBR1	Loss of function of SMAD2/3 in cancer	Activated type I receptor phosphorylates SMAD2/3 directly

The full set of variants can be found by executing example query 15 on the “Canonical RD Pocket” motif, located at http://vulcan.cs.uga.edu/prokino/query/Q15 or in Supp. Table S1.

at PKA equivalent positions 87, 165, and 189. Kinases that contain basic residues (R/K/H) at positions structurally equivalent to the RD pocket positions in PKA are classified as canonical while others, such as JAK2, are classified as non-canonical.

Given the functional role of the RD pocket, one can ask the following questions: “In which kinases, if any, are RD-pocket variants observed in cancer samples?” and “How do these variants alter the RD pocket and which of the variant kinases are regulated by activation loop phosphorylation?” Answering these questions typically involves retrieval of data from various data sources such as COSMIC, UniProt, and PDB followed by post-processing to identify variants mapping to the RD pocket. However, because the concept “Canonical RD Pocket” is represented by the *Structural Motif* class in ProKinO, the above questions can be answered rapidly by querying the ontology. A query requesting variants in the RD pocket revealed multiple kinases with recurrent variants at pocket positions (Table 1). Analysis of amino acid types at the pocket positions in WT and mutant forms indicate that the basic property of the pocket residues is altered in many cancer samples (Fig. 2C–E).

In the cell cycle check point kinase-2 (CHEK2), for example, the RD-pocket variants p.R165C, p.R165H, p.R165G, and p.R165S have been observed in cancers. Queries requesting the reactions and pathways associated with kinases harboring RD pocket variants reveal that CHEK2 controls multiple pathways associated with cell cycle control and DNA damage repair upon activation loop phosphorylation [Xu et al., 2002] (Table 1). Based on this contextual information, one can formulate the testable hypothesis that CHEK2 RD pocket variants impact cell cycle control by impairing CHEK2 regulation via activation loop phosphorylation.

While the majority of variants in the RD pocket replace a basic residue with a polar or hydrophobic residue, in some cases, a potential canonical pocket is formed because of the variant. In PRKCQ and PRKCB, for example, a WT cysteine (at PKA position 87 in the C-helix) is mutated to an arginine (Table 1). This variant is predicted to coordinate with the activation loop phosphate in a manner analogous to a canonical RD pocket.

To investigate how the phosphorylation sites in the activation loop are altered in human cancers, we used the modified residue property in the *Functional Feature* class and instances of the *Motif* class to identify variants that alter phosphorylated serine, threonine or tyrosine residues in the activation loop. Our analysis revealed multiple phosphorylated tyrosine residues that are mutated to a serine or aspartate (Fig. 2B). These variants are interesting because replacement of tyrosine by serine, as observed in ALK and PDGFRA (Supp. Table S1), is predicted to rewire signaling networks by introducing a new phosphorylation site [Tan et al., 2009]. On the other hand, replacement of phosphorylated tyrosine by an aspartate may constitutively activate the kinase, with the negatively charged aspartate functioning as a phosphomimic.

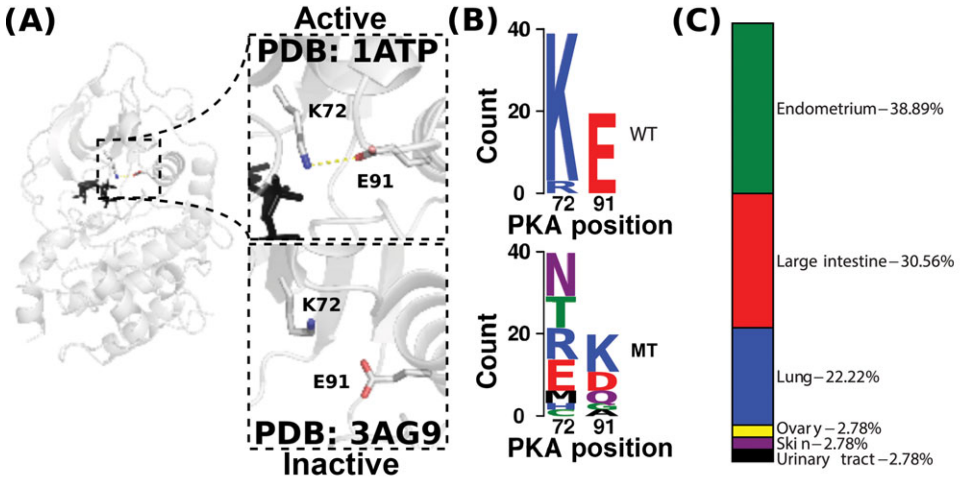

Figure 3. KE salt bridge variants. **A**: Salt bridge between lysine (p.K72) in the β3 strand and glutamate (p.E91) in the C-helix. ATP is shown in black sticks (PDB:1ATP). **B**: Somatic cancer variants mapping to p.K72 and p.E91. The top panel shows amino acids observed at that position in WT human kinases. The bottom panel shows the amino acids observed at the corresponding positions in mutant (MT) kinases. **C**: Sample primary sites with variants in the salt bridge positions.

Variants in the Lysine Glutamate (KE) Salt Bridge Interaction

The lysine glutamate salt bridge is a structural motif formed between a conserved lysine (p.K72) in sub-domain II and a glutamate (p.E91) in sub-domain III (C-helix). Although the KE salt bridge interaction is formed in most active structures, it is broken in many inactive structures due to repositioning of the flexible C-helix (Fig. 3A) [Jeffrey et al., 1995; Wenqing et al., 1997]. The KE salt bridge terminology is widely used to describe kinase activation and regulatory mechanisms, but has not been systematically studied in the context of cancer variants.

To determine if the KE salt bridge is altered in cancers, we incorporated the "KE Salt Bridge" concept in the *Structural Motif* class. Using the *locatedIn* relation between the *Structural Motif* and *Mutation* class, we queried for variants mapping to the KE salt bridge (Table 2). p.K72 is predominately mutated to an asparagine, threonine, arginine or glutamate in many kinases (Fig. 3B) across a variety of tissues (Fig. 3C). These variants are predicted to inactivate the kinase, as mutational studies in PKA and other kinases have shown that mutation of p.K72 to an arginine abrogates kinase catalytic activity [Iyer et al., 2005; Strutz-Seebohm et al., 2005; Zhong et al., 2011]. p.E91 is mutated to a lysine in a significant number of kinases and these variants are also predicted to inactivate the kinase by introducing a repelling electrostatic interaction with p.K72. One may suspect that a double variant, p.[(K72R; E91D)], could establish a similar salt bridge. However, this combination has not yet been observed in a sequenced cancer sample. Only two kinases, PRKG1 and TGFBR2, have somatic variants sequenced at both positions, but these were detected in distinct samples (Supp. Table S2).

Although the majority of variants altering the KE salt bridge are predicted to impair catalytic activity, it is unclear if they will impact kinase scaffolding functions, as demonstrated for some pseudokinases [Kornev and Taylor, 2009; Hu et al., 2011]. Notably, some of the kinases harboring variants at p.72 are predicted pseudokinases, including the kinase suppressor of Ras 1 (KSR1) and ERBB3, while the Vaccinia-related kinase 2 (VRK2) is a predicted pseudokinase that harbors variants at p.E91 (Table 2).

Variants Mapping to the Hydrophobic Spine

The hydrophobic spine is a structural motif encompassing residues from different regions of the kinase domain. The hydrophobic spine terminology was introduced [Kornev and Taylor, 2010] to describe the conserved hydrophobic interactions spanning the ATP and substrate binding lobes of the kinase domain. It is classified into the catalytic (C-) and regulatory (R-) spines based on their proposed role in kinase functions. The R-spine (consisting of PKA residues p.L95, p.L106, p.Y164, p.F185 and p.D220) is dynamically

Table 2. A Subset of KE Salt Bridge Variants Shown with Related Pathway and Reaction Data

Wild type	Position	Mutant type	PKA position	Gene	Pathway	Reaction
K	745	M/R	72	EGFR	EGFR downregulation	Phosphorylation of CBL (EGFR:GRB2:CBL)
K	742	E/M	72	ERBB3	PI3K events in ERBB2 signaling	PIP2 to PIP3 conversion by PI3K bound to phosphorylated heterodimer of ERBB2 and ERBB4 CYT1
R	617	C/H	72	KSR1	–	–
K	51	N/T	72	RIPK4	–	–
K	78	E/R	72	STK11	PI3K cascade	Activation of cytosolic AMPK by phosphorylation
K	277	E/N/R	72	TGFBR2	Loss of function of TGFBR2 in cancer	Activated type I receptor phosphorylates SMAD2/3 directly
E	290	K	91			
E	73	Q	91	VRK2	–	–

The full set of variants can be found by executing example query 16 on the "KE Salt Bridge" motif, located at http://vulcan.cs.uga.edu/prokino/query/Q16 or in Supp. Table S2.

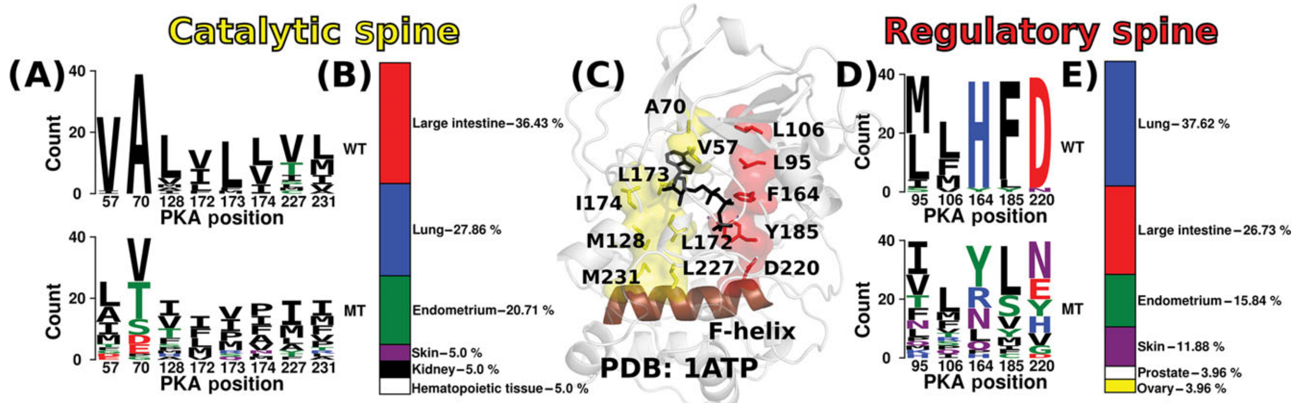

Figure 4. The hydrophobic spines. **A**: Wild type (WT) and mutant (MT) residues for the Catalytic spine. **B**: Distribution of sample primary sites with C-spine variants. **C**: Catalytic (yellow, left) and regulatory (red, right) spines. The F-helix is colored brown and adenosine triphosphate (ATP) is depicted as black sticks (PDB:1ATP). **D**: WT and mutant residues for the R-spine. **E**: Distribution of sample primary sites for Regulatory spine variants.

assembled upon kinase activation and is proposed to play a regulatory role. The C-spine is completed upon ATP binding and is believed to play a role in ATP binding and catalysis (Fig. 4C). To provide structural context for variants that map to the hydrophobic spines, we introduced two new concepts in ProKinO: the "Catalytic Spine" and the "Regulatory Spine"; and related these concepts to the *Mutation* class using the *locatedIn* relation. The concepts are included as part of the *Structural Motif* class and instantiated with residues that define the spine in each kinase (see *Methods*). Utilizing this conceptual representation, we formulated a query to identify kinases with variants in the C- and R-spine positions.

Within the R-spine, p.F185 and p.D220 are among the most frequently mutated residues (Fig. 4D). p.F185 is located in the conserved DFG motif and undergoes a conformational change from a "DFG-out" conformation in the inactive state to a "DFG-in" conformation in the active state, resulting in the assembly of the R-spine [Bukhtiyarova et al., 2007]. Mutation of p.F185 to an aspartate or asparagine results in loss of catalytic activity in PKA and other kinases [Meharena et al., 2013]. Our queries revealed similar recurrent variants in BRAF and EGFR at position p.F185, which we predict to inactivate the kinase by destabilizing the R-spine (Table 3). p.D220, a conserved R-spine residue in the F-helix that is mutated in multiple cancers (Fig. 4E), maintains the backbone conformation of the catalytic HRD motif in a "strained" conformation in the active state and loss of this conformational strain is correlated with the disassembly of the R-spine and kinase inactivation [Oruganty et al., 2013]. Variants that disrupt the hydrogen bonding interaction between p.D220 and the HRD motif backbone (p.D220A and p.D220N) have been shown to abrogate catalytic activity [Meharena et al., 2013; Oruganty et al., 2013]. Based on this information, we

Table 3. A Subset of Regulatory Spine Variants Shown with Related Pathway and Reaction Data

Wild type	Position	Mutant type	PKA position	Gene	Pathway	Reaction
I	1171	N	95	ALK	–	–
H	574	N/Q	164	BRAF	NGF signaling via TRKA from the plasma membrane	(Frs2)Rap1-GTP binds to and activates B-Raf
F	595	L/S	185			
L	777	P/Q	106	EGFR	EGFR transactivation by gastrin	SOS1-mediated nucleotide exchange of RAS (HB-EFG-initiated)
H	835	L	164		EGFR downregulation	Phosphorylation of CBL (EGFR:GRB2:CBL)
F	856	L/S/Y	185		GAB1 signalosome	Gab1 phosphorylation by EGFR kinase
M	77	I	95	JNK1	Apoptosis	Phosphorylation of BIM by JNK
H	149	Y	164		Cell death signalling via NRAGE, NRIF and NADE	NRAGE activates JNK
H	174	R	164	LKB1	PI3K cascade	Activation of cytosolic AMPK by phosphorylation
L	195	M	185			
D	237	G/Y	220			
L	85	P/Q	95	LOK	–	–
F	129	L	106	MAP2K1	Cytokine signaling in immune system	TPL2 phosphorylates MEK1, SEK1
M	128	I	95	ROCK1	–	–
L	139	I	106			
H	196	Y	164			
D	446	N/V	220	TGFBR2	TGFBR1 KD mutants in cancer	TGFBR2 phosphorylates TGFBR1 KD mutants
H	677	R/Y	164	TRKC	–	–
F	209	C/L	106	WNK3	–	–
L	295	I	185			

The full set of variants can be found by executing example query 18 on the "Regulatory Spine" motif, located at http://vulcan.cs.uga.edu/prokino/query/Q18 or in Supp. Table S3.

Table 4. A Subset of Catalytic Spine Variants Shown with Related Pathway and Reaction Data

Wild type	Position	Mutant type	PKA position	Gene	Pathway	Reaction
V	256	L	57	ABL1	Fcgamma receptor (FCGR)-dependent phagocytosis	ABL phosphorylates WAVE2
L	370	P/R	173			
V	371	A/L	174			
L	1,318	M	227	ALK	–	–
V	471	A/F/I	57	BRAF	CREB phosphorylation through the activation of Ras	Raf activation
L	537	S	128			
I	582	M	172		NGF signaling via TRKA from the plasma membrane	(Frs2)Rap1-GTP binds to and activates B-Raf
F	583	S	173			
L	584	F/P	174			
A	247	D/T	70	CHEK2	G2/M checkpoints	Phosphorylation and activation of CHK2 by ATM
L	354	V	173		p53-independent G1/S DNA damage checkpoint	Phosphorylation of Cdc25A at Ser-123 by Chk2
L	355	P	174			
L	477	I/V	128	DCLK2	–	–
A	743	P/S/T/V	70	EGFR	EGFR transactivation by gastrin	SOS1-mediated nucleotide exchange of RAS (HB-EFG-initiated)
L	798	F/H/P/V	128			
V	843	I/L	172			
L	844	P/V	173		EGFR downregulation	Phosphorylation of CBL (EGFR:GRB2:CBL)
V	845	A/M	174			
L	907	M	231			
A	651	T/V	70	EPHA3	–	–
M	256	T	231	MAP2K1	Cytokine signaling in immune system	TPL2 phosphorylates MEK1, SEK1
I	500	M/T	128	MAP3K19	–	–
L	115	F/I	128	MAPK8	Apoptosis	Phosphorylation of BIM by JNK
A	1,126	S	70	MET	Axon guidance	Inactivation of R-Ras by Sema4D-Plexin-B1 GAP activity
V	1,290	L	227		Sema4D in semaphorin signaling	Inactivation of Rho-GTP by p190RhoGAP
V	463	I/L	57	PAK7	Signaling by Robo receptor	Activation of Rac by Sos
M	636	I	231			
V	750	I	57	STK31	–	–
V	860	A/I/L	172			
V	104	A/L	57	STK38L	–	–

The full set of variants can be found by executing example query 17 on the "Catalytic Spine" motif, located at http://vulcan.cs.uga.edu/prokino/query/Q17 or in Supp. Table S4.

predict that the asparagine variants observed at position p.D220, as in TGFBR2 (Supp. Table S3), inactivate the kinase.

Our queries also reveal recurrent variants in the C-spine (Fig. 4A and Supp. Table S4) in a variety of tissue types (Fig. 4B). PKA residue p.A70 forms hydrophobic interactions with the adenosine group of ATP and is one of the most frequently mutated C-spine residues. The small size of the alanine side-chain is important for accommodating ATP [Hu et al., 2011]. As many of the variants observed at p.A70 increase amino acid size (Fig. 4A), they are predicted to sterically block access to the ATP binding site. Recent studies on the RAF kinases have shown that mutation of p.A70 to a phenylalanine impairs ATP binding, but retains the scaffolding functions of the kinase [Hu et al., 2011]. Thus, even though C-spine variants are likely to impair ATP binding and catalysis, scaffolding functions may be retained. We used information from the *Reaction* and *Pathway* classes captured in ProKinO to obtain insights into the scaffolding functions associated with mutated kinases. Kinases harboring C-spine variants are activated by dimerization and interacting partners in signaling pathways (Table 4). This information can be used to generate hypotheses regarding the impact of C-spine variants on kinase catalytic and scaffolding functions. Together, these examples demonstrate how mining cancer variants in the context of structural motifs and pathways can provide new hypotheses for experimental studies.

ProKinO Provides a Framework for Hypothesis Generation and Testing

To validate the utility of ProKinO in knowledge discovery and hypothesis testing, we performed iterative querying while assuming minimal prior knowledge of kinase structure and function. Below we demonstrate how the iterative querying process, followed by experimental studies, resulted in the identification of a novel mutational hotspot in the kinase domain.

Hypothesis Generation Through Iterative Ontology Querying

Since information on wild and mutant type residues is captured for each nonsynonymous variant, we began by asking a simple question: "Are certain amino acid types more frequently mutated in the kinase domain?" We translated this question into a SPARQL query using the information conceptualized in ProKinO. Our query revealed that arginine is the most frequently mutated residue in the kinase domain (Fig. 5A). Based on this knowledge, we next formulated a query to identify the kinases harboring arginine variants. The results depicted in Figure 5B show that arginine variants are found in a diverse array of kinases, but with the greatest frequency in EGFR. Using the *Gene* and *Sequence Motif* classes and the *locatedIn* relation between them, we queried for the structural location

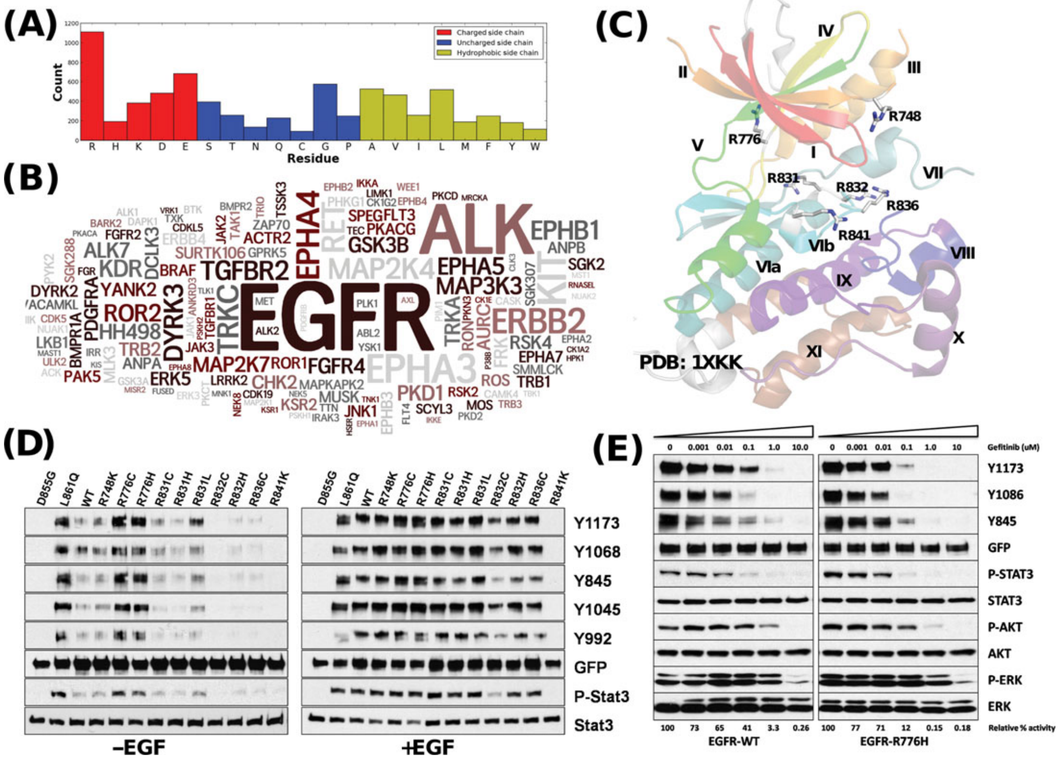

Figure 5. Iterative querying and hypothesis generation. **A**: Plot showing the frequency of each amino acids mutated in cancers in the protein kinase domain. **B**: Wordle image showing kinases harboring arginine variants. The text height is proportional to the number of arginine variants observed in the corresponding kinase domain. **C**: Locations of EGFR arginine variants in the crystal structure (PDB:1XKK). Subdomains are colored and labeled. **D**: Western blot results showing constitutive activity of p.R776^{105PKA} mutants in the absence (–) and presence (+) of activating EGF ligand. **E**: EGFR auto-phosphorylation and downstream signaling of wild type and p.R776H^{105PKA} mutant inhibition with varied concentrations of gefitinib.

of the arginine variants in EGFR (Fig. 5C). Two of the arginine residues (p.R836^{165PKA} and p.R841^{170PKA}) are found in subdomain VIb and are part of the RD pocket and substrate binding pocket, respectively. The other arginine residues (p.R748^{75PKA}, p.R776^{105PKA}, p.R831^{160PKA}, and p.R832^{161PKA}), on the other hand, are not part of any known functional site, but are frequently mutated in cancer samples.

Experimental Characterization of Arginine Variants in EGFR

To understand the functional impact of arginine variants in EGFR, we analyzed the associated pathways and reactions in ProKinO. EGFR is a receptor tyrosine kinase that controls a diverse array of cellular processes associated with cell migration, adhesion and proliferation. Its constitutive activity has been correlated with several cancer types and a variety of commercial inhibitors have been developed to abrogate this activity [Lynch et al., 2004]. Autophosphorylation of EGFR is one of the well-studied reactions in EGFR signaling, in which binding of EGF to the receptor activates the kinase domain and leads to autophosphorylation of Tyr residues (p.Y845^{174PKA}, p.Y992^{0PKA}, p.Y1064^{0PKA}, and p.Y1173^{0PKA}) in the C-terminal tail [Helin and Beguinot, 1991; Margolis et al., 1989; Walton et al., 1990], and downstream phosphorylation of proteins such as the transcription factor Stat3 [Chan et al., 2004]. Based on this knowledge, we formulated a testable hypothesis that causative arginine variants will impact EGFR autophosphorylation and Stat3 phosphorylation. To test this hypothesis, we transfected WT and mutant EGFR in Chinese hamster cells (which express very low levels of EGFR) and probed for phosphorylation of EGFR C-terminal tail and Stat3 tyrosine residues using western blot analysis, as described in the *Methods* section.

The substrate binding pocket variant (p.R841K^{170PKA}) abrogates EGFR activity to the same extent as the catalytically dead variant (p.D855G^{187PKA}) (Fig. 5D). In contrast, mutation of p.R831^{160PKA}, p.R832^{161PKA}, or p.R836^{165PKA} shows no significant change in C-terminal tail and Stat3 phosphorylation compared to WT. Interestingly, however, p.R776C/H^{105PKA} variants increase EGFR activity in the absence of the activating EGF ligand. The extent of EGFR activation by p.R776C/H^{105PKA} is comparable to p.L861Q^{190PKA} (Fig. 5D), a well-known lung cancer variant that also activates EGFR in a ligand independent manner [Choi et al., 2007]. Cancer cells harboring p.R776C/H^{105PKA} variants in EGFR respond better to treatment with inhibitors [Lynch et al., 2004]. Consistent with previous studies, our experimental results indicate increased sensitivity of the EGFR p.R776H^{105PKA} mutant to gefitinib treatment in comparison to WT (Fig. 5E).

Position 776^{105PKA} in the αC-β4 Loop is a Mutational Hotspot

To obtain additional insights on the mechanisms by which p.R776^{105PKA} variants activate EGFR and confer drug sensitivity, we posed the following question: “Is the residue equivalent to p.R776^{105PKA} mutated in other kinases?”. If so, “What is the nature of WT and mutant type amino acids observed at the p.R776^{105PKA}

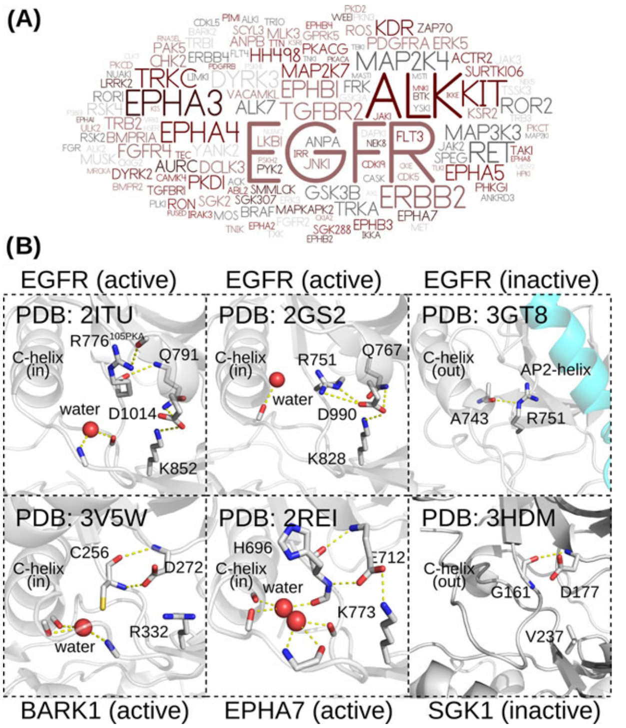

Figure 6. Mechanisms of activation of p.R776H^{105PKA} and p.R776C^{105PKA} variants in EGFR. **A**: Kinases with variants at position p.R776^{105PKA}. The text height is proportional to the number of variants. **B**: Crystal structures (PDBs: 2ITU, 2GS2, and 3GT8) of EGFR showing common p.R776^{105PKA} orientations. Inactive structure shows a common C-helix capping interaction, whereas active structures instead coordinate with the hinge region and C-terminal tail. Bottom shows kinase structures with naturally occurring cysteine (PDB: 3V5W), histidine (PDB: 2REI) and glycine (PDB: 3HDM) at position p.R776^{105PKA}.

position?" Because protein kinases are evolutionarily related and structurally conserved, analysis of kinases with variants at structurally equivalent positions can reveal shared mechanisms of mutational activation and drug inhibition. Likewise, analysis of kinases naturally conserving mutant types at equivalent positions can provide insights into the impact of variants on kinase structure, function and drug binding. We queried for variants at position p.776^{105PKA}. Our queries indicate multiple kinases with disease variants at position p.776^{105PKA} in the αC-β4 loop (Fig. 6A). The αC-β4 loop serves as a hinge point for C-helix and inter-lobe movement and variants in the loop contribute to abnormal kinase regulation in FGFR and PDGFR [Chen et al., 2007; Kannan et al., 2008].

Based on this knowledge and our query results, we hypothesized that p.R776C/H^{105PKA} variants activate EGFR by relieving auto-inhibitory hinge interactions associated with C-helix movement (Fig. 6B). Consistent with this view, comparisons of C-helix "in" (active) and C-helix "out" (inactive) conformations indicates loss of capping interaction between p.R776^{105PKA} side-chain and C-helix backbone upon C-helix movement (Fig. 6B). Furthermore, analysis of kinases that naturally conserve a histidine or cysteine at position p.776^{105PKA} (analogous to the mutant types in EGFR) indicate that, in these kinases, the C-helix is held in an active "in" conformation and the C-helix hinge interactions are partially mediated by conserved water molecules. We also note that in EGFR, p.R776^{105PKA} is within hydrogen bonding proximity to the C-terminal auto-inhibitory AP2-helix, suggesting that both auto-inhibitory C-helix hinge and C-terminal tail interactions may be relieved by the p.R776C/H^{105PKA} variant. Further studies are needed to fully understand the mechanisms by which the p.R776C/H^{105PKA} variant activates EGFR.

Concluding Remarks

We have demonstrated that ProKinO is a valuable resource for mining and annotating the cancer kinome. In particular, the conceptual representation of knowledge related to structural and functional motifs allows effective mining of cancer variants while facilitating hypothesis generation and testing. Our ontological approach is conceptually different from previous structure and machine learning based approaches to predict variant impact [Capriotti and Altman, 2011; Shi and Moult, 2011; Hashimoto et al., 2012; Izarzugaza et al., 2012; Dixit and Verkhivker, 2014]. Aggregate queries such as "the number of activation loop variants in each kinase", or "the number of mutated kinases involved in each pathway" can be rapidly performed using ProKinO and provide the information necessary to generate hypotheses for experimental studies. Through the iterative process of ontology querying, reasoning, hypotheses generation and testing, we have identified a novel mutational hotspot in the αC-β4 loop region of the kinase domain and demonstrated its functional relevance in EGFR activity and drug sensitivity.

Computational approaches, like those available in the Cancer Related Analysis of VAriance Toolkit (CRAVAT) [Douville et al., 2013], have proven useful in separating likely driver variants from passenger. However, by making predictions on all proteins, they necessarily miss the wealth of knowledge stored in domain specific and single locus databases, and further don't provide the structural context necessary to frame a testable hypothesis. The results presented here will serve as a conceptual starting point for experimental studies and help prioritize key variants and mutated kinases for functional characterization and drug discovery [Simpson et al., 2009; Brognard et al., 2011; Eglen and Reisine, 2011; Antal et al., 2014].

While ProKinO offers several utilities for integrative analysis of protein kinase data, it needs to be further developed to fully realize its impact in kinase research. For example, sequence and structural motifs that contribute to the functional specificity of major kinase groups and families can be added in the *Sequence* and *Structural Motif* classes to explore how variants impact family or group specific functions. Network analysis on missense variants has revealed their preponderance in protein–protein, protein–nucleic acid, and protein–ion interfaces and validated that proteins involved in signal transduction are more frequently mutated in cancer [Nishi et al., 2013]. These interaction interfaces can provide the context crucial to predict variant impact. Likewise, incorporating information on kinase substrates and phosphorylation patterns can provide additional functional context for predicting variant impact [Hanks and Hunter, 1995; Ashburner et al., 2000; Lim and Pawson, 2010].

Finally, user-friendly interfaces need to be incorporated to facilitate integrative analysis of ProKinO data by a wide range of scientific users. In particular, SPARQL query construction requires both an in-depth knowledge of the SPARQL query language and a semantic understanding of the ontology. We are working on a graphical query builder, which will allow the formulation of queries by visually inspecting the classes and relations in the ontology schema. This will allow biologists who are not familiar with the SPARQL query language to formulate integrative, hypothesis-oriented queries on ProKinO data. As part of future development, we also plan to incorporate structural visualization tools such as Mutation Position Imaging Toolbox (MuPIT) [Niknafs et al., 2013] and data visualization tools like SGVizler [Skjæveland, 2012]. These tools are expected to enhance the usability of ProKinO and, consequently, accelerate the functional characterization of the cancer kinome.

Acknowledgment

Disclosure statement: The authors declare no conflict of interest.

References

Antal C, Kang E, Stephenson N, Trotter E, Hunter T, Brognard J, Newton A. 2014. Prevalence of inactivating protein kinase C mutations in human cancers (1055.2). FASEB J 28:1055-1052.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, et al. 2000. Gene ontology: tool for the unification of biology. Nat Genet 25:25–29.

Brognard J, Hunter T. 2011. Protein kinase signaling networks in cancer. Curr Opin Genet Dev 21:4–11.

Brognard J, Zhang Y-W, Puto LA, Hunter T. 2011. Cancer-associated loss-of-function mutations implicate DAPK3 as a tumor-suppressing kinase. Cancer Res 71:3152–3161.

Bukhtiyarova M, Karpusas M, Northrop K, Namboodiri HVM, Springman EB. 2007. Mutagenesis of p38α MAP kinase establishes key roles of Phe169 in function and structural dynamics and reveals a novel DFG-OUT state. Biochemistry 46:5687–5696.

Capriotti E, Altman RB. 2011. Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinform 12 Suppl 4:S3.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2:401–404.

Chan KS, Carbajal S, Kiguchi K, Clifford J, Sano S, DiGiovanni J. 2004. Epidermal growth factor receptor-mediated activation of Stat3 during multistage skin carcinogenesis. Cancer Res 64:2382–9.

Chen H, Ma J, Li W, Eliseenkova AV, Xu C, Neubert TA, Miller WT, Mohammadi M. 2007. A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. Mol Cell 27:717–730.

Choi SH, Mendrola JM, Lemmon MA. 2007. EGF-independent activation of cell-surface EGF receptors harboring mutations found in gefitinib-sensitive lung cancer. Oncogene 26:1567–76.

Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D, Zhu J. 2013. Exploring TCGA pan-cancer data at the UCSC cancer genomics browser. Sci Rep 3:2652.

Deininger M, Buchdunger E, Druker BJ. 2005. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood 105:2640–2653.

DeLano WL. 2002. PyMOL.San Carlos, CA: DeLano Scientific. p 700.

Dixit A, Verkhivker GM. 2014. Structure-functional prediction and analysis of cancer mutation effects in protein kinases. Comput Math Methods Med 2014:653487.

Douville C, Carter H, Kim R, Niknafs N, Diekhans M, Stenson PD, Cooper DN, Ryan M, Karchin R. 2013. CRAVAT: cancer-related analysis of variants toolkit. Bioinformatics 29:647–8.

Eglen R, Reisine T. 2011. Drug discovery and the human kinome: recent trends. Pharmacol Ther 130:144–56.

Feinberg J. c2001. Wordle – Beautiful Word Clouds. http://www.wordle.net/.

Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. 2008. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet Chapter 10:Unit 10 11.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. Nature reviews. Cancer 4:177–183.

Gosal G, Kochut KJ, Kannan N. 2011a. ProKinO: an ontology for integrative analysis of protein kinases in cancer. PLoS ONE 6:e28782-e28782.

Gosal GPS, Kannan N, Kochut KJ. 2011b. ProKinO: a framework for protein kinase ontology. 2011 IEEE International Conference on Bioinformatics and Biomedicine: 550–555.

Hanks SK, Hunter T. 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J 9:576–96.

Hashimoto K, Rogozin IB, Panchenko AR. 2012. Oncogenic potential is related to activating effect of cancer single and double somatic mutations in receptor tyrosine kinases. Hum Mutat 33:1566–75.

Helin K, Beguinot L. 1991. Internalization and down-regulation of the human epidermal growth factor receptor are regulated by the carboxyl-terminal tyrosines. J Biol Chem 266:8363–8368.

Hu J, Yu H, Kornev AP, Zhao J, Filbert EL, Taylor SS, Shaw AS. 2011. Mutation that blocks ATP binding creates a pseudokinase stabilizing the scaffolding function of kinase suppressor of Ras, CRAF and BRAF. Proce Natl Acad Sci USA 108:6067–6072.

Huse M, Kuriyan J. 2002. The conformational plasticity of protein kinases. Cell 109:275–82.

Iyer GH, Garrod S, Woods VL Jr, Taylor SS. 2005. Catalytic independent functions of a protein kinase as revealed by a kinase-dead mutant: study of the Lys72His mutant of cAMP-dependent kinase. J Mol Biol 351:1110–1122.

Izarzugaza JM, del Pozo A, Vazquez M, Valencia A. 2012. Prioritization of pathogenic mutations in the protein kinase superfamily. BMC Genomics 13 Suppl 4:S3.

Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP. 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature 376:313–20.

Johnson LN, Noble MEM, Owen DJ. 1996. Active and inactive protein kinases: structural basis for regulation. Cell 85:149–158.

Kannan N, Neuwald AF, Taylor SS. 2008. Analogous regulatory sites within the alphaC-beta4 loop regions of ZAP-70 tyrosine kinase and AGC kinases. Biochim Biophys Acta 1784:27–32.

Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G. 2007. Structural and functional diversity of the microbial kinome. PLoS Biol 5:e17.

Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong N-H, Taylor SS, Sowadski JM. 1991. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science 253:407–414.

Kornev AP, Taylor SS. 2009. Pseudokinases: functional insights gleaned from structure. Structure 17:5–7.

Kornev AP, Taylor SS. 2010. Defining the conserved internal architecture of a protein kinase. Biochim Biophys Acta 1804:440–444.

Lahiry P, Torkamani A, Schork NJ, Hegele RA. 2010. Kinase mutations in human disease: interpreting genotype-phenotype relationships. Nat Rev Genet 11:60–74.

Lassila O, Swick RR. 1998. Resource Description Framework (RDF) model and syntax specification. World Wide Web Consortium Recommendation, W3C.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25:2744–50.

Lim WA, Pawson T. 2010. Phosphotyrosine signaling: evolving a new cellular communication system. Cell 142:661–7.

Lucet IS, Fantino E, Styles M, Bamert R, Patel O, Broughton SE, Walter M, Burns CJ, Treutlein H, Wilks AF. 2006. The structural basis of Janus kinase 2 inhibition by a potent and specific pan-Janus kinase inhibitor. Blood 107:176–183.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, et al. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350:2129–39.

Manning G, Plowman GD, Hunter T, Sudarsanam S. 2002a. Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci 27:514–520.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002b. The protein kinase complement of the human genome. Science (New York, N.Y.) 298:1912–1934.

Margolis BL, Lax I, Kris R, Dombalagian M, Honegger AM, Howk R, Givol D, Ullrich A, Schlessinger J. 1989. All autophosphorylation sites of epidermal growth factor (EGF) receptor and HER2/neu are located in their carboxyl-terminal tails. Identification of a novel site in EGF receptor. J Biol Chem 264:10667–71.

Meharena HS, Chang P, Keshwani MM, Oruganty K, Nene AK, Kannan N, Taylor SS, Kornev AP. 2013. Deciphering the structural basis of eukaryotic protein kinase regulation. PLoS Biol 11.

Neuwald AF. 2009. Rapid detection, classification and accurate alignment of up to a million or more related protein sequences. Bioinformatics (Oxford, England) 25:1869–1875.

Niknafs N, Kim D, Kim R, Diekhans M, Ryan M, Stenson PD, Cooper DN, Karchin R. 2013. MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures. Hum Genet 132:1235–43.

Nishi H, Tyagi M, Teng S, Shoemaker BA, Hashimoto K, Alexov E, Wuchty S, Panchenko AR. 2013. Cancer missense mutations alter binding properties of proteins and their interaction networks. PloS ONE 8:e66273.

Ortutay C, Valiaho J, Stenberg K, Vihinen M. 2005. KinMutBase: a registry of disease-causing mutations in protein kinase domains. Hum Mutat 25:435–42.

Oruganty K, Kannan N. 2012. Design principles underpinning the regulatory diversity of protein kinases. Philos Trans R Soc Lond B Biol Sci 367:2529–39.

Oruganty K, Talathi NS, Wood ZA, Kannan N. 2013. Identification of a hidden strain switch provides clues to an ancient structural mechanism in protein kinases. Proc Natl Acad Sci USA 110:924–929.

Parthiban V, Gromiha MM, Schomburg D. 2006. CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res 34(Web Server issue):W239–242.

Prud E, Seaborne A. 2006. Sparql query language for rdf.

Russo AA, Jeffrey PD, Pavletich NP. 1996. Structural basis of cyclin-dependent kinase activation by phosphorylation. Nat Struct Biol 3:696–700.

Sanner MF. 1999. Python: a programming language for software integration and development. J Mol Graph Model 17:57–61.

Shi Z, Moult J. 2011. Structural and functional impact of cancer-related missense somatic mutations. J Mol Biol 413:495–512.

Sicheri F, Moarefi I, Kuriyan J. 1997. Crystal structure of the Src family tyrosine kinase Hck. Nature 385:602–609.

Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40(Web Server issue):W452–7.

Simpson GL, Hughes JA, Washio Y, Bertrand SM. 2009. Direct small-molecule kinase activation: Novel approaches for a new era of drug discovery. Curr Opin Drug Discov Devel 12:585–96.

Skjæveland MG. 2012. Sgvizler: a javascript wrapper for easy visualization of sparql result sets.

Strutz-Seebohm N, Seebohm G, Mack AF, Wagner HJ, Just L, Skutella T, Lang UE, Henke G, Striegel M, Hollmann M. 2005. Regulation of GluR1 abundance in murine hippocampal neurones by serum-and glucocorticoid-inducible kinase 3. J Physiol 565:381–390.

Talevich E, Kannan N. 2013. Structural and evolutionary adaptation of rhoptry kinases and pseudokinases, a family of coccidian virulence factors. BMC Evolut Biol 13:117-117.

Tan CSH, Bodenmiller B, Pasculescu A, Jovanovic M, Hengartner MO, Jorgensen C, Bader GD, Aebersold R, Pawson T, Linding R. 2009. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. Sci Signal 2:ra39.

Taylor SS, Yang J, Wu J, Haste NM, Radzio-Andzelm E, Anand G. 2004. PKA: a portrait of protein kinase dynamics. Biochim Biophys Acta 1697:259–269.

Taylor SS, Zhang P, Steichen JM, Keshwani MM, Kornev AP. 2013. PKA: lessons learned after twenty years. p 1271–1278.

Walton GM, Chen WS, Rosenfeld MG, Gill GN. 1990. Analysis of deletions of the carboxyl terminus of the epidermal growth factor receptor reveals self-phosphorylation at tyrosine 992 and enhanced in vivo tyrosine phosphorylation of cell substrates. J Biol Chem 265:1750–1754.

Wenqing XU, Harrison SC, Eck MJ. 1997. Three-dimensional structure of the tyrosine kinase c-Src. Nature 385:595–602.

Worth CL, Preissner R, Blundell TL. 2011. SDM–a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 39(Web Server issue):W215–22.

Xu X, Tsvetkov LM, Stern DF. 2002. Chk2 activation and phosphorylation-dependent oligomerization. Mol Cell Biol 22:4419–4432.

Yamaguchi H, Hendrickson WA. 1996. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. Nature 384:484–9.

Zheng J, Knighton DR, Ten Eyck LF, Karlsson R, Xuong NH, Taylor SS, Sowadski JM. 1993. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with magnesium-ATP and peptide inhibitor. Biochemistry 32:2154–2161.

Zhong J, Chaerkady R, Kandasamy K, Gucek M, Cole RN, Pandey A. 2011. The interactome of a PTB domain-containing adapter protein, Odin, revealed by SILAC. J Proteomics 74:294–303.