

Pathogenic mutation hotspots in protein kinase domain structure

Kirill E. Medvedev¹  | R. Dustin Schaeffer¹  | Jimin Pei² | Nick V. Grishin^{1,3}

¹Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas, USA

²Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas, USA

³Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas, USA

Correspondence

Kirill E. Medvedev, 5323 Harry Hines Boulevard, Dallas, TX 75390-9050, USA.
Email: kirill.medvedev@utsouthwestern.edu

Funding information

National Institute of General Medical Sciences of the National Institutes of Health, Grant/Award Number: GM147367; National Science Foundation, Grant/Award Number: DBI2224128; Welch Foundation, Grant/Award Number: I-1505; National Institutes of Health, Grant/Award Number: GM127390

Review Editor: Nir Ben-Tal

Abstract

Control of eukaryotic cellular function is heavily reliant on the phosphorylation of proteins at specific amino acid residues, such as serine, threonine, tyrosine, and histidine. Protein kinases that are responsible for this process comprise one of the largest families of evolutionarily related proteins. Dysregulation of protein kinase signaling pathways is a frequent cause of a large variety of human diseases including cancer, autoimmune, neurodegenerative, and cardiovascular disorders. In this study, we mapped all pathogenic mutations in 497 human protein kinase domains from the ClinVar database to the reference structure of Aurora kinase A (AURKA) and grouped them by the relevance to the disease type. Our study revealed that the majority of mutation hotspots associated with cancer are situated within the catalytic and activation loops of the kinase domain, whereas non-cancer-related hotspots tend to be located outside of these regions. Additionally, we identified a hotspot at residue R371 of the AURKA structure that has the highest number of exclusively non-cancer-related pathogenic mutations (21) and has not been previously discussed.

KEY WORDS

domain, kinases, mutation hotspot, protein structure

1 | INTRODUCTION

Post-translational modification of proteins through phosphorylation is a common process in biology. Protein kinases play an essential role in cellular signaling and regulation and are responsible for phosphorylation. There are more than 500 human protein kinases, which are grouped into several families based on their structure and function and which are encoded by 2% of all human genes (Manning et al., 2002). Major kinase groups from the human kinome include AGC (intracellular signaling

kinases), CMGC (responsible for cell cycle control, MAPK signaling, splicing, etc.), CAMK (Calmodulin/Calcium regulated kinases), CK1 (Cell Kinase 1, originally known as Casein Kinase 1), STE (form the MAPK cascade), TK (Tyrosine kinases), TKL (similar to TK, but whose activities are generally on serine/threonine substrates), and RGC (Receptor Guanylate Cyclases) (Manning et al., 2002). Protein kinases are involved in many cellular pathways, and their dysregulation has been implicated in various diseases, including cancer, diabetes, neurodegenerative and cardiovascular disorders, and rheumatoid

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. Protein Science published by Wiley Periodicals LLC on behalf of The Protein Society.

arthritis, among others (Lahiry et al., 2010; Lee & Paull, 2021). Being one of the largest families of evolutionarily related proteins, kinases phosphorylate particular amino acids of 30% of the human proteome (Hubbard & Cohen, 1993) and therefore are one of the most important group of drug targets together with G-protein-coupled receptors (Cohen, 2002; Melnikova & Golden, 2004). Kinase inhibitors, which block the activity of specific kinases, have been developed to treat various types of cancer, autoimmune diseases, and other disorders (Melnikova & Golden, 2004). The role of kinases in carcinogenesis has been intensively studied: the protein kinase domain is the most frequently encoded domain among cancer-related genes (Futreal et al., 2004). In normal cells

kinases function as both tumor suppressors and proto-oncogenes, therefore mutations in these proteins facilitate oncogenic processes such as inhibition of DNA damage response, promotion of angiogenesis, deactivation of apoptotic pathways, and others (Torkamani et al., 2009).

The secondary structure of a kinase domain includes two structurally and functionally distinct subdomains or lobes: N- and C-terminal (Taylor & Kornev, 2011). The smaller N-terminal lobe adopts a five-stranded β -sheet and includes highly conserved sequence motif—Gly-rich loop between β 1 and β 2. This loop interacts with adenine ring of ATP (Madhusudan et al., 2002). The AGC group of typical protein kinases shares the distinctive helix α B in the N-terminal region (Figure 1). This helix is not

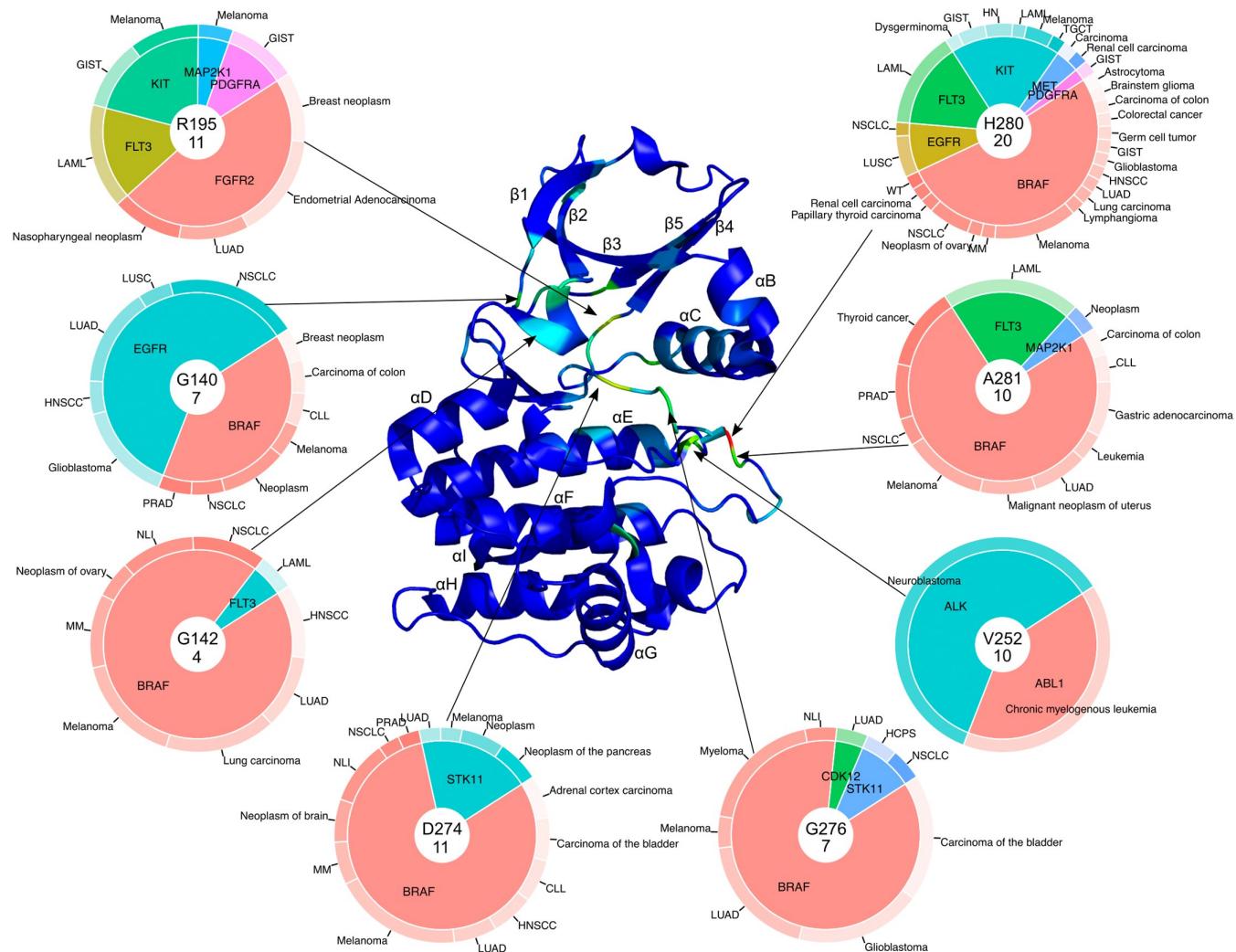


FIGURE 1 Cancer-related mutations mapped to kinase domain of Aurora kinase A (PDB: 3E5A). Pie charts represent genes (inner layer) and medical conditions and diseases (outer layer) that are related to mutations in a particular position. One mutation can be related to several diseases. In the center of each pie chart the top item represents the number of amino acid in 3E5A structure, the bottom item represents the total number of SNVs in this position in MSA. CLL, B-cell chronic lymphocytic leukemia; GIST, gastrointestinal stromal tumor; HCPS, hereditary cancer-predisposing syndrome; HN, hematologic neoplasm; HNSCC, squamous cell carcinoma of the head and neck; LAML, acute myeloid leukemia; LUAD, lung adenocarcinoma; LUSC, squamous cell lung carcinoma; MM, multiple myeloma; NLI, neoplasm of the large intestine; NSCLC, non-small cell lung carcinoma; PRAD, prostate adenocarcinoma; TGCT, germ cell tumor of testis; WT, nephroblastoma (Wilms tumor).

observed in any of the other type of typical kinases (Scheeff & Bourne, 2005). The majority of kinases typically exhibit a noticeable kink near the beginning of the strand $\beta 4$. This kink alters the arrangement and structure of a significant portion of the hydrophobic pocket formed by strand $\beta 4$, helices αC , and αE (Scheeff & Bourne, 2005). The larger C-terminal lobe contains mostly alpha helices. This lobe also includes the catalytic loop that bridges $\beta 6$ and $\beta 7$ and activation loop right after the catalytic region (Taylor & Kornev, 2011). The catalytic region of many kinase families is reinforced by intricate hydrogen-bond networks that stabilize the structure of the active site. Moreover, within this region, a distinctive “crossing loops” structure is formed, where the catalytic loop intersects with the loop between strands $\beta 8$ and $\beta 9$. This motif is uncommon in protein structures and serves as one of the distinguishing features of the kinase superfamily (Grishin, 1999). Helix αE plays a stabilizing role in the ATP binding pocket through its interactions with strands $\beta 7$ and $\beta 8$. In the majority of kinases, αE is oriented at approximately 45° to these strand elements (Scheeff & Bourne, 2005). Helix αF , following the diverse loop structures, represents the final segment of structural similarity shared by all kinases. However, the similarity in this region diminishes rapidly. In certain instances, it could be argued that this helix exhibits poor superposition among superfamily structures and that it should not be regarded as part of the “universal core” (Scheeff & Bourne, 2005). However, αF consistently appears with a similar orientation in most of the kinases' structures, and in many cases, seems to serve a similar role: the stabilization of the backbone of the catalytic loop.

Here we identified pathogenic mutation hotspots using data from the ClinVar database (Landrum et al., 2020) and an alignment of 497 human protein kinase domains. We defined hotspot mutations based on the frequency of mutations among kinase domains. Our analysis showed that most of the cancer-related mutation hotspots are located within catalytic and activation loops of kinase domain, whereas most of non-cancer-related hotspots are mainly located outside of these loops. We also identified the hotspot with the highest number of exclusively non-cancer-related pathogenic mutations (21) that corresponds to R371 of the AURKA structure and has not been discussed previously.

2 | RESULTS AND DISCUSSION

We mapped all pathogenic mutations in 497 human protein kinase domains (Modi & Dunbrack Jr., 2019) from ClinVar database (Landrum et al., 2020) to the reference structure of Aurora kinase A (AURKA) and grouped them by the relevance to the disease type (cancer-related and others).

Cancer-related mutations mapped to the Aurora kinase A 3D structure revealed eight major hotspots (Figure 1). Most of the hotspots are located within catalytic and activation loops of kinase domain (Dixit et al., 2009). The hotspot with the highest number of mutations (20) corresponds to H280 of AURKA, located in the activation loop. This hotspot includes mutations from six kinase genes with the BRAF gene being responsible for the largest number of various cancer types (Figure 1). BRAF V600 is a known driver mutation that is related to poor prognosis in colorectal cancer (Chen et al., 2014), lung cancer (Cardarella et al., 2013), melanoma (Flaherty et al., 2012), and glioblastoma (Chang et al., 2016). Driver mutations from other kinases localized in this hotspot include EGFR L861 (lung cancer [Mitsudomi & Yatabe, 2010]), FLT3 D835 (leukemia [Martelli et al., 2013]), KIT D816 (leukemia [Smith & Shah, 2013]), melanoma [Willmore-Payne et al., 2006]), MET D1228 (carcinoma [Lorenzato et al., 2002]), PDGFRA D842 (gastrointestinal tumors [Heinrich et al., 2012]). Two cancer-related hotspots with 11 total mutations correspond to R195 and D274 of the AURKA structure (Figure 1). R195 includes mutations in five genes: FGFR2 (endometrial carcinoma [Dutt et al., 2008], breast neoplasm [Chang et al., 2016]), FLT3 (leukemia [Zhang et al., 2014]), KIT (gastrointestinal stromal tumors [Fornasarig et al., 2020]), MAP2K1 (melanoma [Emery et al., 2009]), and PDGFRA (gastrointestinal stromal tumors [Corless et al., 2005]). R195 also includes 13 non-cancer-related mutations in five genes (Figure 2): FGFR3 (hypochondroplasia [Xue et al., 2014]), PDGFRB (myofibromatosis [Arts et al., 2016]), MAP2K1 and MAP2K2 (cardio-facio-cutaneous syndrome [Schulz et al., 2008]), and ACVR1 (telangiectasia [Koenighofer et al., 2019]). D274 locates in activation loop of kinase domain and includes mutations from two genes: BRAF (melanoma [Maldonado et al., 2003]) and STK11 (lung cancer [Ji et al., 2007]). Two cancer-related hotspots with 10 mutations correspond to V252 (catalytic loop) and A281 (activation loop) (Figure 1). V252 includes mutations in two genes ALK (neuroblastoma [Schonherr et al., 2011]) and ABL1 (chronic myelogenous leukemia [Cho et al., 2013]). A281 includes mutations in three genes: BRAF (melanoma [Chang et al., 2016]), FLT3 (leukemia [Mills et al., 2006]), and MAP2K1 (neoplasms [Morris et al., 2013]). The final cancer-related hotspot located within the activation loop is G276 which has 7 mutations in three genes: BRAF (glioblastoma [Chang et al., 2016]), STK11 (non-small cell lung carcinoma [MacConaill et al., 2014]), and CDK12 (lung adenocarcinoma [Biswas et al., 2016]).

In the contrast to the cancer-related mutations, non-cancer-related hotspots are mostly located outside of catalytic and activation loops of kinase domain and include larger number of involved genes (Figure 2). The hotspot with the highest number of exclusively non-

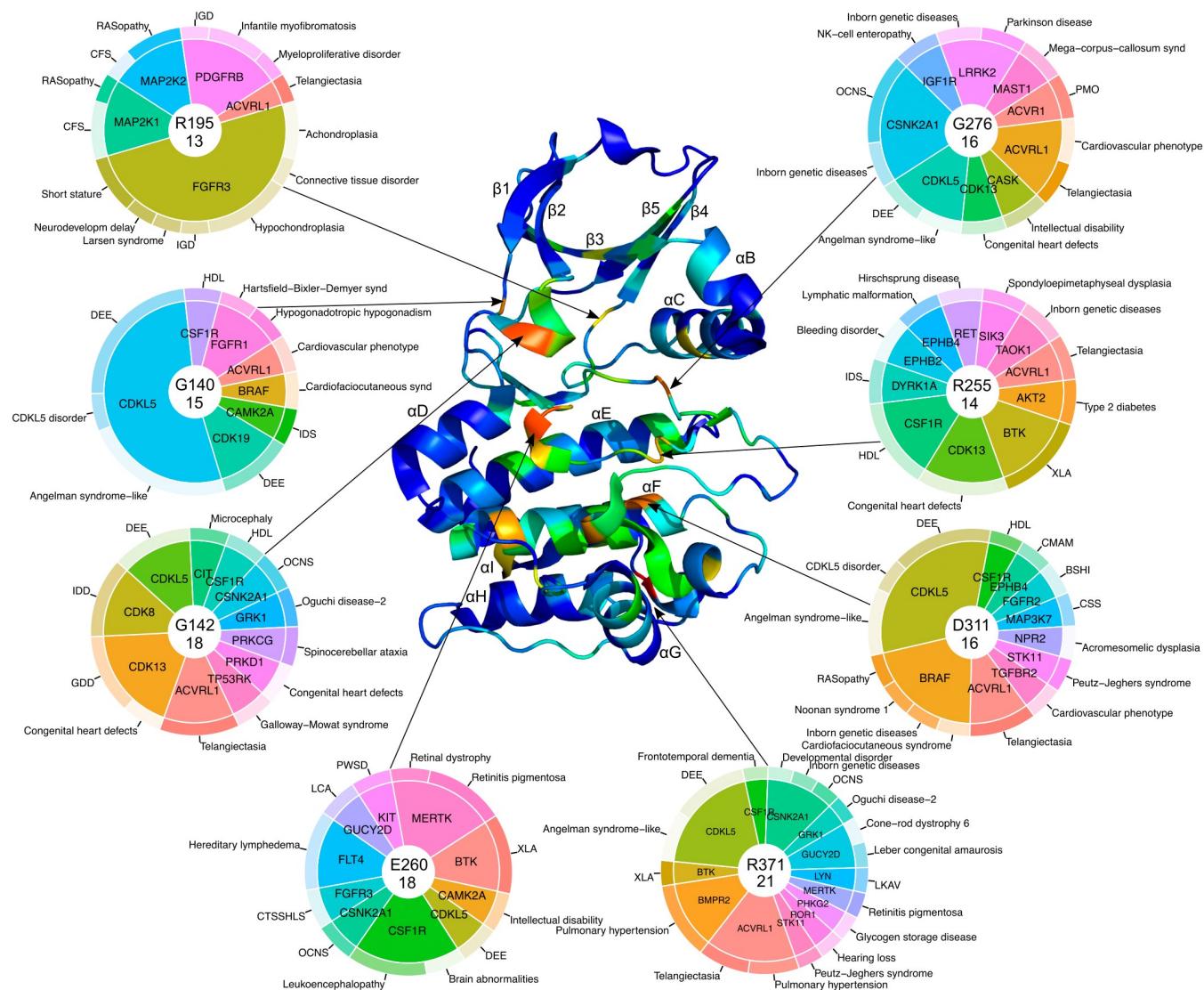


FIGURE 2 Non-cancer-related mutations mapped to kinase domain of Aurora kinase A (PDB: 3E5A). Pie charts represent genes (inner layer) and medical conditions and diseases (outer layer) that are related to mutations in a particular position. One mutation can be related to several diseases. In the center of each pie chart the top item represents the number of amino acid in 3E5A structure, the bottom item represents the total number of SNVs in this position in MSA. BSHI, bilateral sensorineuronal hearing impairment; CFS, cardiofaciocutaneous syndrome; CMAM, capillary malformation-arteriovenous malformation; CSS, cardiospondylocarpofacial syndrome; CTSSHLs, camptodactyly-tall stature-scoliosis-hearing loss syndrome; DEE, developmental and epileptic encephalopathy; GDD, global developmental delay; HDL, hereditary diffuse leukoencephalopathy; IDD, intellectual developmental disorder; IDS, intellectual disability syndrome; IGD, inborn genetic diseases; LCA, Leber congenital amaurosis; LKAV, LYN kinase-associated vasculopathy; OCNS, Okur-Chung neurodevelopmental syndrome; PMO, progressive myositis ossificans; PWSD, piebaldism with sensorineuronal deafness; XLA, X-linked agammaglobulinemia.

cancer-related pathogenic mutations (21) corresponds to R371 of the AURKA structure and is located at the C-terminal part of α H of the kinase domain. To the best of our knowledge, this hotspot has not been described previously. R371 includes mutations in 15 genes that correspond to seven kinase groups (Table 1). The majority of these kinase groups (five out of seven) are Ser/Thr protein kinases (AGC, CAMK, CMGC, STE, TKL). The rest are Receptor Guanylate Cyclase (RGC) and Tyr protein

kinases (TYR). Tyrosine kinases are the prevalent group among proteins with mutations in the R371 hotspot (6 out of 15). Based on their Medical Subject Heading (MeSH) disease classification (Rogers, 1963) the majority of diseases caused by mutations from these hotspots belong to classes of nervous system diseases, such as encephalopathy (Hector et al., 2017), Okur-Chung syndrome (Owen et al., 2018), and X-linked agammaglobulinemia (Lopez-Herrera et al., 2008); as well as eye

TABLE 1 Disease classes that are caused by non-cancerous mutations corresponded to R371 hotspot of Aurora kinase A.

Gene name	UniProt ID	Kinase group	Disease	MeSH class
GRK1	Q15835	AGC	Oguchi disease	Eye diseases
PHKG2	P15735	CAMK	Glycogen storage disease IXC	Nutritional and metabolic diseases
STK11	Q15831	CAMK	Peutz-Jeghers syndrome	Digestive system diseases; skin and connective tissue diseases
CDKL5	O76039	CMGC	Developmental and epileptic encephalopathy	Nervous system diseases
			Angelman syndrome-like	Nervous system diseases
CSNK2A1	P68400	CMGC	Okur-Chung neurodevelopmental syndrome	Nervous system diseases
GUCY2D	Q02846	RGC	Cone-rod dystrophy 6	Eye diseases
			Leber congenital amaurosis 1	Eye diseases
TAOK1	Q7L7X3	STE	n/a	n/a
ACVRL1	P37023	TKL	Pulmonary hypertension	Respiratory tract diseases
			Telangiectasia, hereditary hemorrhagic	Hemic and lymphatic diseases; cardiovascular diseases
BMPR2	Q13873	TKL	Pulmonary hypertension	Respiratory tract diseases
BTK	Q06187	TYR	X-linked agammaglobulinemia with growth hormone deficiency	Immune system diseases; nervous system diseases; endocrine system diseases
CSF1R	P07333	TYR	Frontotemporal dementia	Nervous system diseases; mental disorders
LYN	P07948	TYR	LYN kinase associated vasculopathy	Diseases of the musculoskeletal system and connective tissue
MERTK	Q12866	TYR	Retinitis pigmentosa 38	Eye diseases
ROR1	Q01973	TYR	Hearing loss, autosomal recessive 108	Nervous system diseases; otorhinolaryngologic diseases
ZAP70	P43403	TYR	n/a	n/a

Abbreviation: MeSH, Medical Subject Headings.

diseases such as Leber congenital amaurosis (Thompson et al., 2017). Enrichment analysis of 15 kinases related to the R371 hotspot using ShinyGO (Ge et al., 2020) revealed enrichment of these proteins in the NF- κ B signaling pathway, which is distinctive when using the whole human proteome as a background (Enrichment FDR = 2.3E-05, Fold Enrichment = 62.6) and our dataset of kinases (Enrichment FDR = 4.6E-02, Fold Enrichment = 8.2). The NF- κ B signaling pathway is known for its key role in inflammatory processes (Liu et al., 2017); however, it also plays an important role in the function of the nervous system (Dresselhaus & Meffert, 2019; Kaltschmidt & Kaltschmidt, 2009) and the regulation of ocular surface inflammation (Lan et al., 2012).

Two non-cancer-related hotspots with 18 mutations correspond to G142 and E260. G142 is located within the Gly-rich loop of the N-terminal lobe of the kinase domain and includes non-cancer-related mutations in 11 genes such as: CIT (microcephaly [Li et al., 2016]), CDK8 (intellectual developmental disorder [Calpena et al., 2019]),

TP53RK (Galloway-Mowat syndrome [Braun et al., 2017]). This hotspot also includes four cancer-related mutations in two genes (Figure 1): BRAF (non-small cell lung carcinoma [Houben et al., 2004]) and FLT3 (acute myeloid leukemia [Zhang et al., 2014]). E260 locates within catalytic loop and includes non-cancer-related mutations in 10 genes, for example: BTK (agammaglobulinemia [Sigmon et al., 2008]), CSF1R (leukoencephalopathy [Kinoshita et al., 2012]), FGFR3 (hearing loss syndrome [Toydemir et al., 2006]). The hotspot D311 locates in the N-terminal part of α F and includes 16 mutations in 10 genes such as: CDKL5 (encephalopathy [Melikishvili et al., 2019]), ACVRL1 (telangiectasia [Letteboer et al., 2005]), MAP3K7 (cardiospondylocarpofacial syndrome [Le Goff et al., 2016]).

Another hotspot with 16 mutations corresponds to G276 of the AURKA structure. It is the only non-cancer-related hotspot located within the activation loop and includes mutations in 9 genes such as MAST1 (mega-corpus-callosum syndrome [Tripathy et al., 2018]), ACVRL1 (telangiectasia [Schulte et al., 2005]). This

hotspot also includes the seven cancer-related mutations discussed above (Figure 1). Hotspot G140 is located in the Gly-rich loop region of kinase domains and includes 15 non-cancer-related pathogenic mutations in seven genes, for example: CDK19 (encephalopathy [Zarate et al., 2021]), CSF1R (leukoencephalopathy [Rademakers et al., 2011]). This hotspot also includes seven cancer-related mutations in two genes (Figure 1): BRAF (melanoma [Carlino et al., 2014]) and EGFR (non-small cell

TABLE 2 Top kinase genes with largest number of pathogenic cancer-related mutations.

Gene name	UniProt ID	Number of mutations
BRAF	P15056	53
ALK	Q9UM73	27
ABL1	P00519	26
FLT3	P36888	26
STK11	Q15831	21
EGFR	P00533	19
MAP2K1	Q02750	18
KIT	P10721	18
FGFR2	P21802	17
MET	P08581	17
ERBB2	P04626	14

lung carcinoma [Berge et al., 2013]). Finally, R255 is located within catalytic loop and includes non-cancer-related mutations in 11 genes such as AKT2 (type 2 diabetes [George et al., 2004]), DYRK1A (intellectual disability [Ji et al., 2015]), and CSF1R (leukoencephalopathy [Guerreiro et al., 2013]).

A previous pan-cancer study revealed that among protein domains significantly enriched in mutations, the protein kinase domain (Pfam ID: PF07714) is the second most mutated, considering all detected mutations (Miller et al., 2015). The top three kinase genes with the largest number of mutations in pan-cancer dataset included BRAF, EGFR and ERBB2 (Miller et al., 2015). Our analysis of cancer-related pathogenic mutations from the ClinVar database showed that the top three genes (containing the most pathogenic cancer-related mutations) are BRAF, ALK, ABL1, and EGFR and ERBB2 are in the top 11 (Table 2).

Furthermore, we compared the statistics of cancerous and non-cancerous mutations by calculating the over and underrepresentation of these mutations at each position (see Materials and Methods). Our analysis revealed 19 positions with significant (p -value <0.05) overrepresentation of cancerous mutations and seven positions with significant overrepresentation of non-cancerous mutations (Figure 3, Table S1). Mapping these positions to the Aurora kinase A structure revealed that

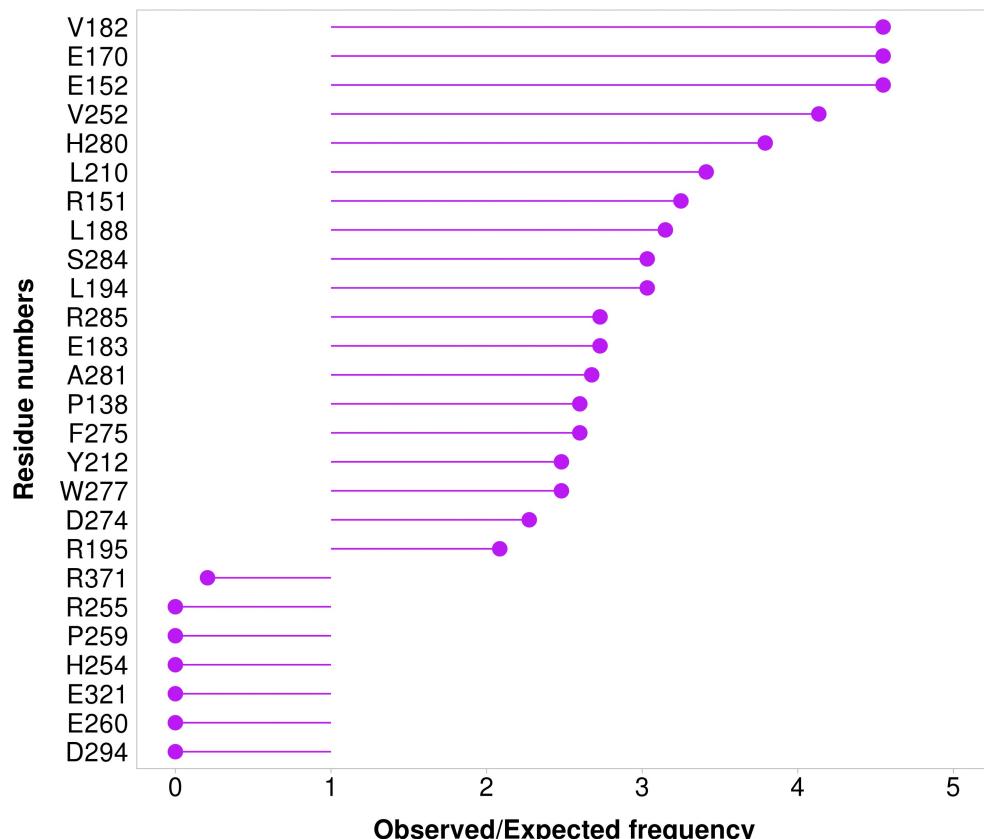


FIGURE 3 The ratio of observed and expected frequencies of cancer-related mutations in specific positions of Aurora kinase A (p -value < 0.05). Ratio > 1 defines overrepresentation of cancer-related and underrepresentation of non-cancer-related mutations in a position. Ratio < 1 defines underrepresentation of cancer-related and overrepresentation of non-cancer-related mutations in a position.

positions with an overrepresentation of cancerous mutations are predominantly located in the N-terminal region of the kinase domain and within the catalytic loop (Figure 4). Only one position with overrepresented cancer-related mutations is located within the activation loop. On the other hand, positions with overrepresentation of non-cancerous mutations are mostly located in C-terminal region of the domain and within the activation loop (Figure 4).

We investigated the intrinsic tendency of each residue to mutate by collecting all missense mutations with benign clinical relevance and a population frequency greater than or equal to 1% from the gnomAD database (Gudmundsson et al., 2022) for all kinases in our dataset. Overall, we identified 26 such mutations with maximum four mutations per position of Aurora kinase A structure (Table S1). Benign mutations are predominantly located on the surface of the protein structure and outside of the catalytic and activation loops (Figure S1). Moreover, the majority of cancerous and non-cancerous mutation hotspots do not share positions in common with benign missense mutations (Table S1).

Thus, our analysis revealed major cancer-related pathogenic mutation hotspots in the structure of the

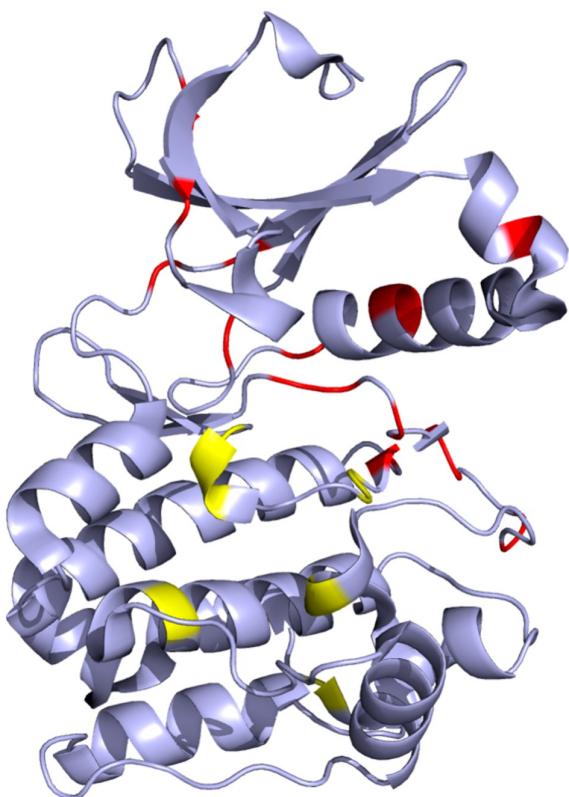


FIGURE 4 Positions of Aurora kinase A with an overrepresentation of cancer-related (red) and non-cancer-related (yellow) mutations.

kinase domain located mostly within the catalytic and activation loops. Non-cancer-related hotspots are mostly located outside of catalytic and activation loops of kinase domain and include a larger number of involved genes. Protein kinases play dual roles as tumor suppressors and proto-oncogenes in normal, healthy cells. As a consequence, mutations in protein kinases often promote a diverse array of tumorigenic activities. Due to their crucial involvement in DNA damage response and cell cycle checkpoints, protein kinases are associated with numerous loss-of-function mutations that can lead to the inhibition of apoptosis, the acquisition of a “mutator” phenotype, and eventual tumorigenesis (Torkamani et al., 2009). Structures of protein kinases have unveiled significant structural distinctions between the closely related active and highly specific inactive forms of kinases (Nolen et al., 2004). For instance, the conformational landscape of ABL kinase, which encompasses active, inactive, intermediate, and inactive-like conformations, has provided evidence that the various structures of the kinase activation loop may indeed reflect natural kinase conformations (Cowan-Jacob et al., 2005). Hence, it is likely that activating mutations disturb this equilibrium, favoring the active conformation by either destabilizing the inactive state or stabilizing the active state of the protein kinase. This mechanism could potentially contribute to certain mutations' ability to promote tumorigenesis and could be the reason for the locations of cancerous mutations within the catalytic and activation loops.

3 | MATERIALS AND METHODS

3.1 | Data collection

We used a structure-based multiple sequence alignment (MSA) of 497 human protein kinase domains from Modi and Dunbrack Jr. (2019). This MSA does not consider atypical kinases. For each protein in the MSA “Pathogenic” and “Likely Pathogenic” single nucleotide variants (SNVs) that result in single amino acid substitutions were retrieved from ClinVar database (Landrum et al., 2020). Retrieved SNVs were considered as pathogenic and were mapped to MSA and to the reference structure of Aurora kinase A (Gene name: AURKA, UniProt: O14965, PDB: 3E5A [Zhao et al., 2008]). The reference structure was selected based on previous study (Modi & Dunbrack Jr., 2019). Total numbers of SNVs were summarized for each amino acid position in AURKA and were treated as B-factors in the PDB file for corresponding C-alpha atoms. The resulting PDB file was visualized using PyMOL. SNVs located in the gap regions of AURKA in

MSA were excluded from consideration. In this study, we determined hotspot mutations by counting the frequency of their occurrence in kinase domains. The top 5% positions with the highest frequency of mutation for cancer-related and non-cancer-related cases were considered as hotspots. Medical conditions and diseases that are related to SNVs in human protein kinase domains were retrieved from ClinVar database. Cancer-related medical conditions were sorted out manually from the total list. Pie charts were created using R package, v4.2.1, library `webr` and `PieDonut` function with following settings for `PieDonut` arguments: `showRatioPie = FALSE`, `showRatioDonut = FALSE`, `labelpositionThreshold = 0.5`, `donutLabelSize = 3.0`, `pieLabelSize = 3.5`, `r0 = 0.2`, `r1 = 0.7`, `r2 = 0.8`, `maxx = 1.7`, `start = 1`.

3.2 | Statistical analysis

Comparison of the number of cancerous and non-cancerous mutations for each position was conducted using over and underrepresentation, which was calculated as ratio of observed and expected frequencies. The observed frequency for cancer mutations in each Aurora kinase A position was calculated as a ratio of the number of the cancerous mutations in this position over the sum of all cancerous mutations mapped to Aurora kinase A structure. The observed frequency for non-cancer mutations in each Aurora kinase A position was calculated as a ratio of the number of the non-cancerous mutations in this position over the sum of all non-cancerous mutations mapped to Aurora kinase A structure. The expected frequency for each position was calculated as ratio of total number of mutation (cancerous and non-cancerous) in this position of Aurora kinase A over the total number of mutations. Significance of over and underrepresentation was checked using chi-square test (p -value <0.05 is considered significant). Statistical analysis was conducted using the R package, v4.2.1.

3.3 | Analysis of benign mutations

To examine the inherent propensity of each residue to mutate, we obtained all missense mutations with benign clinical relevance and frequency in population greater than or equal to 1% from gnomAD v2.1.1 database (Gudmundsson et al., 2022) for all human protein kinases in our dataset. Obtained benign mutations were mapped to human kinases MSA and to the reference structure of Aurora kinase A (Figure S1).

AUTHOR CONTRIBUTIONS

Kirill E. Medvedev: Conceptualization; methodology; software; validation; formal analysis; investigation; data curation; visualization; writing—original draft; project administration. R. Dustin Schaeffer: Writing—review and editing; funding acquisition. Jimin Pei: Writing—review and editing. Nick V. Grishin: Conceptualization; resources; funding acquisition; writing—review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no competing interests associated with the manuscript.

ORCID

Kirill E. Medvedev  <https://orcid.org/0000-0002-7982-4242>

R. Dustin Schaeffer  <https://orcid.org/0000-0001-6502-1425>

REFERENCES

- Arts FA, Chand D, Pecquet C, Velghe AI, Constantinescu S, Hallberg B, et al. Pdgfrb mutants found in patients with familial infantile myofibromatosis or overgrowth syndrome are oncogenic and sensitive to imatinib. *Oncogene*. 2016;35(25):3239–48.
- Berge EM, Aisner DL, Doebele RC. Erlotinib response in an nsclc patient with a novel compound g719d+l861r mutation in egfr. *J Thorac Oncol*. 2013;8(9):e83–4.
- Biswas R, Gao S, Cultraro CM, Maity TK, Venugopalan A, Abdullaev Z, et al. Genomic profiling of multiple sequentially acquired tumor metastatic sites from an "exceptional responder" lung adenocarcinoma patient reveals extensive genomic heterogeneity and novel somatic variants driving treatment response. *Cold Spring Harb Mol Case Stud*. 2016; 2(6):a001263.
- Braun DA, Rao J, Mollet G, Schapiro D, Daugeron MC, Tan W, et al. Mutations in keops-complex genes cause nephrotic syndrome with primary microcephaly. *Nat Genet*. 2017;49(10):1529–38.
- Calpena E, Hervieu A, Kaserer T, Swagemakers SMA, Goos JAC, Popoola O, et al. De novo missense substitutions in the gene encoding cdk8, a regulator of the mediator complex, cause a syndromic developmental disorder. *Am J Hum Genet*. 2019; 104(4):709–20.
- Cardarella S, Ogino A, Nishino M, Butaney M, Shen J, Lydon C, et al. Clinical, pathologic, and biologic features associated with braf mutations in non-small cell lung cancer. *Clin Cancer Res*. 2013;19(16):4532–40.
- Carlino MS, Haydu LE, Kakavand H, Menzies AM, Hamilton AL, Yu B, et al. Correlation of braf and nras mutation status with outcome, site of distant metastasis and response to chemotherapy in metastatic melanoma. *Br J Cancer*. 2014;111(2):292–9.
- Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol*. 2016;34(2):155–63.

Chen D, Huang JF, Liu K, Zhang LQ, Yang Z, Chuai ZR, et al. Brafv600e mutation and its association with clinicopathological features of colorectal cancer: a systematic review and meta-analysis. *PloS One*. 2014;9(3):e90607.

Cho YU, Kim SO, Chi HS, Park SJ, Jang S, Park CJ, et al. Detection of abl1 kinase mutations in Philadelphia-positive patients exhibiting an inadequate molecular response using restriction fragment mass polymorphism and its clinical significance: a single-center experience in Korea. *Int J Lab Hematol*. 2013; 35(6):589–600.

Cohen P. Protein kinases—the major drug targets of the twenty-first century? *Nat Rev Drug Discov*. 2002;1(4):309–15.

Corless CL, Schroeder A, Griffith D, Town A, McGreevey L, Harrell P, et al. Pdgfra mutations in gastrointestinal stromal tumors: frequency, spectrum and in vitro sensitivity to imatinib. *J Clin Oncol*. 2005;23(23):5357–64.

Cowan-Jacob SW, Fendrich G, Manley PW, Jahnke W, Fabbro D, Liebetanz J, et al. The crystal structure of a c-src complex in an active conformation suggests possible steps in c-src activation. *Structure*. 2005;13(6):861–71.

Dixit A, Yi L, Gowthaman R, Torkamani A, Schork NJ, Verkhivker GM. Sequence and structure signatures of cancer mutation hotspots in protein kinases. *PloS One*. 2009;4(10): e7485.

Dresselhaus EC, Meffert MK. Cellular specificity of nf-kappab function in the nervous system. *Front Immunol*. 2019;10:1043.

Dutt A, Salvesen HB, Chen TH, Ramos AH, Onofrio RC, Hatton C, et al. Drug-sensitive fgfr2 mutations in endometrial carcinoma. *Proc Natl Acad Sci U S A*. 2008;105(25):8713–7.

Emery CM, Vijayendran KG, Zipser MC, Sawyer AM, Niu L, Kim JJ, et al. Mek1 mutations confer resistance to mek and b-raf inhibition. *Proc Natl Acad Sci U S A*. 2009;106(48): 20411–6.

Flaherty KT, Robert C, Hersey P, Nathan P, Garbe C, Milhem M, et al. Improved survival with mek inhibition in braf-mutated melanoma. *N Engl J Med*. 2012;367(2):107–14.

Fornasarig M, Gasparotto D, Foltran L, Campigotto M, Lombardi S, del Savio E, et al. A novel kindred with familial gastrointestinal stromal tumors caused by a rare kit germline mutation (n655k): Clinico-pathological presentation and tki sensitivity. *J Pers Med*. 2020;10(4):234.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. *Nat Rev Cancer*. 2004; 4(3):177–83.

Ge SX, Jung D, Yao R. Shinygo: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. 2020;36(8):2628–9.

George S, Rochford JJ, Wolfrum C, Gray SL, Schinner S, Wilson JC, et al. A family with severe insulin resistance and diabetes due to a mutation in akt2. *Science*. 2004;304(5675):1325–8.

Grishin NV. Phosphatidylinositol phosphate kinase: a link between protein kinase and glutathione synthase folds. *J Mol Biol*. 1999; 291(2):239–47.

Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant interpretation using population databases: lessons from gnomad. *Hum Mutat*. 2022;43(8): 1012–30.

Guerreiro R, Kara E, le Ber I, Bras J, Rohrer JD, Taipa R, et al. Genetic analysis of inherited leukodystrophies: genotype-phenotype correlations in the csflr gene. *JAMA Neurol*. 2013; 70(7):875–82.

Hector RD, Kalscheuer VM, Hennig F, Leonard H, Downs J, Clarke A, et al. Cdkl5 variants: improving our understanding of a rare neurologic disorder. *Neurol Genet*. 2017;3(6):e200.

Heinrich MC, Griffith D, McKinley A, Patterson J, Presnell A, Ramachandran A, et al. Crenolanib inhibits the drug-resistant pdgfra d842v mutation associated with imatinib-resistant gastrointestinal stromal tumors. *Clin Cancer Res*. 2012;18(16): 4375–84.

Houben R, Becker JC, Kappel A, Terheyden P, Brocker EB, Goetz R, et al. Constitutive activation of the ras-raf signaling pathway in metastatic melanoma is associated with poor prognosis. *J Carcinog*. 2004;3:6.

Hubbard MJ, Cohen P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem Sci*. 1993;18(5):172–7.

Ji H, Ramsey MR, Hayes DN, Fan C, McNamara K, Kozlowski P, et al. Lkb1 modulates lung cancer differentiation and metastasis. *Nature*. 2007;448(7155):807–10.

Ji J, Lee H, Argiroopoulos B, Dorrani N, Mann J, Martinez-Agosto JA, et al. Dykrla haploinsufficiency causes a new recognizable syndrome with microcephaly, intellectual disability, speech impairment, and distinct facies. *Eur J Hum Genet*. 2015;23(11):1473–81.

Kaltschmidt B, Kaltschmidt C. Nf-kappab in the nervous system. *Cold Spring Harb Perspect Biol*. 2009;1(3):a001271.

Kinoshita M, Yoshida K, Oyanagi K, Hashimoto T, Ikeda S. Hereditary diffuse leukoencephalopathy with axonal spheroids caused by r782h mutation in csflr: case report. *J Neurol Sci*. 2012; 318(1–2):115–8.

Koenighofer M, Parzefall T, Frohne A, Allen M, Unterberger U, Laccone F, et al. Spectrum of novel hereditary hemorrhagic telangiectasia variants in an austrian patient cohort. *Clin Exp Otorhinolaryngol*. 2019;12(4):405–11.

Lahiry P, Torkamani A, Schork NJ, Hegele RA. Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat Rev Genet*. 2010;11(1):60–74.

Lan W, Petznick A, Heryati S, Rifada M, Tong L. Nuclear factor-kappab: central regulator in ocular surface inflammation and diseases. *Ocul Surf*. 2012;10(3):137–48.

Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. Clinvar: improvements to accessing data. *Nucleic Acids Res*. 2020;48(D1):D835–44.

Le Goff C, Rogers C, Le Goff W, Pinto G, Bonnet D, Chrabieh M, et al. Heterozygous mutations in map3k7, encoding tgf-beta-activated kinase 1, cause cardiospondylocarpofacial syndrome. *Am J Hum Genet*. 2016;99(2):407–13.

Lee JH, Paull TT. Cellular functions of the protein kinase atm and their relevance to human disease. *Nat Rev Mol Cell Biol*. 2021; 22(12):796–814.

Letteboer TG, Zewald RA, Kamping EJ, de Haas G, Mager JJ, Snijder RJ, et al. Hereditary hemorrhagic telangiectasia: Eng and alk-1 mutations in dutch patients. *Hum Genet*. 2005; 116(1–2):8–16.

Li H, Bielas SL, Zaki MS, Ismail S, Farfara D, Um K, et al. Biallelic mutations in citron kinase link mitotic cytokinesis to human primary microcephaly. *Am J Hum Genet*. 2016;99(2):501–10.

Liu T, Zhang L, Joo D, Sun SC. Nf-kappab signaling in inflammation. *Signal Transduct Target Ther.* 2017;2:17023.

Lopez-Herrera G, Berron-Ruiz L, Mogica-Martinez D, Espinosa-Rosales F, Santos-Argumedo L. Characterization of bruton's tyrosine kinase mutations in mexican patients with x-linked agammaglobulinemia. *Mol Immunol.* 2008;45(4):1094–8.

Lorenzato A, Olivero M, Patane S, Rosso E, Oliaro A, Comoglio PM, et al. Novel somatic mutations of the met oncogene in human carcinoma metastases activating cell motility and invasion. *Cancer Res.* 2002;62(23):7025–30.

MacConaill LE, Garcia E, Shivedasani P, Ducar M, Adusumilli R, Breneiser M, et al. Prospective enterprise-level molecular genotyping of a cohort of cancer patients. *J Mol Diagn.* 2014;16(6):660–72.

Madhusudan AP, Xuong NH, Taylor SS. Crystal structure of a transition state mimic of the catalytic subunit of camp-dependent protein kinase. *Nat Struct Biol.* 2002;9(4):273–7.

Maldonado JL, Fridlyand J, Patel H, Jain AN, Busam K, Kageshita T, et al. Determinants of braf mutations in primary melanomas. *J Natl Cancer Inst.* 2003;95(24):1878–90.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science.* 2002;298(5600):1912–34.

Martelli MP, Sportoletti P, Tiacci E, Martelli MF, Falini B. Mutational landscape of aml with normal cytogenetics: biological and clinical implications. *Blood Rev.* 2013;27(1):13–22.

Melikishvili G, Epitashvili N, Tabatadze N, Chikvinidze G, Dulac O, Bienvenu T, et al. New insights in phenomenology and treatment of epilepsy in cdkl5 encephalopathy. *Epilepsy Behav.* 2019;94:308–11.

Melnikova I, Golden J. Targeting protein kinases. *Nat Rev Drug Discov.* 2004;3(12):993–4.

Miller ML, Reznik E, Gauthier NP, Aksoy BA, Korkut A, Gao J, et al. Pan-cancer analysis of mutation hotspots in protein domains. *Cell Syst.* 2015;1(3):197–209.

Mills KI, Walsh V, Gilkes AF, Agrawal SG, Knapper S. Novel observation of three flt3 codons mutated in tandem in an elderly acute myeloid leukaemia patient. *Br J Haematol.* 2006;132(1):116–7.

Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: Egfr gene and cancer. *FEBS J.* 2010;277(2):301–8.

Modi V, Dunbrack RL Jr. A structurally-validated multiple sequence alignment of 497 human protein kinase domains. *Sci Rep.* 2019;9(1):19790.

Morris EJ, Jha S, Restaino CR, Dayananth P, Zhu H, Cooper A, et al. Discovery of a novel erk inhibitor with activity in models of acquired resistance to braf and mek inhibitors. *Cancer Discov.* 2013;3(7):742–50.

Nolen B, Taylor S, Ghosh G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell.* 2004;15(5):661–75.

Owen CI, Bowden R, Parker MJ, Patterson J, Patterson J, Price S, et al. Extending the phenotype associated with the csnk2a1-related okur-chung syndrome-a clinical study of 11 individuals. *Am J Med Genet A.* 2018;176(5):1108–14.

Rademakers R, Baker M, Nicholson AM, Rutherford NJ, Finch N, Soto-Ortolaza A, et al. Mutations in the colony stimulating factor 1 receptor (csf1r) gene cause hereditary diffuse leukoencephalopathy with spheroids. *Nat Genet.* 2011;44(2):200–5.

Rogers FB. Medical subject headings. *Bull Med Libr Assoc.* 1963; 51(1):114–6.

Scheeff ED, Bourne PE. Structural evolution of the protein kinase-like superfamily. *PLoS Comput Biol.* 2005;1(5):e49.

Schonherr C, Ruuth K, Yamazaki Y, Eriksson T, Christensen J, Palmer RH, et al. Activating alk mutations found in neuroblastoma are inhibited by crizotinib and nvp-tae684. *Biochem J.* 2011;440(3):405–13.

Schulte C, Geisthoff U, Lux A, Kupka S, Zenner HP, Blin N, et al. High frequency of eng and alk1/acvr1l mutations in german hht patients. *Hum Mutat.* 2005;25(6):595.

Schulz AL, Albrecht B, Arici C, van der Burgt I, Buske A, Gillessen-Kaesbach G, et al. Mutation and phenotypic spectrum in patients with cardio-facio-cutaneous and costello syndrome. *Clin Genet.* 2008;73(1):62–70.

Sigmon JR, Kasasbeh E, Krishnaswamy G. X-linked agammaglobulinemia diagnosed late in life: case report and review of the literature. *Clin Mol Allergy.* 2008;6:5.

Smith CC, Shah NP. The role of kinase inhibitors in the treatment of patients with acute myeloid leukemia. *Am Soc Clin Oncol Educ Book.* 2013;33(1):313–8.

Taylor SS, Kornev AP. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem Sci.* 2011;36(2):65–77.

Thompson JA, De Roach JN, McLaren TL, Montgomery HE, Hoffmann LH, Campbell IR, et al. The genetic profile of leber congenital amaurosis in an australian cohort. *Mol Genet Genomic Med.* 2017;5(6):652–67.

Torkamani A, Verkhivker G, Schork NJ. Cancer driver mutations in protein kinase genes. *Cancer Lett.* 2009;281(2):117–27.

Toydemir RM, Brassington AE, Bayrak-Toydemir P, Krakowiak PA, Jorde LB, Whitby FG, et al. A novel mutation in fgfr3 causes camptodactyly, tall stature, and hearing loss (catshl) syndrome. *Am J Hum Genet.* 2006;79(5):935–41.

Tripathy R, Leca I, van Dijk T, Weiss J, van Bon BW, Sergaki MC, et al. Mutations in mast1 cause mega-corpus-callosum syndrome with cerebellar hypoplasia and cortical malformations. *Neuron.* 2018;100(6):1354–1368.e5.

Willmore-Payne C, Holden JA, Chadwick BE, Layfield LJ. Detection of c-kit exons 11- and 17-activating mutations in testicular seminomas by high-resolution melting amplicon analysis. *Mod Pathol.* 2006;19(9):1164–9.

Xue Y, Sun A, Mekikian PB, Martin J, Rimoin DL, Lachman RS, et al. Fgfr3 mutation frequency in 324 cases from the international skeletal dysplasia registry. *Mol Genet Genomic Med.* 2014;2(6):497–503.

Zarate YA, Uehara T, Abe K, Oginuma M, Harako S, Ishitani S, et al. Cdk19-related disorder results from both loss-of-function and gain-of-function de novo missense variants. *Genet Med.* 2021;23(6):1050–7.

Zhang W, Gao C, Konopleva M, Chen Y, Jacamo RO, Borthakur G, et al. Reversal of acquired drug resistance in flt3-mutated acute myeloid leukemia cells via distinct drug combination strategies. *Clin Cancer Res.* 2014;20(9):2363–74.

Zhao B, Smallwood A, Yang J, Koretke K, Nurse K, Calamari A, et al. Modulation of kinase-inhibitor interactions by auxiliary protein binding: crystallography studies on aurora a

interactions with vx-680 and with tpx2. *Protein Sci.* 2008; 17(10):1791–7.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Medvedev KE, Schaeffer RD, Pei J, Grishin NV. Pathogenic mutation hotspots in protein kinase domain structure. *Protein Science*. 2023;32(9):e4750. <https://doi.org/10.1002/pro.4750>