



OPEN ACCESS

EDITED BY

Wenbo Ma,
Tulane University, United States

REVIEWED BY

Jürgen Dönitz,
University of Göttingen, Germany
Junya Tabata,
Jikei University School of Medicine, Japan
Vivek Panwar,
Chitkara University (Himachal Pradesh), India

*CORRESPONDENCE

Beth Apsel Winger
✉ beth.winger@ucsf.edu

RECEIVED 24 March 2025

ACCEPTED 26 June 2025

PUBLISHED 04 August 2025

CITATION

Rangarajan A, Sviezhentseva I, Gunderson E, Pikman Y, Jacobson MP and Apsel Winger B (2025) A structure-based tool to interpret the significance of kinase mutations in clinical next generation sequencing in cancer. *Front. Oncol.* 15:1599389. doi: 10.3389/fonc.2025.1599389

COPYRIGHT

© 2025 Rangarajan, Sviezhentseva, Gunderson, Pikman, Jacobson and Apsel Winger. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A structure-based tool to interpret the significance of kinase mutations in clinical next generation sequencing in cancer

Amith Rangarajan¹, Ilona Sviezhentseva², Emma Gunderson³, Yana Pikman^{2,4}, Matthew P. Jacobson¹ and Beth Apsel Winger^{3,5*}

¹Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, United States, ²Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, United States, ³Department of Clinical Pharmacy, University of California, San Francisco, San Francisco, CA, United States, ⁴Division of Hematology/Oncology, Boston Children's Hospital, Boston, MA, United States, ⁵Department of Pediatrics, Division of Hematology, University of California, San Francisco, San Francisco, CA, United States

Introduction: Clinical workflows to analyze variants of unknown significance (VUSs) found in clinical next generation sequencing (NGS) are labor intensive, requiring manual analysis of published data for each variant. There is a strong need for tools and resources that provide a consistent way to analyze variants. With the explosion of clinical NGS data and the concurrent availability of protein structures through the Protein Data Bank and protein models through programs such as AlphaFold, there exists an unprecedented opportunity to use structural information to help standardize NGS analysis with the overall goal of advancing personalized cancer therapy.

Methods: Using the Catalogue of Somatic Mutations in Cancer (COSMIC), the largest curated database of clinical cancer mutations, we mapped thousands of missense mutations in the kinase and juxtamembrane (JM) domains of 48 receptor tyrosine kinases (RTKs) onto structurally aligned kinase structures, then clustered known activating mutations along with VUSs based on proximity in three-dimensional structure. Using cell-based models we demonstrate that our resource can be used to aid in identification of activating mutations while providing insight into mechanisms of kinase activation and regulation.

Results: We provide a database of structurally aligned and functionally annotated mutations that can be used as a tool to evaluate kinase VUSs based on their structural alignment with known activating mutations. The tool can be accessed through a user-friendly website in which one can input a kinase mutation of interest, and the system will output a list of structurally analogous mutations in other kinases, as well as their functional annotations.

Discussion: Though our tool is not expected to be used as an isolated source for variant functional prediction, we expect our database will be a valuable addition to the current tools and resources used to analyze clinical NGS, with important clinical implications to guide recommendations for personalized cancer therapy.

KEYWORDS

cancer, targeted therapy, molecular pathology, next generation sequencing, kinases, structural biology, variants of unknown significance

Introduction

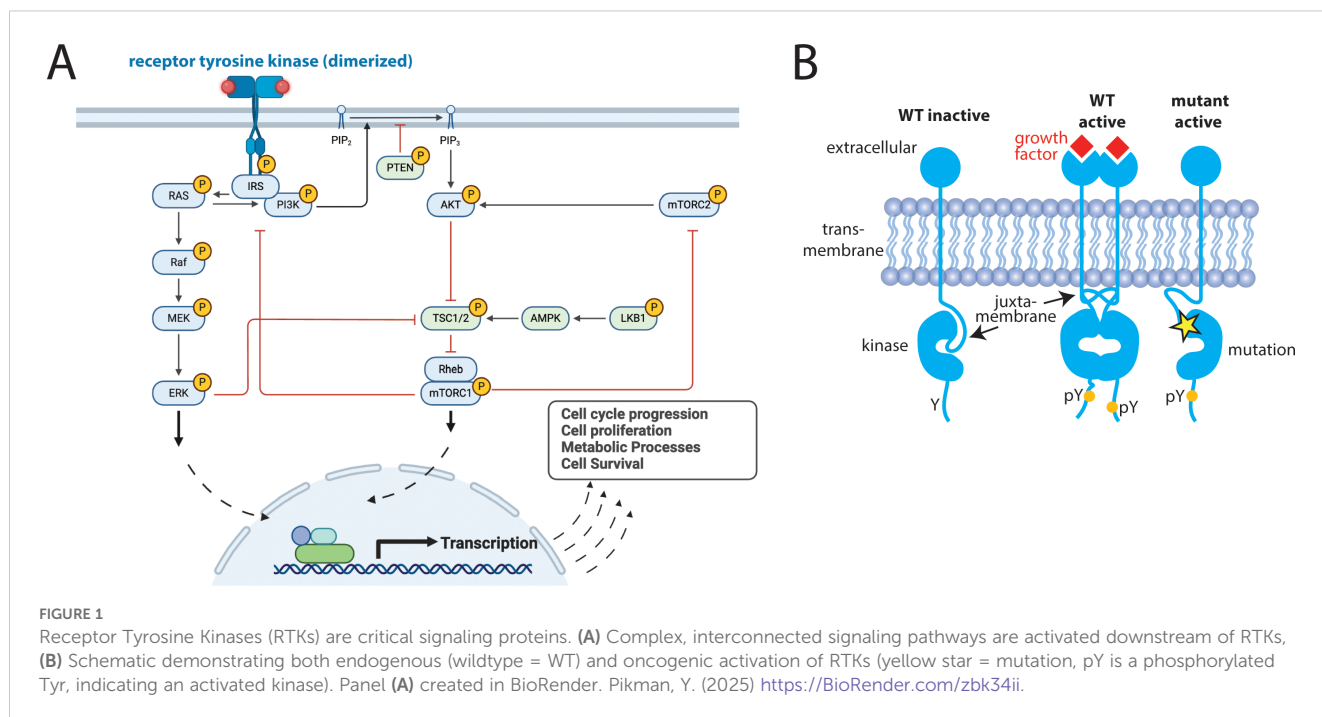
Protein kinases comprise the most targetable class of oncogenic proteins, with at least 80 FDA-approved kinase inhibitors available (1). Some of the greatest successes of kinase inhibitors include ABL kinase inhibitors in *BCR::ABL1* positive chronic myelogenous leukemia (CML) which transformed CML from a deadly disease curable only by allogeneic stem cell transplant to a chronic disease curable with an oral kinase inhibitor (2). In a more recent example, oral ALK kinase inhibitors became first line therapy for ALK-fusion positive non-small cell lung cancer after several studies showed similar efficacy and significantly improved side effects and quality of life for patients on ALK inhibitors compared to chemotherapy (3). There are many other examples, and extensive precision medicine efforts using clinical next generation sequencing (NGS) have now expanded the use of kinase inhibitors to virtually all cancer types (4). However, most mutations detected through NGS are variants of unknown significance (VUSs), which are genetic mutations that have an unknown effect on human health (5, 6). Kinase VUSs create a particular challenge for oncologists who have to decide whether to use kinase inhibitors for uncharacterized mutations in patients with otherwise limited treatment options (4).

Over 100 variant prediction algorithms have been created to address the challenge of VUS interpretation in NGS (7). Such algorithms, some of which are specific to disease or protein class, generally aim to predict the functional significance of mutations by incorporating multiple elements into a score and/or using machine learning/artificial intelligence (7–11). Most utilize protein sequence data, such as the degree of chemical change and evolutionary conservation at the mutated amino acid site; some also incorporate aspects of protein structure (12–15). However, the specificity of available algorithms is only ~60–80%, leading the American Association of Pathology to recommend against applying them to clinical care (11, 16, 17). Clinical workflows to analyze NGS continue to be labor intensive, requiring in-depth analysis of published data by molecular pathologists and other content experts who often come together as a Molecular Tumor Board (10, 18, 19). For each protein and each mutation, the data available for variant analysis differs, making NGS interpretation impossible

to standardize. There is a strong need for clinical and experimental data across variants that could help standardize NGS analysis.

Thousands of experimental structures are now available for human protein kinases, and these experimental structures are supplemented by models generated by powerful structure prediction algorithms such as AlphaFold (20). These advances have stimulated discussion of how protein structure can advance clinical practice (21). The functional impact of a given mutation, if any, depends sensitively on its location within the three-dimensional structure of the protein. For example, a mutation that changes the net charge of the protein (e.g., Lys->Glu) might have no significant impact on function if the amino acid is located on a surface loop distant from the active site but could cause the protein to become unfolded if the amino acid is buried in the core or could cause loss of function if the amino acid is located in the active site. Thus, the effect of a mutation on three-dimensional protein structure can provide important insights into the functional impact of mutations (19, 22, 23). Incorporating protein structure into NGS analysis is most effective if there is a clear understanding of how protein regions contribute to protein function. Since kinase structure-function relationships are well defined, understanding the impact of kinase mutations on protein structure can be particularly useful in providing functional insights such as whether a mutation is activating (24). This assertion is supported by the fact that many well characterized oncogenic mutations are concentrated in structurally analogous kinase regions. Examples of structurally analogous oncogenic mutations include juxtamembrane domain mutations in FLT3 and KIT (25–27), and activation loop mutations in EGFR and KIT, among others (28–30).

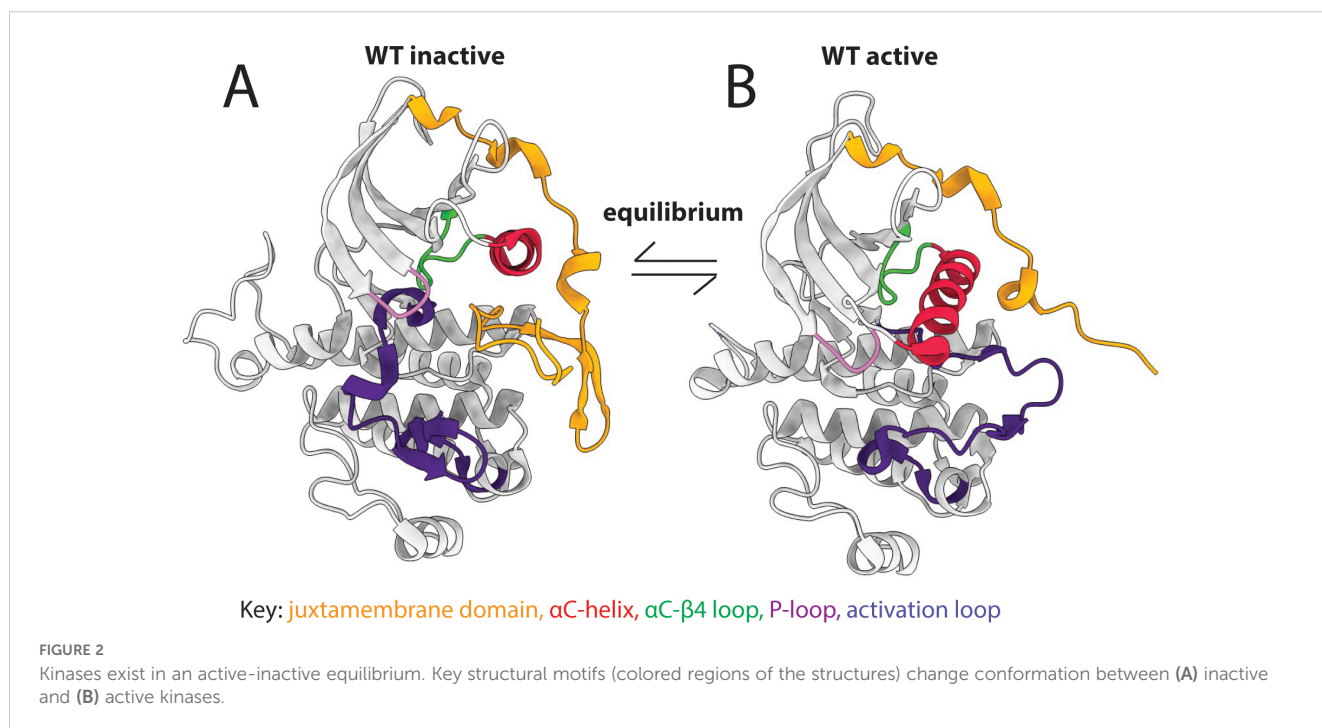
Receptor tyrosine kinases (RTKs) constitute an important family of oncogenic kinases. Endogenously, RTKs are activated by growth factor binding that stimulates complex and interconnected downstream signaling pathways that lead to cell growth and proliferation (Figure 1A) (31). In cancer, RTKs are frequently mutated to drive cancer growth (Figure 1B) and they are targets for the majority of kinase inhibitors (31). RTKs have four main domains: an extracellular ligand-binding domain, a transmembrane domain, a juxtamembrane (JM) domain, and a catalytic kinase domain (Figure 1B). They exist in equilibrium between active and



inactive states (Figure 2) (31, 32). Active and inactive states are structurally distinct, with key conserved motifs (motif = small structural element in a protein) rearranging in the active state to position catalytic residues in a geometry that enables catalysis (Figure 2) (24, 32). For example, the activation loop, which provides a platform for substrate binding, blocks the active site in the inactive conformation and extends out of the active site in the active conformation. Also during activation, a motif called the α C-

helix shifts inward to create a salt bridge that coordinates ATP for catalysis (Supplementary Figure S1). For a subset of kinases (e.g. KIT, PDGFRA, FLT3), the JM domain is known to be autoinhibitory, blocking the active site in the inactive conformation and moving away from the active site in the active conformation (28, 33, 34).

The equilibrium between inactive and active states is regulated by ligand binding and post-translational modifications, and can



also be modulated by oncogenic mutations that push the equilibrium of RTKs toward the active state through mechanisms that are sometimes, but not always, known (Figure 1B) (32). For example, in lung cancer, short in-frame deletions in the $\beta 3$ - αC loop of EGFR (exon 20) pull the αC -helix toward the active position, leading to kinase activation (35). In contrast, KIT D816V, FLT3 D835Y, and PDGFRA D842V are activation loop mutations commonly found in acute leukemias, gastrointestinal stromal tumors, and other cancer types (28, 36–38). They are highly oncogenic and are located in structurally analogous positions in their respective kinases, but their mechanisms of activation are not well understood.

Given the therapeutic importance of kinases in oncology, our deep understanding of kinase structure-function relationships, and the need for consistent data across mutations for NGS analysis, we generated a structural alignment of all missense mutations in the kinase and JM domains of 48 RTKs based on mutations reported in the Catalogue of Somatic Mutations in Cancer (COSMIC), the largest curated database of clinical cancer mutations (39). The results were functionally annotated and entered into a database that can be used as a tool to evaluate novel RTK missense mutations based on their structural position relative to known activating missense mutations. The database can be accessed either in full through GitHub (https://github.com/quantumdolphin/kinase_paper) or via a user-friendly web application (<https://kinase-mutation-atlas.streamlit.app>). We provide experimental data validating the use of this tool by showing several VUSs that structurally align with oncogenic mutations in other kinases are activating. Lastly, we show how this database can be used to develop testable mechanistic hypotheses regarding previously uncharacterized mutations in several kinase regions.

Materials and methods

Structural alignment and data collection

Forty eight RTKs that are mutated in cancer were selected for structural alignment (Supplementary Table S1, Figure 3A) (39). Thirty-six of the 48 kinases had kinase domain structures in the Protein Data Bank (40, 41) with <2.7 Å resolution, which we selected as a cut-off for adequate resolution per standard resolution guidelines (42). This resolution is sufficient to clearly determine the orientations of most protein side chains, which we used subsequently for clustering. Using a much more stringent cutoff would have resulted in eliminating all structures for several of the kinases. Many kinases had multiple structures therefore a representative structure was selected for each kinase that had (1): maximum coverage of kinase and JM domains, and (2) high quality (high resolution <2.7 Å and low R-free, a crystallographic quality metric). For the 12 kinases without structures, homology models were obtained from the Kinamatrix web server, which has been previously extensively validated for generating kinase homology models (43–45).

Mutational data and corresponding protein sequences were downloaded from COSMIC (39). UniProt numbering obtained from the PDBrenum server was used for residue numbering of the representative crystal structures (46, 47). The sequences were aligned between PDB and COSMIC database for each kinase to verify that data was mapped correctly from the COSMIC data to PDB structure. This was an important step to harmonize the residue number between the mutational and structural data.

To address structural differences in the 48 kinases, we limited our analysis to the set of residues that are well conserved across the kinases. The more highly variable regions in the JM domain and the

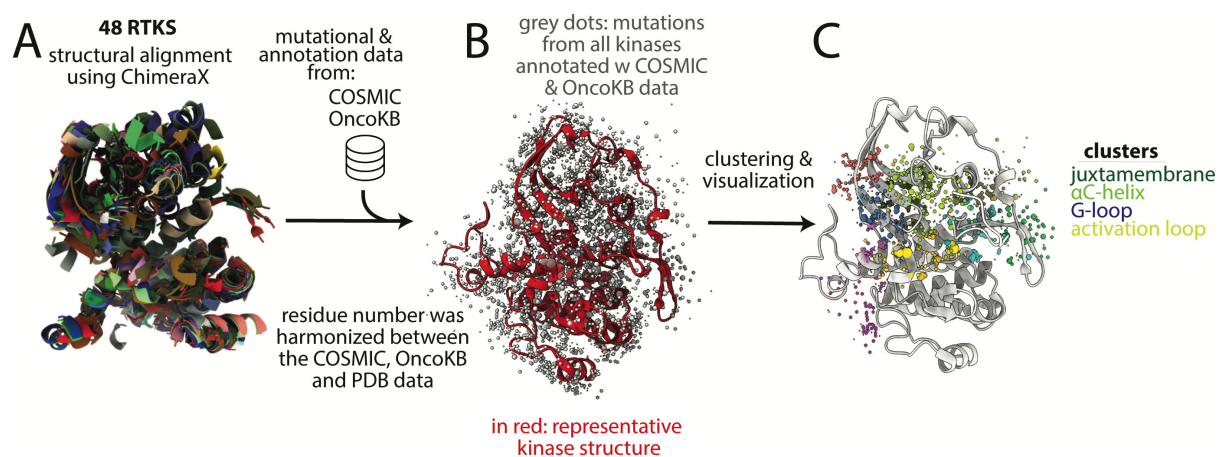


FIGURE 3

Basic methodology for structural alignment of mutations and clustering. (A) 48 kinase structures were aligned structurally using ChimeraX. (B) Mutational data from COSMIC and annotation data from OncoKB were added to the positions in the aligned structure. For clarity, only one kinase structure is shown (in red) after the initial multi-structure alignment; grey dots represent mutations. (C) Lastly, the datapoints were clustered based on proximity in 3-dimensional space, using the center of mass of the side chain for clustering. Colored dots are clusters of different mutations, with each group of different colored dots comprising a single cluster. Four of the clusters are labeled based on the motifs in which most mutations in the cluster are found; however these are not exact and the clusters overlap multiple motifs. Dots do not all sit on the backbone because only a single representative structure is shown and because the mutations are on sidechains, which are not depicted, for clarity. COSMIC is the mutational database used, OncoKB is the annotation database used.

activation loop were not included in the analysis. Kinase structures were aligned using the Matchmaker tool in ChimeraX with default parameters (48). This alignment was used to calculate the center of mass of side chains, which were then used as input for clustering (see below).

From the kinase mutational data downloaded from COSMIC, there were 20,772 residue positions that were affected by somatic cancer mutations. Of these, 5,778 residue positions were represented in our kinase structures (Supplementary Figure S2). Mutations at 300 of these residue positions had annotations in OncoKB, a precision oncology knowledge database that classifies mutations as gain-of-function (GOF), likely gain-of-function (LGOF), loss-of-function (LOF), likely loss-of-function (LLOF), neutral, and likely neutral (49, 50). The OncoKB annotations were added to the mutational information in the database and all missense mutations in the kinase and JM domains of the proteins were mapped to the aligned kinase structures (Figure 3B).

Structure-based clustering

Custom Python scripts and the MDAnalysis package were used for clustering and analysis (code and data: https://github.com/quantumdolphin/kinase_paper) (51, 52). Mutations from COSMIC were mapped to the three-dimensional structure using side chain center of masses for subsequent clustering. Hierarchical clustering was conducted using the Euclidean distance and Ward linkage method, with the X, Y, and Z coordinates of the side chain center of masses as input parameters (53, 54). The Euclidean distance, which is the shortest line between two points in any dimension, was selected as a parameter because it is commonly used when comparing spatial relationships in protein structures (54). Since clustering was based on three-dimensional side-chain center of masses, the Euclidean distance provided a readily-understandable geometric measure, related to commonly used metrics like root-mean-squared deviation. Ward linkage was chosen to minimize loss of information and within-cluster variance (53). Overall, the goal in selecting these parameters was to use a robust and reproducible method that aligns with best practices in structural biology. The combination of Euclidean distance and Ward linkage could provide spatially meaningful clusters that could be inspected visually and interpreted in the context of known kinase structure-function relationships.

The number of clusters was the other main parameter used. The goal was to group mutations in three-dimensional space to aid in visual inspection — not for classification based on the cluster membership. The dataset was split empirically into 40 clusters. We calculated the silhouette score (55) to investigate whether there was an optimal number of clusters. The silhouette score is a measure between -1 and 1 that indicates how well each point fits within its assigned cluster. A score close to 1 means the point is well-matched to its cluster and far from others (ideal). Scores near 0 suggest overlap between clusters, and negative scores suggest misclassification. In general, a higher silhouette score means better clustering. The variation in the silhouette score using different numbers of clusters was minimal (Supplementary

Figure S3) suggesting that there is no optimal number of clusters. Increasing beyond 40 clusters modestly increased the score, but at the cost of greatly increased complexity of analysis and interpretation. Empirically, using other numbers of clusters, between 20–40, did not qualitatively change the interpretation.

For further details on clustering see [Supplementary Methods](#).

Cluster analysis

The top 10 clusters that maximized GOF mutations and minimized LOF mutations were selected for further analysis (Supplementary Figure S4). ChimeraX was used for visualization (48). A user-friendly, publicly available web application was built using Streamlit that allows the user to input a kinase mutation and output a list of mutations in other kinases that are either structurally aligned or in close proximity to the initial mutation (<https://kinase-mutation-atlas.streamlit.app>) (56).

Sequence alignment

For visualization of the data, a multiple sequence alignment of kinase and JM domain sequences was generated using MUSCLE (Data Sheet 2, Appendix 1) (57). This sequence alignment was used to generate sequence logos for data interpretation in figures.

Ba/F3 cell assays

Ba/F3 cells were maintained in RPMI 1640 (Cellgro) supplemented with 1% penicillin-streptomycin (PS), murine IL-3 and 10% fetal bovine serum (FBS). pHAGE-PDGFRα (Addgene) and pLVX-Puro-FLT3-eBFP were used for generating described mutations using a site directed mutagenesis kit (Agilent) (19, 22). For primers, see [Supplementary Methods](#). HEK293T cells were transfected with 8 µg of either PDGFRα or FLT3 vector, 2 µg of pCMV-VSVG, and 4 µg of PAX2 vectors using the X-tremeGENE HP DNA transfection agent (Roche #6366244001). The viral supernatants were collected after 48 hours and used for Ba/F3 cell transduction. Three million cells were infected with lentiviral supernatant and polybrene (Santa Cruz Biotechnology #SC-134220) using spinfection followed by incubation for 5 hours at 37°C. After infection, cells were selected using 1 µg/ml puromycin (Invitrogen #ant-pr-1). Following selection, IL-3 withdrawal was performed, and viability assessed in IL-3 free media. All variants were tested in triplicate and repeated in 3 distinct biological replicates.

Results

Top mutation clusters are within key regulatory regions

After structurally aligning and clustering kinase and JM domain mutations from 48 RTKs, we found that the top 10 clusters of

mutations that maximized GOF and minimized LOF mutations were concentrated in key regulatory regions close to the kinase active site (Figure 3C). These regulatory regions include the activation loop, the glycine-rich loop (G-loop), the JM domain, the α C-helix, and the loops N-terminal and C-terminal to the α C-helix (the β 3- α C loop and the α C- β 4 loop). To demonstrate the utility of our database as a tool, we focused our analysis on a sample of mutations from several of the top clusters. The results below are organized based on the regulatory regions in which the sample mutations are found.

The activation loop is a key site of activating mutations

The activation loop recognizes and binds substrate, undergoing large conformational changes between inactive and active states, with phosphorylation stabilizing the active state. Several well-characterized GOF mutations were found within the activation loop (Figures 4A, B), consistent with known literature (28–30). Mutations in our database were absent from the N-terminal portion of the activation loop, called the DFG motif (Figure 4C), likely because the DFG is critical for catalysis and DFG mutations would be predicted to be LOF. However many mutations, including VUSs, were found in the ~10 amino acids immediately following the DFG, and to a lesser extent, portions of the protein surrounding/

contacting this area of the activation loop. Two aspartic acid (Asp, D) residues in the activation loop (positions 7 and 11, Figures 4B–D) had very high numbers of GOF mutations in the type 3 RTKs (28): FLT3, KIT, and PDGFRA (Figure 4D). These Asp residues sit on either end of a small helix known as a 3–10 helix, which we hypothesize may be involved in the mechanism of activation by the Asp mutations. The Asp at position 7 is the Asp^{B9} (FLT3 D835, KIT D816, and PDGFRA D842) which is recurrently mutated in many cancer types (28). Only the type 3 RTKs had large numbers of mutations that structurally aligned to this site (Figure 4D), however there were small numbers of mutations at this position in several other kinases, including KDR D1052X mutations that have been shown to increase KDR activity in a tissue culture model (58). Though these KDR mutations are not annotated as GOF in OncoKB, we hypothesize that based on their structural alignment with other highly oncogenic, recurrent mutations in KIT, PDGFRA, and FLT3, that they are in fact GOF.

The Asp (D) at the C-terminal end of the 3–10 helix in KIT, KDR, PDGFRA, and FLT3 (Figure 4B, position 11) is also frequently mutated, but less well characterized than Asp^{B9}. We hypothesize that most mutations at this position will be activating by disrupting key electrostatic interactions (dotted lines between R815 and D820 in the case of KIT shown in Figure 4B) that hold the activation loop in the inactive conformation. Additionally, there is a tyrosine (Tyr, Y) (position 14) phosphorylation site C-terminal on the activation loop that is highly mutated. Given that this Tyr is a

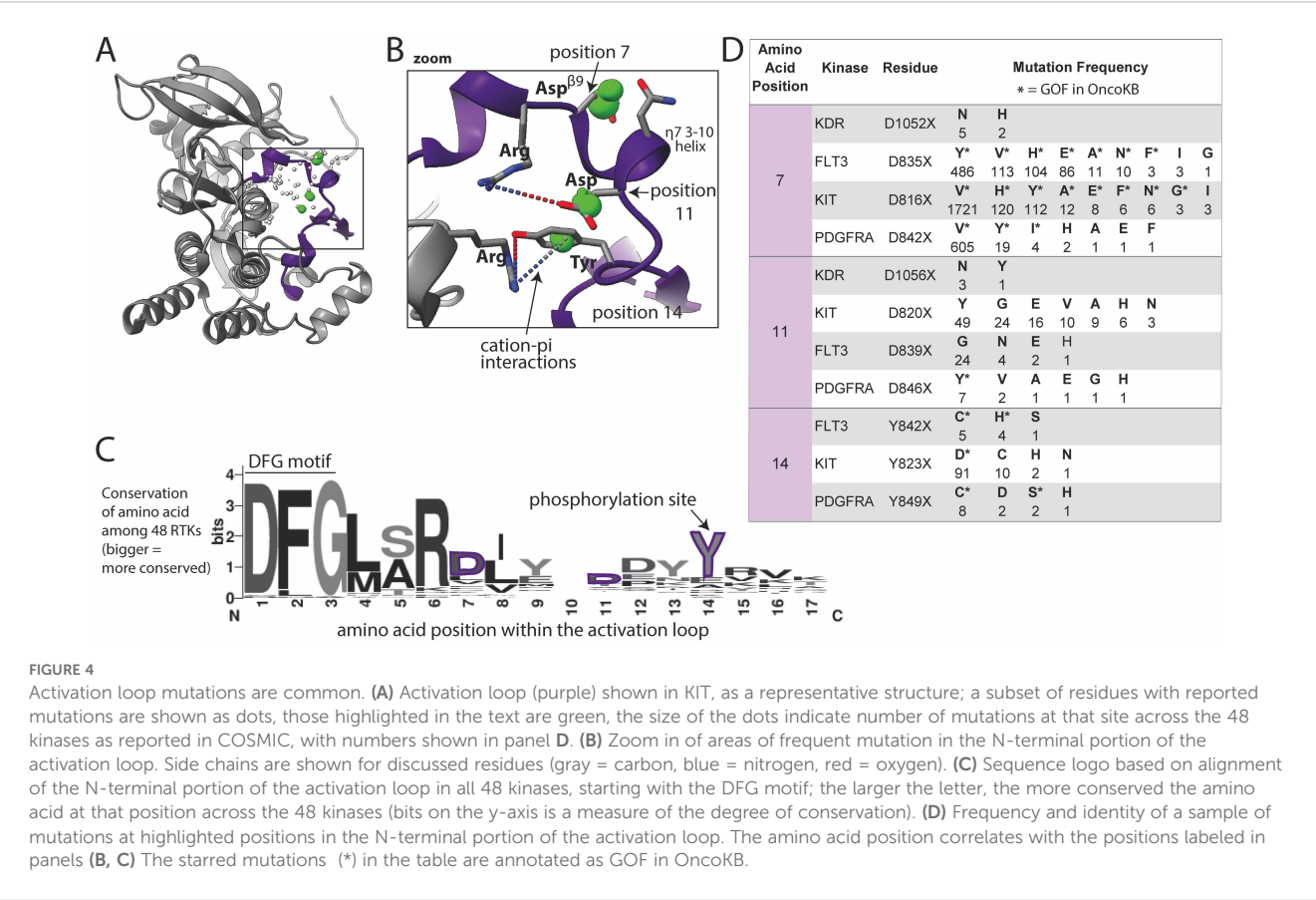


FIGURE 4 Activation loop mutations are common. (A) Activation loop (purple) shown in KIT, as a representative structure; a subset of residues with reported mutations are shown as dots, those highlighted in the text are green, the size of the dots indicate number of mutations at that site across the 48 kinases as reported in COSMIC, with numbers shown in panel D. (B) Zoom in of areas of frequent mutation in the N-terminal portion of the activation loop. Side chains are shown for discussed residues (gray = carbon, blue = nitrogen, red = oxygen). (C) Sequence logo based on alignment of the N-terminal portion of the activation loop in all 48 kinases, starting with the DFG motif; the larger the letter, the more conserved the amino acid at that position across the 48 kinases (bits on the y-axis is a measure of the degree of conservation). (D) Frequency and identity of a sample of mutations at highlighted positions in the N-terminal portion of the activation loop. The amino acid position correlates with the positions labeled in panels (B, C). The starred mutations (*) in the table are annotated as GOF in OncoKB.

phosphorylation site, mutations may be phospho-mimetics as several of the mutant residues are Asp, which has a negative charge resembling a phosphate group.

Mutations in the G-loop may impact function by changing loop flexibility

A large number of mutations were located around the G-loop, which is characterized by the consensus sequence (GXGXΦG), where Φ is usually a Tyr or phenylalanine (Phe, F) (Figures 5A–D) (24, 59). The G-loop is highly flexible, positioning ATP for phosphoryl transfer in the active site.

Mutations reported in COSMIC within the G-loop are almost exclusively at the first and last glycine (Gly, G), positions 7 and 12, at the two ends of the loop (Figure 5B). Though mutations in EGFR at these positions are well characterized as GOF, there are uncharacterized mutations in at least 10 other RTKs at these positions (Figure 5D) (60). Gly, lacking a side chain, plays a unique role in protein flexibility, and the first and last Gly in the G-loop act as hinge points for the movement of the loop. Mutating these to any other amino acid is predicted to decrease loop flexibility, potentially locking it into an active-like conformation. Thus, we hypothesize that all mutations of these 2 critical glycines could impact loop flexibility

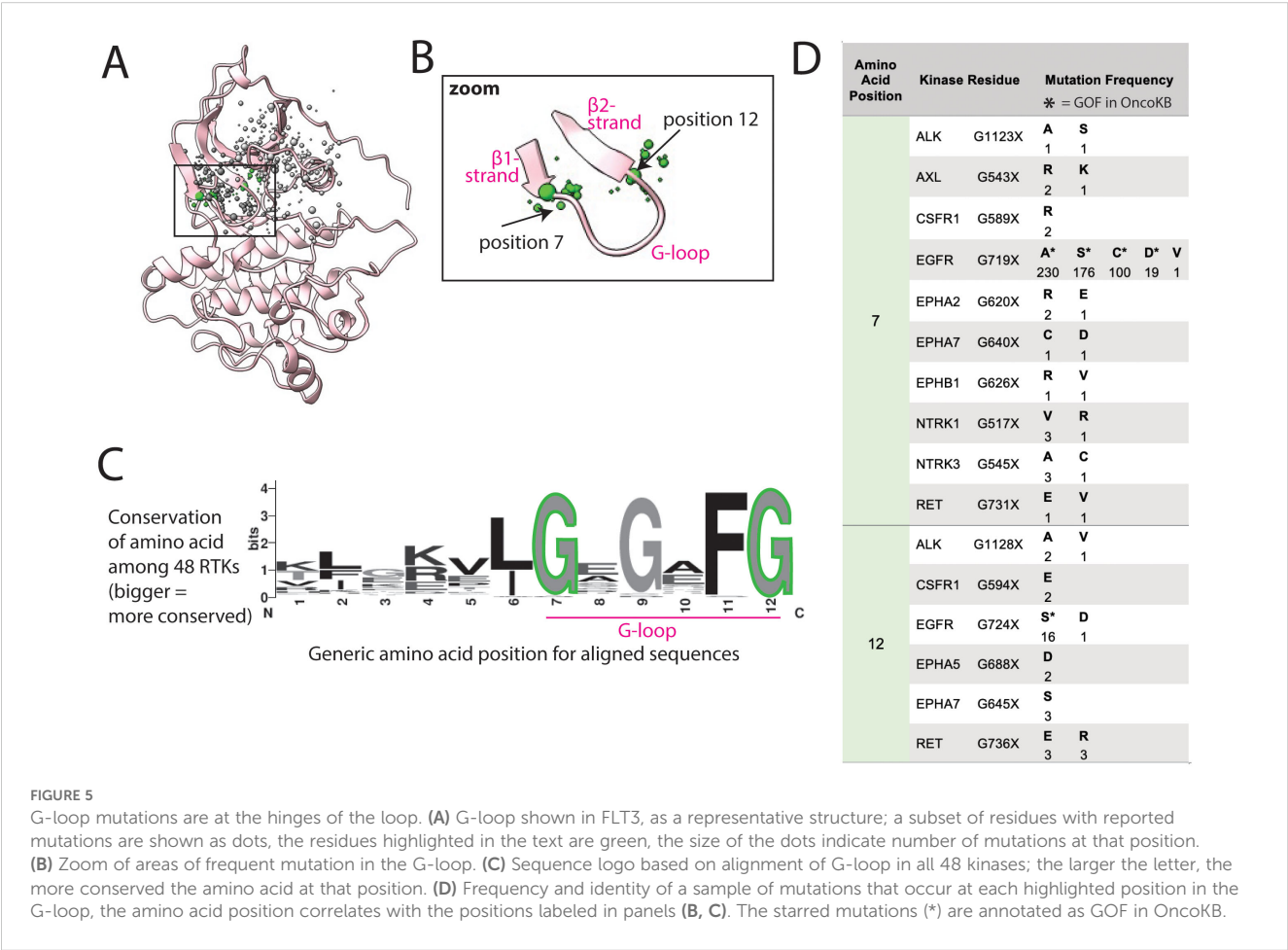
and thus activation, as previously demonstrated for G-loop mutations in EGFR in lung adenocarcinoma (61).

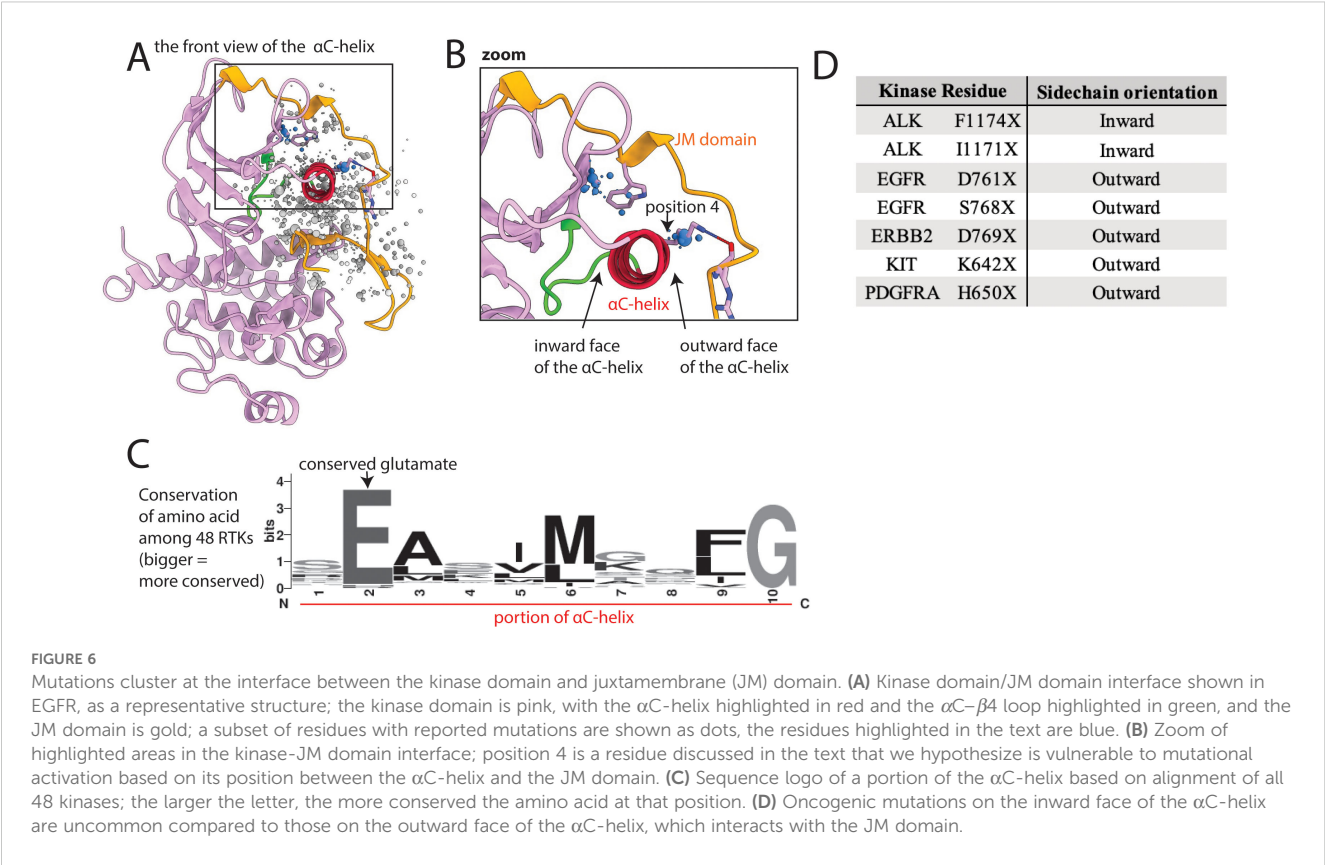
Within the same cluster, there are also numerous mutations in the β₁ and β₂ strands of the N-lobe flanking the G-loop, which could also plausibly, but more speculatively, be activating through modifying the structure and/or dynamics of the G-loop.

Mutations in the JM domain likely disrupt autoinhibitory interactions with the kinase domain

The αC-helix, which abuts the JM domain in some kinases, is important in kinase regulation and undergoes significant conformational change between active and inactive states (Figure 2) (24). The αC-helix contains a conserved glutamate (Glu, E) that faces inward and helps align the catalytic lysine (Lys, K) necessary for phosphoryl transfer (Supplementary Figure S1, Figure 6). This conserved Glu, and other amino acids on the inward-facing side of the αC-helix, have few mutations. However, mutations of outward-facing amino acids on the αC-helix that interact with the JM domain are more common (Figures 6A–D).

The JM domain has autoinhibitory interactions with the kinase domain in the inactive state, especially in type 3 RTKs. There are





numerous well-characterized JM domain GOF mutations that abolish JM-to-kinase domain autoinhibitory interactions, such as exon 11 mutations in KIT or exon 14 or 15 mutations in FLT3 (25–27). We found 2 large areas of mutations involving the JM domain and the contacting kinase domain, including outward-facing portions of the α C-helix (Figures 6A, B). KIT K642E is an example of a characterized GOF mutation previously identified on the outward face of the α C-helix (62). K642 in wildtype KIT forms a hydrogen bond with the backbone of the JM domain; therefore mutating it would break this interaction. We hypothesize that mutations in the analogous Lys of other kinases, such as EGFR shown in Figure 6B, position 4 will also be activating. Many other mutations in the interface between the JM and kinase domains may also be GOF by disrupting autoinhibition. For example, we previously reported that a kinase domain VUS, FLT3 A680V, which sits at this interface was activating (19). As this mutation is on the interface between JM and kinase domains, the mechanism of activation may involve disruption of autoinhibition.

Mutations in the α C- β 4 loop impact kinase function through allostery

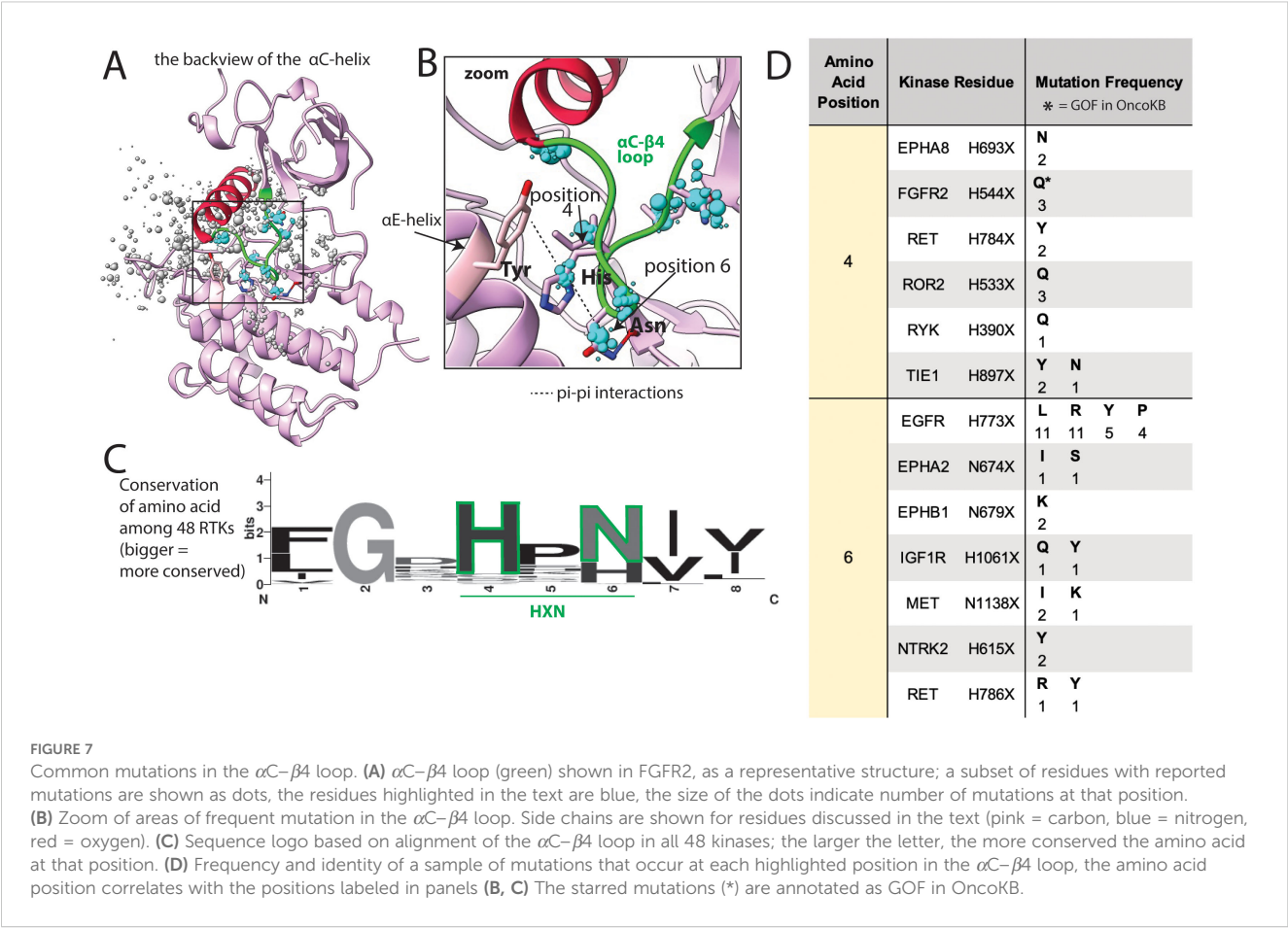
In addition to mutations on portions of the α C-helix interacting with the JM domain, there is also a large number of mutations in the 2 loops flanking the α C-helix. These may lead to GOF by changing the orientation of the helix. On one side of the α C-helix is the β 3- α C loop, which harbors well characterized mutations in BRAF,

EGFR, and HER2 (35). On the other side of the helix is the α C- β 4 loop (Figures 7A–C) which is involved in long-range allosteric coupling with the catalytic kinase core by orchestrating movement of the α C-helix with kinase activation (63). This α C- β 4 loop has a large number of mutations in COSMIC, though relatively few have been characterized as activating (Figure 7D).

Uncharacterized mutations that align with known oncogenic mutations are activating

We hypothesized that our database of structurally aligned mutations could be used to help identify previously uncharacterized mutations as activating. A set of 5 PDGFRA mutations and 5 structurally analogous FLT3 mutations, including 2 activation loop mutations, 2 α C- β 4 loop mutations, and 6 JM domain mutations that interact with the kinase domain at different locations were selected for experimental validation (Figures 8A, B). These mutants were expressed in Ba/F3 cells, which become cytokine-independent in the presence of an activated kinase and have been used extensively to evaluate for kinase activation (64). All selected mutations were within key regulatory regions, though only 3 (PDGFRA Y849C, V561D, Y555C) were designated GOF in OncoKB (Figure 8B). PDGFRA D842V and FLT3 D835Y mutations were used as positive controls, and wildtype PDGFRA and FLT3 were used as negative controls.

Two of the selected FLT3 mutations were designated LGOF in OncoKB but had been previously shown to be activating in tissue



culture models. These mutations include, 1) the activation loop mutation FLT3 Y842C, which is structurally analogous to the PDGFRA GOF mutation Y849C, and, 2) FLT3 Y572C, which is structurally analogous to the PDGFRA GOF mutation Y555C (65–67). All of these mutations were also activating in our Ba/F3 assays (Figures 8C, D).

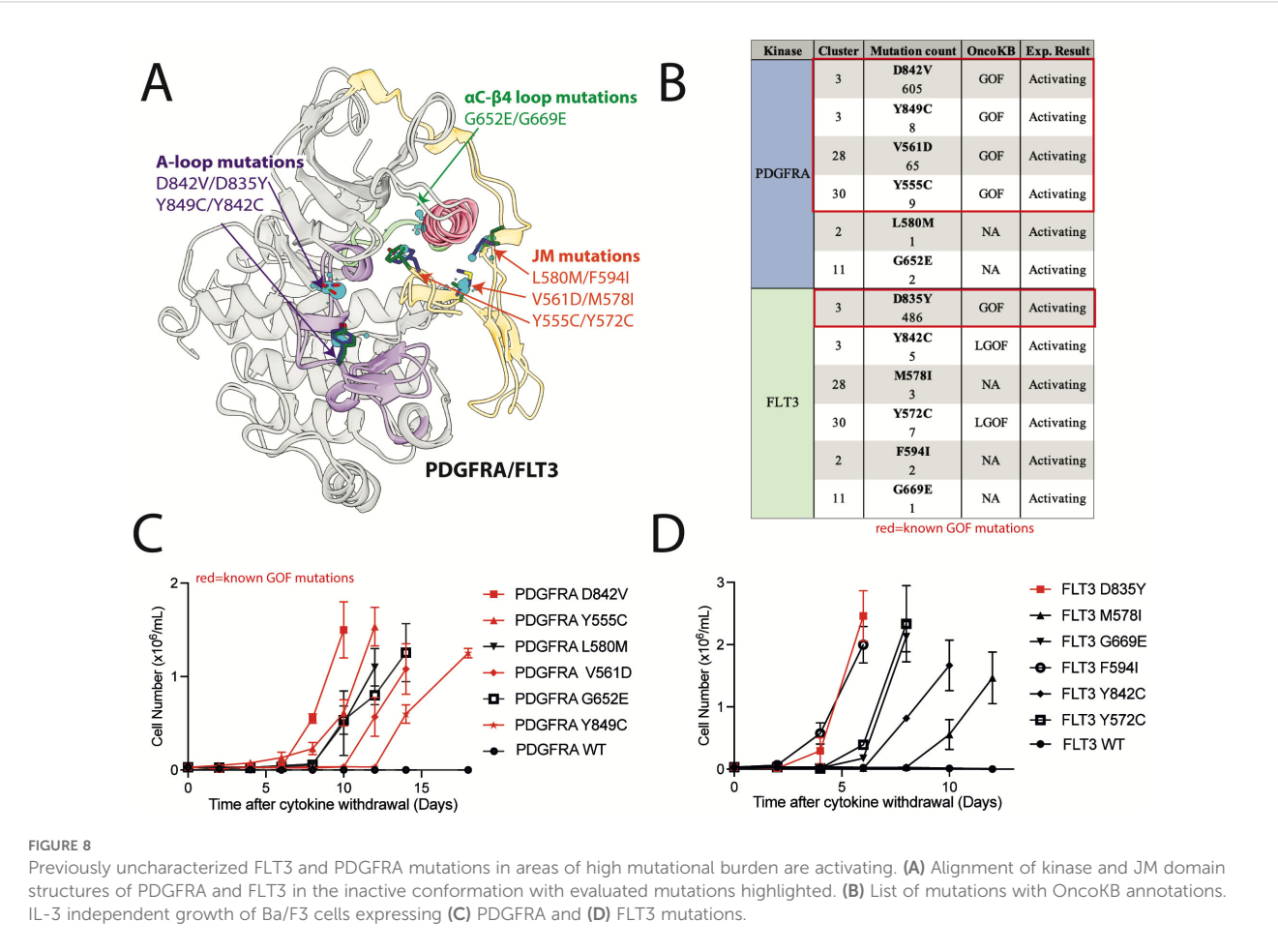
Two of the selected PDGFRA mutations (L580M and G652E) and 3 of the selected FLT3 mutations (M578I, F594I, and G669E) had no annotations in OncoKB and were not characterized in the literature. However, these mutations were in top clusters and within regulatory regions of high mutational burden. When expressed in Ba/F3 cells, we found all 5 mutations were activating (Figures 8C, D). These data support the use of this database as a resource in the analysis of kinase mutations found in clinical NGS.

Discussion

With the explosion of next-generation sequencing data and the concurrent availability of protein structures through both the Protein Data Bank and models generated by programs such as AlphaFold, we have an unprecedented opportunity to use protein structure to develop tools to understand the biological impact of clinical cancer mutations (20, 41). Here, we describe the generation of a database of structurally aligned and functionally annotated

cancer mutations across 48 kinases that can be used as a resource to analyze kinase VUSs, with potential clinical implications to guide targeted therapy. Though this tool has not yet been prospectively evaluated and is not intended to be used as a sole predictor of functional significance for clinical kinase mutations, it provides concrete experimental data across mutations and kinases that can aid clinical NGS analysis. We have focused our analysis on clusters of mutations centered on 4 regions of the kinase domain that are known to be important for regulating activity, and thus enriched in both known GOF mutations and numerous VUSs that we propose are also likely to be activating:

1. Activation loop. Here, we found mutations absent from the N-terminal portion of the activation loop, called the DFG motif. In the Ser/Thr kinase, BRAF, oncogenic, inactivating DFG mutations are well described (68, 69). These DFG-mutant BRAF proteins activate cell signaling by causing the inactive BRAF to heterodimerize and transactivate CRAF (68, 69). However, inactivating DFG mutations that activate oncogenic signaling have not been described in other kinases. The lack of DFG mutations in our database was expected given that mutation of the DFG prevents kinase catalytic function. In contrast, mutations of Asp^{β9}, which occupies a central position on the activation loop and is highly mutated in cancer (e.g. KIT D816V, PDGFRA



D842V, FLT3 D835Y, position 7 in Figure 4), comprised one of the top mutation clusters, providing proof of principle that structural areas of high mutational burden across kinases harbor oncogenic mutations. Asp^{β9} mutations that introduce hydrophobic or aromatic side chains (Val, Tyr, Ala, Phe, Ile) are known to be activating (22, 70, 71). Even a conservative Asp to Glu substitution, where the Glu differs from the wildtype Asp by only a single methyl group, can be activating (e.g. PDGFRA D842E, FLT3 D835E) indicating precise charge placement is required to stabilize the inactive state (22, 71). Though the mechanism by which Asp^{β9} mutations activate kinases is not well understood, our analysis of the structures of multiple kinases simultaneously suggests the orientation and position of the Asp^{β9} may stabilize the charge distribution (or the dipole moment) of the 3–10 helix (Figure 4B), which could preserve the inactive conformation. Therefore, mutation of the Asp^{β9} could shift the charge distribution and destabilize the inactive state. Several other mutations in the activation loop C-terminal to the Asp^{β9} likely have a similar mechanism, i.e., destabilizing the inactive state, such as the Y842C mutation in FLT3, which we confirm to be activating in Ba/F3 cells.

2. G-loop. Mutations in this key structural and functional element have been previously described in EGFR in non-small cell lung cancer, in BCR::ABL1 in CML, and in BRAF in colon cancer (60, 72, 73). Interestingly, the EGFR and BCR::ABL1 mutations confer poor prognosis with aggressive, drug-resistant disease (60, 72). Other G-loop mutations have been reported, but with unknown prognostic significance. For example, ALK G1128A, a G-loop mutation studied in the context of neuroblastoma, has intermediate oncogenic activity compared to other ALK mutations, but there is not enough data to determine prognostic significance (23). Our analysis showed mutations within the G-loop, including the EGFR and ALK mutations discussed above, are located almost exclusively at the first and last Gly residues at the two ends of the loop. Analogous mutations have been reported in at least 10 other RTKs, including RET, NTRK1/3, and several ephrin receptors, albeit with low prevalence (Figure 5D). We hypothesize that these increase kinase activity by rigidifying the loop at the critical glycine hinge-points, potentially locking the loop into an active-like conformation. There are also numerous mutations in the short β₁ and β₂ strands that flank the G-loop. It is possible

that some of these could also be activating through modulating the structure and/or dynamics of the G-loop, but little information is currently available to support this hypothesis.

3. JM domain. Interactions between the JM and kinase domains stabilize the autoinhibited states, and thus mutations in amino acids involved in these interactions are frequently activating. Such mutations include well-characterized GOF mutations on the JM domain itself, such as Y555C in PDGFRA, and mutations on the 'outward-facing' portion of the α C-helix that interact with the JM domain, such as K642E in KIT. We identified additional clusters of uncharacterized mutations at the interface between the JM and α C-helix that we postulated would also destabilize the autoinhibited state and confirmed that L580M in PDGFRA and M578I and F549I in FLT3 were in fact activating in Ba/F3 cells.
4. The 2 loops flanking the α C-helix, the β 3- α C loop and the α C- β 4 loop. It has been shown that small deletions in the β 3- α C loop change the orientation of the α C-helix, leading to activation (35). We hypothesize that missense mutations in the α C- β 4 loop also alter the orientation of the α C-helix and can cause kinase activation. This hypothesis is based on literature showing the α C- β 4 loop (Figures 7A–C) is involved in long-range allosteric coupling with the catalytic kinase core and can orchestrate movement of the α C-helix with kinase activation (63). In particular, the HXN motif (X=any amino acid) is the region of the α C- β 4 loop that forms the pivot point for α C-helix movement (Figure 7B). GOF mutations have been identified in the HXN as well as in other portions of the α C- β 4 loop. Substitution of the histidine (His) within the HXN, such as the GOF mutation FGFR2 H544Q (Figures 7B, C, position 4), eliminates a favorable pi-pi stacking interaction between the His and a conserved Tyr in the α E-helix (Figure 7B), potentially altering the structure of the inactive state (74). In addition to FGFR2, mutations involving the His have been reported in EPHA8, RET, ROR2, RYK, and TIE1 (Figure 7D), and we predict these are similarly activating by eliminating pi-pi stacking interactions and changing the conformation of the α C- β 4 loop and the α C-helix. The Asn (Figures 7B, C, position 6) of the HXN motif is less well conserved but the Asn side chain is hypothesized to be part of a water-mediated hydrogen bond network aiding allosteric communication between the α C-helix and the catalytic core (75). Substitution of Asn with amino acids that change the hydrogen bonding network can affect allosteric signaling and impact kinase activation. Consistent with this assertion, EGFR has a known GOF mutation (EGFR H773L) in this position that eliminates hydrogen-bonding and has been shown to be sensitive to EGFR kinase inhibitors in non-small cell lung cancer (76). IGF1R, NTRK2, RET, EPHA2, EPHB1, and MET also have mutations in this position, which we predict to be activating (Figure 7D). Overall, the α C- β 4 loop appears to be highly

susceptible to GOF mutations, and most mutations in this loop, even seemingly conservative ones, we hypothesize to be likely GOF. To begin to explore this hypothesis, we tested VUSs in PDGFRA and FLT3 at another position in the loop that is 2 residues N-terminal to the HXN motif, and both (G652E and G669E, respectively) were activating in Ba/F3 cells.

In summary, there are many kinase VUSs that align structurally with characterized kinase mutations. Given the complexity of trying to predict protein function based on point mutations, oncologists and tumor boards are likely to continue relying on integrated information from multiple sources to interpret clinical NGS and make molecularly targeted therapy recommendations. Here, we provide a database of structurally aligned and functionally annotated mutations across 48 RTKs that can be used as a consistent resource across RTKs to evaluate VUSs for structural alignment with known activating mutations, while also providing insight into potential mechanisms of activation and regulation. Future work will include more experimental evaluation of the dataset, including evaluating mutations spatially far from known oncogenic mutations to establish specificity of the dataset. Overall, we expect our database to be an important addition to the current tools and resources used to analyze clinical NGS.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

AR: Data curation, Writing – original draft, Methodology, Validation, Visualization, Investigation, Software, Project administration, Resources, Writing – review & editing, Supervision, Formal analysis, Conceptualization. IS: Writing – review & editing, Visualization, Data curation. EG: Writing – review & editing, Project administration, Visualization. YP: Formal analysis, Project administration, Investigation,

Visualization, Writing – review & editing, Data curation, Resources, Funding acquisition, Validation, Methodology, Supervision, Conceptualization, Writing – original draft, Software. MJ: Project administration, Formal analysis, Supervision, Methodology, Software, Data curation, Validation, Visualization, Writing – original draft, Resources, Funding acquisition, Writing – review & editing, Conceptualization, Investigation. BA: Software, Conceptualization, Writing – review & editing, Resources, Visualization, Investigation, Funding acquisition, Writing – original draft, Project administration, Methodology, Validation, Formal analysis, Data curation, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is generously supported by Julia's Legacy of Hope and the St. Baldrick's Foundation Consortium Research Grant (YP) and the Bachrach Family Foundation (BAW).

Conflict of interest

MJ has been a consultant to Schrodinger Inc. which licenses software that can be used to visualize protein structures. MJ is a co-founder of, consultant to, and shareholder of Relay Therapeutics, Initial Therapeutics, Terremoto Biosciences, Circle Pharma, and

Cedilla Therapeutics, which are involved in drug discovery for oncology. BA receives clinical research funding from Novartis.

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

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2025.1599389/full#supplementary-material>.

References

- Roskoski R Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2024 update. *Pharmacol Res.* (2024) 200:107059. doi: 10.1016/j.phrs.2024.107059
- Etienne G, Guilhot J, Rea D, Rigal-Huguet F, Nicolini F, Charbonnier A, et al. Long-term follow-up of the french stop imatinib (STIM1) study in patients with chronic myeloid leukemia. *J Clin Oncol.* (2017) 35:298–305. doi: 10.1200/JCO.2016.68.2914
- Cameron LB, Hitchen N, Chandran E, Morris T, Manser R, Solomon BJ, et al. Targeted therapy for advanced anaplastic lymphoma kinase (<I>ALK</I>)-rearranged non-small cell lung cancer. *Cochrane Database Syst Rev.* (2022) 1:CD013453. doi: 10.1002/14651858.CD013453.pub2
- Schwartzberg L, Kim ES, Liu D, Schrag D. Precision oncology: who, how, what, when, and when not? *Am Soc Clin Oncol Educ Book.* (2017) 37:160–9. doi: 10.1200/EDBK_174176
- Spielmann M, Kircher M. Computational and experimental methods for classifying variants of unknown clinical significance. *Cold Spring Harb Mol Case Stud.* (2022) 8. doi: 10.1101/mcs.a006196
- National Cancer Institute. Available online at: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/variant-of-uncertain-significance>. (Accessed July 8, 2025).
- Garcia FAO, de Andrade ES, Palmero EI. Insights on variant analysis in silico tools for pathogenicity prediction. *Front Genet.* (2022) 13:1010327. doi: 10.3389/fgene.2022.1010327
- Rodrigues CH, Ascher DB, Pires DE. Kinact: a computational approach for predicting activating missense mutations in protein kinases. *Nucleic Acids Res.* (2018) 46:W127–W32. doi: 10.1093/nar/gky375
- Kuntz CP, Woods H, McKee AG, Zelt NB, Mendenhall JL, Meiler J, et al. Towards generalizable predictions for G protein-coupled receptor variant expression. *Biophys J.* (2022) 121:2712–20. doi: 10.1016/j.bpj.2022.06.018
- Johnson A, Ng PK, Kahle M, Castillo J, Amador B, Wang Y, et al. Actionability classification of variants of unknown significance correlates with functional effect. *NPJ Precis Oncol.* (2023) 7:67. doi: 10.1038/s41698-023-00420-w
- Riccio C, Jansen ML, Guo L, Ziegler A. Variant effect predictors: a systematic review and practical guide. *Hum Genet.* (2024) 143:625–34. doi: 10.1007/s00439-024-02670-5
- Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat.* (2013) 34:57–65. doi: 10.1002/humu.22225
- Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun.* (2020) 11:5918. doi: 10.1038/s41467-020-19669-x
- Lopez-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpi JL. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res.* (2017) 45:W222–W8. doi: 10.1093/nar/gkx313
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* (2010) 7:248–9. doi: 10.1038/nmeth0410-248
- Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the association for molecular pathology, american society of clinical oncology, and college of american pathologists. *J Mol Diagn.* (2017) 19:4–23. doi: 10.1016/j.jmoldx.2016.10.002
- Qorri E, Takacs B, Graf A, Enyedi MZ, Pinter L, Kiss E, et al. A comprehensive evaluation of the performance of prediction algorithms on clinically relevant missense variants. *Int J Mol Sci.* (2022) 23. doi: 10.3390/ijms23147946
- Keller RB, Mazor T, Sholl L, Aguirre AJ, Singh H, Sethi N, et al. Programmatic precision oncology decision support for patients with gastrointestinal cancer. *JCO Precis Oncol.* (2023) 7:e2200342. doi: 10.1200/PO.22.00342
- Pikman Y, Tasian SK, Sulis ML, Stevenson K, Blonquist TM, Apsel Winger B, et al. Matched targeted therapy for pediatric patients with relapsed, refractory, or high-risk leukemias: A report from the LEAP consortium. *Cancer Discov.* (2021) 11:1424–39. doi: 10.1158/2159-8290.CD-20-0564

20. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. (2021) 596:583–9. doi: 10.1038/s41586-021-03819-2
21. Thornton JM, Laskowski RA, Borkakoti N. AlphaFold heralds a data-driven revolution in biology and medicine. *Nat Med*. (2021) 27:1666–9. doi: 10.1038/s41591-021-01533-0
22. Paolino J, Dimitrov B, Apse Winger B, Sandoval-Perez A, Rangarajan AV, Ocasio-Martinez N, et al. Integration of genomic sequencing drives therapeutic targeting of PDGFRA in T-cell acute lymphoblastic leukemia/lymphoblastic lymphoma. *Clin Cancer Res*. (2023) 29:4613–26. doi: 10.1158/1078-0432.CCR-22-2562
23. Bresler SC, Weiser DA, Huwe PJ, Park JH, Krytska K, Ryles H, et al. ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. *Cancer Cell*. (2014) 26:682–94. doi: 10.1016/j.ccr.2014.09.019
24. Huse M, Kuriyan J. The conformational plasticity of protein kinases. *Cell*. (2002) 109:275–82. doi: 10.1016/S0092-8674(02)00741-9
25. Hirota S, Isozaki K, Moriyama Y, Hashimoto K, Nishida T, Ishiguro S, et al. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science*. (1998) 279:577–80. doi: 10.1126/science.279.5350.577
26. Nakao M, Yokota S, Iwai T, Kaneko H, Horiike S, Kashima K, et al. Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia*. (1996) 10:1911–8.
27. Sakurai S, Fukasawa T, Chong JM, Tanaka A, Fukayama M. C-kit gene abnormalities in gastrointestinal stromal tumors (tumors of interstitial cells of Cajal). *Jpn J Cancer Res*. (1999) 90:1321–8. doi: 10.1111/j.1349-7006.1999.tb00715.x
28. Klug LR, Kent JD, Heinrich MC. Structural and clinical consequences of activation loop mutations in class III receptor tyrosine kinases. *Pharmacol Ther*. (2018) 191:123–34. doi: 10.1016/j.pharmthera.2018.06.016
29. Yun CH, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, et al. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell*. (2007) 11:217–27. doi: 10.1016/j.ccr.2006.12.017
30. Kiel C, Benisty H, Llorens-Rico V, Serrano L. The yin-yang of kinase activation and unfolding explains the peculiarity of Val600 in the activation segment of BRAF. *Elife*. (2016) 5:e12814. doi: 10.7554/eLife.12814
31. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell*. (2010) 141:1117–34. doi: 10.1016/j.cell.2010.06.011
32. Johnson LN, Noble ME, Owen DJ. Active and inactive protein kinases: structural basis for regulation. *Cell*. (1996) 85:149–58. doi: 10.1016/S0092-8674(00)81092-2
33. Mol CD, Lim KB, Sridhar V, Zou H, Chien EY, Sang BC, et al. Structure of a c-kit product complex reveals the basis for kinase transactivation. *J Biol Chem*. (2003) 278:31461–4. doi: 10.1074/jbc.C300186200
34. Liang L, Yan XE, Yin Y, Yun CH. Structural and biochemical studies of the PDGFRA kinase domain. *Biochem Biophys Res Commun*. (2016) 477:667–72. doi: 10.1016/j.bbrc.2016.06.117
35. Foster SA, Whalen DM, Ozen A, Wongchenko MJ, Yin J, Yen I, et al. Activation mechanism of oncogenic deletion mutations in BRAF, EGFR, and HER2. *Cancer Cell*. (2016) 29:477–93. doi: 10.1016/j.ccr.2016.02.010
36. Heinrich MC, Corless CL, Duensing A, McGreevey L, Chen CJ, Joseph N, et al. PDGFRA activating mutations in gastrointestinal stromal tumors. *Science*. (2003) 299:708–10. doi: 10.1126/science.1079666
37. Thiede C, Steudel C, Mohr B, Schaich M, Schakel U, Platzbecker U, et al. Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood*. (2002) 99:4326–35. doi: 10.1182/blood.V99.12.4326
38. Nagata H, Worobec AS, Oh CK, Chowdhury BA, Tannenbaum S, Suzuki Y, et al. Identification of a point mutation in the catalytic domain of the protooncogene c-kit in peripheral blood mononuclear cells of patients who have mastocytosis with an associated hematologic disorder. *Proc Natl Acad Sci U S A*. (1995) 92:10560–4. doi: 10.1073/pnas.92.23.10560
39. Sondka Z, Dhiri NB, Carvalho-Silva D, Jupp S, Madhumita, McLaren K, et al. COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res*. (2024) 52:D1210–D7. doi: 10.1093/nar/gkad986
40. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res*. (2000) 28:235–42. doi: 10.1093/nar/28.1.235
41. Protein Data Bank. Available online at: <https://www.rcsb.org>. (Accessed July 8, 2025).
42. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J*. (2008) 275:1–21. doi: 10.1111/j.1742-4658.2007.06178.x
43. Ung PM, Rahman R, Schlessinger A. Redefining the protein kinase conformational space with machine learning. *Cell Chem Biol*. (2018) 25:916–24 e2. doi: 10.1016/j.chembiol.2018.05.002
44. Ung PM, Schlessinger A. DFGmodel: predicting protein kinase structures in inactive states for structure-based discovery of type-II inhibitors. *ACS Chem Biol*. (2015) 10:269–78. doi: 10.1021/cb500696t
45. Rahman R, Ung PM, Schlessinger A. KinaMetrix: a web resource to investigate kinase conformations and inhibitor space. *Nucleic Acids Res*. (2019) 47:D361–D6. doi: 10.1093/nar/gky916
46. UniProt C. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res*. (2023) 51:D523–D31. doi: 10.1093/nar/gkac1052
47. PDBrenum. Available online at: <http://dunbrack.fccc.edu/PDBrenum/>. (Accessed July 8, 2025).
48. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*. (2021) 30:70–82. doi: 10.1002/pro.3943
49. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: A precision oncology knowledge base. *JCO Precis Oncol*. (2017) 2017. doi: 10.1200/PO.17.00011
50. Murciano-Goroff YR, Suehnholz SP, Drilon A, Chakravarty D. Precision oncology: 2023 in review. *Cancer Discov*. (2023) 13:2525–31. doi: 10.1158/2159-8290.CD-23-1194
51. Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, Domanski J, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. of the 15th Python in Science Conf*. (2016) 98–105. doi: 10.25080/Majors-629e541a-00e
52. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. (2011) 32:2319–27. doi: 10.1002/jcc.21787
53. Fernandes AAR, Solimun S. Comparison of the use of linkage in cluster integration with path analysis approach. *Front Appl Mathematics Stat*. (2022) 8. doi: 10.3389/fams.2022.790010
54. Liberti L, Lator C, Maculan N, Mucherino A. Euclidean distance geometry and applications. *SIAM Review*. (2014) 56:3–69. doi: 10.1137/120875909
55. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Mathematics*. (1987) 20:53–65. doi: 10.1016/0377-0427(87)90125-7
56. Streamlit. Available online at: <https://streamlit.io>. (Accessed July 8, 2025).
57. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. (2004) 32:1792–7. doi: 10.1093/nar/gkh340
58. Grillo E, Corsini M, Ravelli C, di Somma M, Zammataro L, Monti E, et al. A novel variant of VEGFR2 identified by a pan-cancer screening of recurrent somatic mutations in the catalytic domain of tyrosine kinase receptors enhances tumor growth and metastasis. *Cancer Lett*. (2021) 496:84–92. doi: 10.1016/j.canlet.2020.09.027
59. Taylor SS, Meharena HS, Kornev AP. Evolution of a dynamic molecular switch. *IUBMB Life*. (2019) 71:672–84. doi: 10.1002/iub.2059
60. Robichaux JP, Le X, Vijayan RSK, Hicks JK, Heeke S, Elamin YY, et al. Structure-based classification predicts drug response in EGFR-mutant NSCLC. *Nature*. (2021) 597:732–7. doi: 10.1038/s41586-021-03898-1
61. Fassunke J, Muller F, Keul M, Michels S, Dammert MA, Schmitt A, et al. Overcoming EGFR(G724S)-mediated osimertinib resistance through unique binding characteristics of second-generation EGFR inhibitors. *Nat Commun*. (2018) 9:4655. doi: 10.1038/s41467-018-07078-0
62. Lux ML, Rubin BP, Biase TL, Chen CJ, Maclure T, Demetri G, et al. KIT extracellular and kinase domain mutations in gastrointestinal stromal tumors. *Am J Pathol*. (2000) 156:791–5. doi: 10.1016/S0002-9440(10)64946-2
63. Olivieri C, Wang Y, Walker C, Subrahmanian MV, Ha KN, Bernlohr D, et al. The alphaC-beta4 loop controls the allosteric cooperativity between nucleotide and substrate in the catalytic subunit of protein kinase A. *Elife*. (2024) 12. doi: 10.7554/eLife.91506
64. Daley GQ, Baltimore D. Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. *Proc Natl Acad Sci U S A*. (1988) 85:9312–6. doi: 10.1073/pnas.85.23.9312
65. Kindler T, Breitenbuecher F, Kasper S, Estey E, Giles F, Feldman E, et al. Identification of a novel activating mutation (Y842C) within the activation loop of FLT3 in patients with acute myeloid leukemia (AML). *Blood*. (2005) 105:335–40. doi: 10.1182/blood-2004-02-0660
66. de Raedt T, Cools J, Debicq-Rychter M, Brems H, Mentens N, Sciot R, et al. Intestinal neurofibromatosis is a subtype of familial GIST and results from a dominant activating mutation in PDGFRA. *Gastroenterology*. (2006) 131:1907–12. doi: 10.1053/j.gastro.2006.07.002
67. Frohling S, Scholl C, Levine RL, Loriaux M, Boggon TJ, Bernard OA, et al. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell*. (2007) 12:501–13. doi: 10.1016/j.ccr.2007.11.005
68. Moretti S, De Falco V, Tamburrino A, Barbi F, Tavano M, Avenia N, et al. Insights into the molecular function of the inactivating mutations of B-Raf involving the DFG motif. *Biochim Biophys Acta*. (2009) 1793:1634–45. doi: 10.1016/j.bbamcr.2009.09.001
69. Yao Z, Yaeger R, Rodrik-Outmezguine VS, Tao A, Torres NM, Chang MT, et al. Tumours with class 3 BRAF mutants are sensitive to the inhibition of activated RAS. *Nature*. (2017) 548:234–8. doi: 10.1038/nature23291

70. Ma Y, Zeng S, Metcalfe DD, Akin C, Dimitrijevic S, Butterfield JH, et al. The c-KIT mutation causing human mastocytosis is resistant to STI571 and other KIT kinase inhibitors; kinases with enzymatic site mutations show different inhibitor sensitivity profiles than wild-type kinases and those with regulatory-type mutations. *Blood*. (2002) 99:1741–4. doi: 10.1182/blood.V99.5.1741
71. Yamamoto Y, Kiyoi H, Nakano Y, Suzuki R, Kidera Y, Miyawaki S, et al. Activating mutation of D835 within the activation loop of FLT3 in human hematologic Malignancies. *Blood*. (2001) 97:2434–9. doi: 10.1182/blood.V97.8.2434
72. Kaleem B, Shahab S, Zaidi U, Shamsi TS. P-Loop mutations-Negative prognosticators in tyrosine kinase inhibitors resistant chronic myeloid leukemia patients. *Int J Lab Hematol*. (2022) 44:538–46. doi: 10.1111/ijlh.13798
73. Ikenoue T, Hikiba Y, Kanai F, Aragaki J, Tanaka Y, Imamura J, et al. Different effects of point mutations within the B-Raf glycine-rich loop in colorectal tumors on mitogen-activated protein/extracellular signal-regulated kinase kinase/extracellular signal-regulated kinase and nuclear factor kappaB pathway and cellular transformation. *Cancer Res*. (2004) 64:3428–35. doi: 10.1158/0008-5472.CAN-03-3591
74. Brown LM, Ekert PG, Fleuren EDG. Biological and clinical implications of FGFR aberrations in paediatric and young adult cancers. *Oncogene*. (2023) 42:1875–88. doi: 10.1038/s41388-023-02705-7
75. Yeung W, Ruan Z, Kannan N. Emerging roles of the alphaC-beta4 loop in protein kinase structure, function, evolution, and disease. *IUBMB Life*. (2020) 72:1189–202. doi: 10.1002/iub.2253
76. Zhang Q, Jiang T, Xiao M, Zhang J, Ying X. Identification of a novel osimertinib-sensitive mutation, EGFR H773L, in a chinese patient with NSCLC. *J Thorac Oncol*. (2020) 15:e46–e8. doi: 10.1016/j.jtho.2019.11.018