RESEARCH ARTICLE

Human Mutation

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

Activating Mutations Cluster in the "Molecular Brake" Regions of Protein Kinases and Do Not Associate with Conserved or Catalytic Residues

Miguel A. Molina-Vila,[1]*† Nuria Nabau-Moretó,[2]† Cristian Tornador,[3,4]† Amit J. Sabnis,[5,6] Rafael Rosell,[1] Xavier Estivill,[3,4] Trever G. Bivona,[6] and Cristina Marino-Buslje[7]†

[1]Breakthrough Cancer Research Unit, Dexeus University Hospital, Barcelona, Spain; [2]Computational Genomics Laboratory and Genetics Department, Institut de Biologia Universitat de Barcelona (IBUB), Barcelona, Spain; [3]Genetic Causes of Disease Group, Bioinformatics and Genomics Program, Center for Genomic Regulation (CRG), Barcelona, Spain; [4]Pompeu Fabra University (UPF), Barcelona, Spain; [5]Pediatric Hematology-Oncology, UCSF Benioff Children's Hospital, San Francisco, California; [6]Division of Hematology/Oncology, Department of Medicine, USCF Helen Diller Family Comprehensive Cancer Center, San Francisco, California; [7]Fundación Instituto Leloir, Buenos Aires, Argentina

Communicated by Bruce R. Gottlieb

Received 24 July 2013; accepted revised manuscript 3 December 2013.
Published online 9 December 2013 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22493

ABSTRACT: Mutations leading to activation of proto-oncogenic protein kinases (PKs) are a type of drivers crucial for understanding tumorogenesis and as targets for antitumor drugs. However, bioinformatics tools so far developed to differentiate driver mutations, typically based on conservation considerations, systematically fail to recognize activating mutations in PKs. Here, we present the first comprehensive analysis of the 407 activating mutations described in the literature, which affect 41 PKs. Unexpectedly, we found that these mutations do not associate with conserved positions and do not directly affect ATP binding or catalytic residues. Instead, they cluster around three segments that have been demonstrated to act, in some PKs, as "molecular brakes" of the kinase activity. This finding led us to hypothesize that an auto inhibitory mechanism mediated by such "brakes" is present in all PKs and that the majority of activating mutations act by releasing it. Our results also demonstrate that activating mutations of PKs constitute a distinct group of drivers and that specific bioinformatics tools are needed to identify them in the numerous cancer sequencing projects currently underway. The clustering in three segments should represent the starting point of such tools, a hypothesis that we tested by identifying two somatic mutations in *EPHA7* that might be functionally relevant.

Hum Mutat 35:318–328, 2014. © 2013 Wiley Periodicals, Inc.

KEY WORDS: protein kinases; proto-onocogenes; activating mutations; driver mutations; targeted therapies; cancer

INTRODUCTION

More than 1% of human genes contribute to cancer, a disease that arises as a result of somatic mutations that confer a growth advantage to tumor cells [Futreal et al., 2004]. These mutations are known as "driver" mutations, whereas those that appear incidentally and do not contribute to the tumor phenotype are called "passenger" or "neutral" mutations. Driver mutations can be divided in two groups: (1) "loss-of-function" mutations and (2) activating mutations [Fearon and Vogelstein, 1990]. The "loss-of-function" mutations are those that lead to a complete or partial inactivation of tumor suppressors. Examples of tumor suppressor genes, also called antioncogenes, are *TP53* (MIM #191171), *RB1* (MIM #614041), and *PTEN* (MIM #601728). In contrast, activating or gain-of-function mutations transform proto-oncogenes into oncogenes by inducing increased activity of the corresponding protein, which can be accompanied by a loss of regulation. Proto-oncogenic transcription factors (such as *MYC*; MIM #190080), regulatory GTPases (*RAS* family) and receptor, and cytoplasmic protein kinases (PKs) are activated by this type of mutation [Croce, 2008]. In particular, activating mutations in proto-oncogenic PKs are frequent driver events in many human tumor types [Manning et al., 2002; Futreal et al., 2004; Greenman et al., 2007] and inhibitors of mutated PKs are effective anticancer drugs. Imatinib in gastrointestinal stromal tumors harboring *KIT* (MIM #164920) or *PDGFRA* (MIM #173490) activating mutations [Dagher et al., 2002; Siddiqui and Scott, 2007], gefinitib and erlotinib in lung cancer with *EGFR* (MIM #131550) activating mutations [Pao et al., 2004; Rosell et al., 2009], and vemurafenib and dabrafenib in melanoma-carrying *BRAF* (MIM #164757) activating mutations [Bollag et al., 2010; Hauschild et al., 2012] are all used in clinical practice. Germline gain-of-function mutations of some PKs also give rise to human hereditary disorders, such as raniosynostosis [Pollock et al., 2007] or inherited lymphoedema [Karkkainen et al., 2000].

Thousands of changes in DNA are being identified through the genomic sequencing of human malignancies and other diseases. However, most are likely to be passenger mutations or even polymorphisms [Greenman et al., 2007]. Driver mutations often occur at frequencies indistinguishable from those of passenger mutations [Fröhling et al., 2007; Wood et al., 2007; Loriaux et al., 2008] and discriminating between the two types of mutations is a significant

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

†These authors contributed equally to this work.

*Correspondence to: Miguel A Molina-Vila, Breakthrough Cancer Research Unit, Dexeus University Hospital, C. Sabino Arana 5-19, Barcelona 08028, Spain. E-mail: mamolina@pangaeabiotech.com.

Contract grant sponsors: CONICET (grants PIP1936 and PIP0087).

© 2013 WILEY PERIODICALS, INC.

challenge facing the fields of genomics, cancer biology, and therapeutics that is further compounded by the lack of curated sets of true driver and passenger cancer mutations. Two types of methods have been developed to identify driver mutations, but both fail to discriminate most activating mutations of proto-oncogenic PKs. The first type of method relies on the detection of recurrently altered positions and is limited by the difficulty of assessing the background mutation rate due to the marked variation in mutation frequency among individual tumors [Gonzalez-Perez et al., 2012]. In addition, this type of method is usually unable to detect mutations with a low rate of recurrence, present in only a small fraction of tumors. The second type of method attempts to discriminate driver from passenger mutations by considering evolutionary conservation as well as structural conformation, functional relevance, and other properties of both the original and substituted residues [Reva et al., 2011; Gonzalez-Perez et al., 2012; Hashimoto et al., 2012; Shihab et al., 2013]. The principle underlying this type of method is that a mutation that changes a conserved, functionally, or structurally relevant residue is likely to be a driver. However, this principle can only be properly applied to inactivating mutations and has never been demonstrated in the case of gain-of-function mutations [Shi and Moult, 2011]. In fact, although the activation of some particular PKs by specific mutations has been described in detail, it is not known whether there is a general mechanism through which they exert their functional effects.

Here, we report the results of a novel evidence-based approach whereby we comprehensively analyzed all activating mutations of PKs described in the literature. We found that they constitute a distinct group of driver mutations that do not directly affect conserved, catalytic, or ATP-binding residues. Instead, they cluster in three segments that act as "molecular brakes" of the kinase activity in class III–IV receptor tyrosine kinases (RTKs) [Chen et al., 2007, 2013]. This finding led us to hypothesize that an autoinhibitory mechanism mediated by structurally equivalent "molecular brakes" is present in all PKs. Our results also indicate that current methods to identify driver mutations are not useful in the case of activating mutations of proto-oncogenic PKs and that new, specific tools are needed.

Material and Methods

Systematic Search for Primary Activating Mutations in Human PKs

To compile our comprehensive list of activating mutations in PKs, we performed a two-step systematic search for each of the 518 PKs present in the "complete kinase" study of the Catalogue of Somatic Mutations in Cancer (COSMIC) database [Forbes et al., 2011]. First, we introduced the word "mutation" together with the name of the kinase in PubMed. If the number of articles retrieved was higher than 500, we added the terms "activating," "gain-of-function," and "constitutive activation." We then reviewed all the articles that appeared. Next, we computationally calculated a relative frequency for all mutations in the 518 PKs by dividing the number of tumors carrying the mutation by the total number of tumors where the corresponding gene has been sequenced (according to COSMIC) and then multiplying the resulting figure by 1,000. Finally, all mutations with a relative frequency above two (0.2%) were checked in PubMed by introducing the name of the mutation (e.g., p.P267R).

Only primary mutations with experimental evidence demonstrating their activating role were included in our database. EGFR mutations conferring a response rate to erlotinib higher than 50%, according to the EGFR somatic mutations database (http://www.somaticmutations-egfr.info/), were also added. As a result of our search, we found 407 primary activating mutations in 41 PKs (Supp. Table S1). For each mutation, we included the PubMed reference describing it as activating, as well as the relative frequency we had calculated for it, and the disease(s) where it was first described. When a mutation was absent in human cancers (according to the COSMIC database), it was assigned a relative frequency of 0. The relative frequency of the T790M mutation of EGFR could not be calculated since COSMIC does not differentiate primary and secondary mutations. To facilitate further analyses, for each PK in our list, we compiled its Uniprot [Consortium, 2012] and NCBI Reference Sequence accession, E.C. number (http://www.chem.qmul.ac.uk/iubmb/enzyme/), and Pfam accession [Punta et al., 2012] of the kinase domain.

During the compilation of our list and subsequent analyses, we detected some inconsistencies in the databases that could easily lead to confusion. For example, in the case of the hepatocyte growth factor receptor (MET), the canonical sequence according to Uniprot is isoform 1, but COSMIC uses isoform 2 as a reference. Moreover, in other cases, including that of EGFR, the numbering of residues in the protein data base (pdb) differs from the numbering in Uniprot. In addition, some genes have different names in COSMIC and Uniprot (e.g., TIE2 = TEK and KPCG = PRKCG). In our study, we systematically used the Uniprot names, canonical forms, and numbering.

Sequence Retrieval, Multiple Sequence Alignment, and Frequency Calculations

We retrieved 1,377 tyrosine kinase (TK) sequences (E.C number: 2.7.10., classification) and 4,420 serine/threonine kinase (STK) sequences (E.C number 2.7.11.) belonging to all species from the Uniprot database. Kinase domain boundaries were those defined in the database.

For both families of kinases, multiple sequence alignment (MSA) was performed with muscle, provided in T_coffee version 9.02.r1228 (www.tcoffee.org) [Taly et al., 2011]. To map the mutations in the kinase domain, the EGFR_HUMAN receptor was taken as the reference sequence and 3D structure for TKs (Uniprot accession P00533, pdb code 1M14), and the BRAF_HUMAN as the reference for STKs (Uniprot accession P15056, pdb code 4E26). The juxtamembrane (JM) region of RTKs was defined as the sequence between the first cytoplasmic residue and the first kinase domain residue. For the JM region of the PDGFR subfamily, the KIT_HUMAN was selected as the reference sequence and structure (Uniprot accession P10721, pdb code 1T45). Although hundreds of kinase domain structures are known, none of the STKs have the activation loop complete and only a few crystallographic structures include the complete JM region, so partial mapping was necessary in some cases.

A position was labeled as mutated when looking at the entire family alignment, at least one activating mutation was found in this position. When a deletion or insertion was present, all the altered positions were labeled as mutated (e.g., for EGFR_HUMAN deletion p.K746_E750del, the positions 746, 747, 748, 749, and 750 were labeled). Mutated positions were then mapped by projecting the MSA onto the 3D reference structure (EGFR, BRAF, or KIT). The number of different activating mutations and the number of mutated kinases were then computed for every position, and a normalized frequency of activating mutations per position was calculated. First, an accumulated relative frequency was obtained by adding the relative frequencies of all activating mutations affecting a particular position. This frequency was then normalized according to the following

formula:

$$\text{Normalized frequency} = \frac{\log_{10}([\text{accumulated relative frequency in a position}] \times 100{,}000)}{\Sigma \text{ accumulated relative frequencies in all positions}}$$

Normalized frequencies of activating mutations in each position were represented, as well as the absolute numbers of mutated kinases and different activating mutations in each position. Molecular graphics were generated with the UCSF Chimera package [Pettersen et al., 2004].

The same process was repeated for mutations of unknown effects. We retrieved all COSMIC mutations from the 41 PKs under study and eliminated the activating mutations using a perl curated script. We also eliminated the mutations located in activating positions that were poorly defined in the COSMIC database (e.g., mutations p.L858? in EGFR or p.V600? in *BRAF*). The total number of mutations of unknown effects, the number of kinases affected, and the normalized frequencies were calculated for every position and represented. The normalized frequencies of activating mutations and mutations of unknown effects were also used to calculate central moving averages with $n = 13$, which were subsequently plotted.

Conservation

Conservation was calculated as the Kullback–Leibler relative entropy [Cover and Thomas, 1991] using an amino acid background frequency distribution obtained from the Uniprot database [Consortium, 2012]. The MSAs were first corrected for sequence redundancy using sequence weighting by 62% identity clustering and including pseudocounts to correct for low counts [Marino-Buslje et al., 2010].

Statistical Calculations

We performed a multistep statistical analysis of the distribution of mutations within the kinase domain of TKs, using the R package program. First, we represented box plots from the normalized frequencies of activating mutations and mutations of unknown effects. The plots were subsequently recalculated excluding EGFR (see Supp. Figs.).

We then used two nonparametric tests to determine whether the distributions of the normalized frequencies of mutations were significantly different: Wilcoxon and Kolmogorov–Smirnov tests. We were unable to perform a normality test for two reasons. First, all frequency values were associated with a particular position within the amino acid sequence of the kinase domain and were not interchangeable. Second, there were many positions with no activating mutations and, therefore, with a value of zero. The results of the Wilcoxon and Kolmogorov–Smirnov tests are shown in the Supporting Information.

To examine whether, in contrast to putative passenger mutations, activating mutations are preferentially located in segments HS2 and HS3, we calculated 2×2 contingency tables with the sum of the normalized frequencies and the number of different mutations per position and applied a two-tailed Fisher's exact test. Finally, we studied the correlation of the distribution of mutations within each HS segment using the Spearman test (see Supporting Information).

Colony Transformation Assays

A human cDNA for *EPHA7* (MIM #602190; GenBank accession: BC126151.1) was purchased from Open Biosystems (Waltham, MA). The p.D776N and p.S684I mutations were created using a QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA) as per the manufacturer's instructions with the following primers:

EPHA7-D776N-F: 5′-ctcgtttgtaaagtgtcaaattttggcctgtcccgag-3′
EPHA7-D776N-R: 5′-ctcgggacaggccaaaatttgacactttacaaacgag-3′
EPHA7-S684I-F: 5′-gacttttgtgtgaagcaatcatcatggggcagtttgac-3′
EPHA7-S684I-R: 5′-gtcaaactgccccatgatgattgcttcacacaaaaagtc-3′

Wild-type *EPHA7* and the *EPHA7-D776N* and *EPHA7-S684I* constructs were then cloned into the pBabe construct. These constructs, as well as a pBabe-*HRAS-G12V* and an empty pBabe vector, were then transduced into 293-GPG cells using FuGene (Promega, Madison, WI) as per the manufacturer's instructions. Supernatant containing amphotropic retrovirus was harvested 72 hr following transfection and used to transduce NIH3T3 cells with 4 μg/mL of polybrene. After transduced cells reached confluence, they were selected with 1 μg/mL puromycin for 7 days.

7.5×10^3 cells were then plated in 10 replicates in a CytoSelect 96-well soft agar colony transformation assay (Cell BioLabs, San Diego, CA) as per the manufacturer's instructions. After 1 week, the agar was dissolved and cells were lysed and then incubated with CyQuant dye. Fluorescence was quantitated with a plate reader using a 485/520-nm filter set. The significance of differences in fluorescence intensity was calculated using a student's *t* test. Equivalent levels of EphA7 expression in 3T3 cells were confirmed by Western blotting. Whole cell lysates from transfected cells were loaded onto a polyacrylamide gel and blotted with anti-EphA7 (sc-1015, Santa Cruz Biotechnology, Santa Cruz CA) and anti-actine (A2228, Sigma–Aldrich, St. Louis, MO) antibodies. For the analysis of signal transducers, cells were plated in DMEM with 0.5% fetal bovine serum. After 12 hr of serum starvation, whole cell lysates were harvested and loaded on a polyacrylimide gel, and blotted with antibodies against ERK, p-ERK (Thr202/Tyr204), STAT3, p-STAT3 (Ser727), Akt and p-Akt (all from Cell Signaling Biotechnologies, Danvers, MA), and anti-EPHA7.

Online Prediction Tools

Three online tools to predict the functional impact of mutations were tested: MutationAssesor (http://mutationassessor.org/) version 1.0, release 1 [Reva et al., 2011], TRANSformed Functional Impact for Cancer (transFIC) (http://bg.upf.edu/transfic/home) version 1.0 [Gonzalez-Perez et al., 2012], and Functional Analysis through Hidden Markov Models (FATHMM) (http://fathmm.biocompute.org.uk./index.html) version 2.3 [Shihab et al., 2013]. In all cases, the instructions provided in the corresponding Websites were followed. These tools can only be applied to missense mutations. In consequence, they were run against the 250 activating missense mutations of PKs of our curated set. The results obtained on the Websites were downloaded and analyzed.

The Mutation Assesor categorizes predicted functional impact of mutations as "low," "medium," or "high." The transFIC uses the same categorization but offers three different results per mutation, based on the transformation of the scores given by three well-known tools (SIFT, http://sift.jcvi.org/; PoliPhen 2, http://genetics.bwh.harvard.edu/pph2/dokuwiki/downloads; and MutationAssesor). Finally, FATHMM classifies mutations as "cancer"-associated or "passenger/other." It is the only tool that offers the option of changing one of the parameters, the prediction threshold. Using the default value (–0.75), basically all mutations are considered "cancer,"

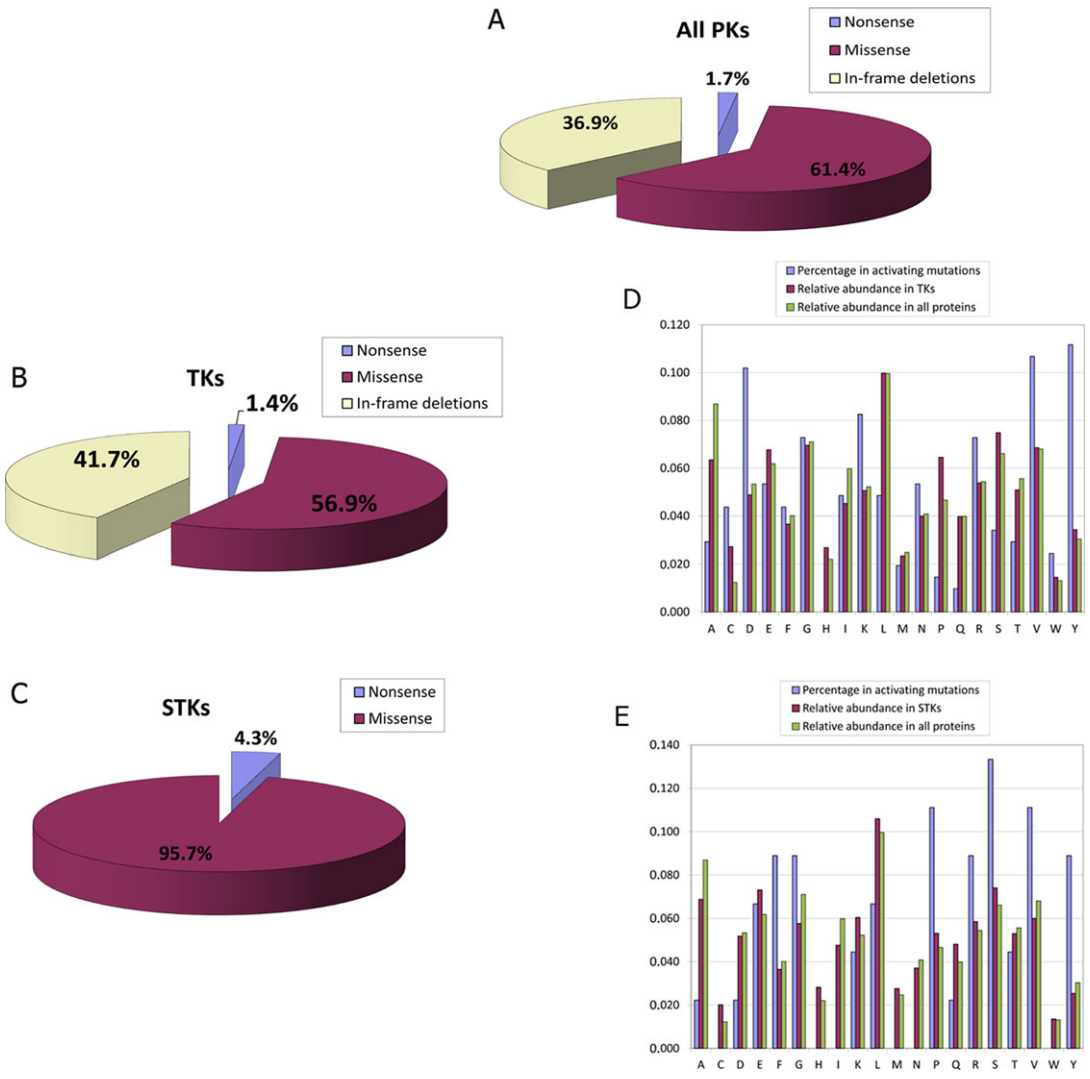

Figure 1. Description of activating mutations in PKs. Classification of activating mutations in (**A**) all PKs, (**B**) TKs, and (**C**) STKs. Amino acids affected by missense activating mutations in (**D**) TKs and (**E**) STKs. The blue bars indicate the percentage of activating mutations that alter each amino acid. The purple bars represent the relative abundance of the amino acid in TKs or STKs. Capital letters at the bottom of the bars represent the amino acid.

and the instructions of the Website suggest a more discriminating threshold of –3.0, which we selected for our analysis.

Results

Description of the Activating Mutations Reported in PKs

Based on our search of the literature and the COSMIC database [COSMIC; Forbes et al., 2011], we compiled a curated, comprehensive list of all the primary point mutations and short insertion/deletions in PKs that have been shown to lead to constitutive activation. We found 407 primary activating mutations: 360 in 27 TKs, 41 in 11 STKs, and six in three dual-specificity kinases (Supp. Table S1). The majority of these mutations are cancer associated. As expected, gain-of-function mutations in "druggable" kinases, such as BRAF (18 mutations), KIT (79 mutations), and EGFR (66 mutations), have been extensively studied and reported in the literature and in COSMIC. Some of the cancer-associated mutations have frequently been detected, especially in specific malignancies, whereas others have been found in only a few tumor samples. For this reason, a relative frequency was calculated for each activating mutation (Supp. Table S1). The 79 activating mutations that did not appear in COSMIC, some related to rare genetic diseases and a few artificially generated, were assigned a relative frequency of zero.

Of the 407 activating mutations, 250 (61%) are missense, 150 (37%) in-frame insertions/deletions, and 7 (2%) nonsense mutations, which are much more likely to produce an inactive protein (Fig. 1A–C). Activating in-frame insertions/deletions occur exclusively in seven TKs. Interestingly, 45% of the activating missense mutations in TKs affect a C, D, K, V, or Y, although these five residues represent only 22% of the total amino acid content of the TKs under study. In contrast, only 17% of the activating missense mutations in TKs affect A, H, L, P, Q, S, or T, which represent 42% of the total amino acid content (Fig. 1D). The spectrum seems to be somewhat different in STKs, where 62% of activating missense mutations affect F, G, P, S, V, or Y, which together constitute 31% of the total amino acid content, and only 13% affect A, C, D, H, I, L, M, N, Q, or W, which constitute 45% of the total amino acid content (Fig. 1E).

Table 1. The Three Hypermutated Segments (HS) Clustering Activating Mutations in TKs

Hypermutated segment	HS1	HS2	HS3
Position (referred to EGFR)	673–720[a]	745–791	833–871
Subdomains[b]	JM + I (N-terminal)	II (C-terminal) + III + IV + V (N-terminal)	VIB + VII + VIII (two amino acids N-terminal)
Total number of residues	48	47	39
Number of residues affected by activating mutations	45 (94%)	33 (70%)	23 (59%)
Number of TKs with activating mutations	6	15	15
Number of different activating mutations described	133	76	71

[a]Residues 550–597 of c-KIT.
[b]Subdomains in the kinase domain are defined according to Hanks and Hunter (1995)

Activating Mutations of TKs Cluster in Two Hypermutated Segments Within the Kinase Domain (HS2 and HS3), Whereas Mutations of Unknown Effect are Almost Uniformly Scattered

In TKs, 179 activating mutations in 21 proteins were located in the kinase domain. We mapped these mutations by projecting the MSA of the entire TK family onto the 3D structure of the human EGFR. We selected EGFR as a reference because it is a well-established genetic driver of tumorigenesis in several human tumor types, it is a TK in which numerous activating mutations have been described, and it has been characterized structurally in numerous crystallographic studies. An accumulated frequency of activating mutations was then calculated for each position in the amino acid sequence and subsequently normalized. We successfully mapped 176 of the 179 mutations; the other three mutations, affecting residues with no equivalent position in EGFR, were excluded from the analysis. We found that 147 of the 176 mutations (84%) were clustered in two hypermutated segments (HS2 and HS3) (Table 1). This clustering was apparent both in terms of normalized frequencies (Fig. 2A–C)

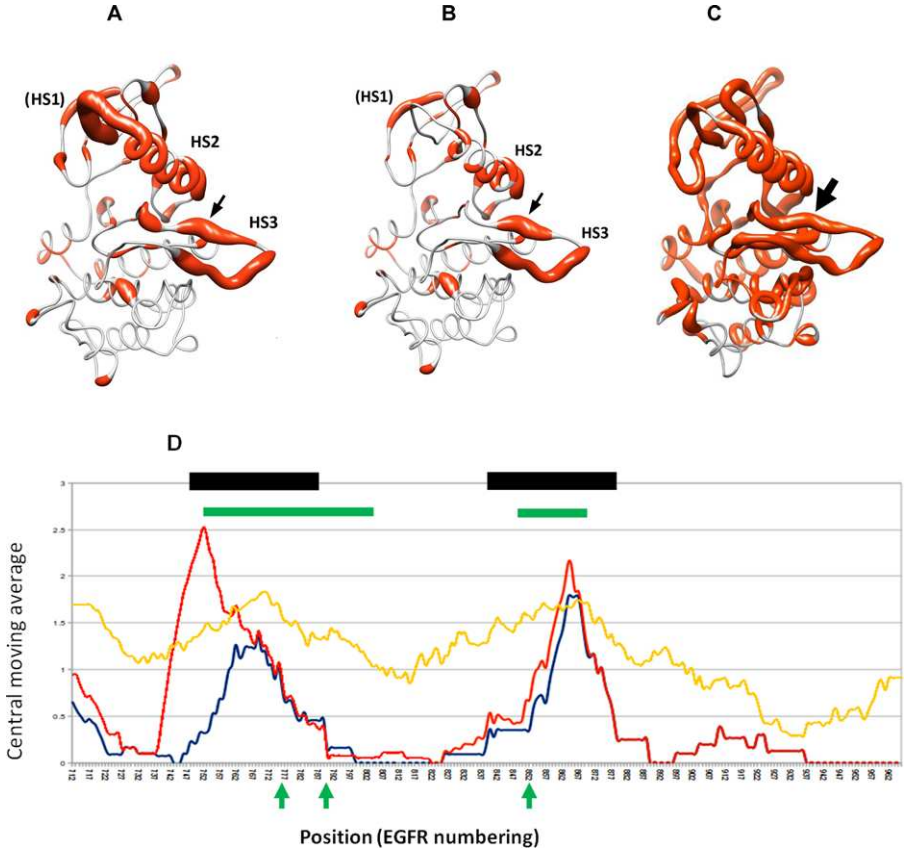

Figure 2. Activating mutations cluster in two hypermutated segments of the kinase domain. **A**–**C**: Distribution of mutations in the kinase domain of TKs, depicted in the ribbon representation of the human EGFR structure (pdb code 1M14). Mutated positions are shown in orange, and the diameter of the ribbon is proportional to the normalized frequency of mutations affecting each position, ranging from 0 to 4.43. **A**: Activating mutations in all TKs analyzed. **B**: Activating mutations in TKs excluding EGFR. **C**: Mutations of unknown effect. The position of hypermutated segments HS2, HS3, and the C-terminus of the segment HS1 is shown. The arrow indicates the hotspot L861 in EGFR (V600 in BRAF). **D:** Central moving average (n = 13) plot of the normalized frequencies of mutations in each position of the kinase domain. In red, activating mutations in all TKs analyzed; in blue, activating mutations in TKs excluding EGFR; in yellow, mutations of unknown effects. The black bars show the position of the hypermutated segments HS2 and NS3. A minor clustering was also apparent in the case of mutations of unknown effect, perhaps because some may be as yet unrecognized activating mutations. Significant differences were observed between the distribution of activating mutations and that of mutations of unknown effect as well as between the distribution of activating mutations excluding EGFR and that of mutations of unknown effect excluding EGFR (Wilcoxon and Kolmogorov–Smirnov $P < 2.2\ 10–16$). The red bars show the location of the molecular brake regions that have been described in class III–IV RTKs. The arrows indicate the three key residues involved in the network of hydrogen bonds of the brake (positions 549, 565, and 641 in FGFR2; corresponding to 776, 791, and 852 in EGFR).

Table 2. The Three Hypermutated Segments (HS) Clustering Activating Mutations in TKs, Excluding EGFR Mutations

Hypermutated segment	HS1	HS2	HS3
Position (referred to EGFR)	673–720[a]	755–791	846–871
Subdomains[b]	JM (all) + I (N-terminal)	III + IV + V (N-terminal)	VIB (two amino acid C-terminal) + VII + VIII (two amino acid N-terminal)
Total number of residues	48	37	26
Number of residues affected by activating mutations	38 (90%)	18 (47%)	15 (58%)
Number of TKs with activating mutations	5	14	14
Number of different activating mutations described	120	40	56

[a]Residues 550–597 of c-KIT.
[b]Subdomains in the kinase domain are defined according to Hanks and Hunter (1995)

and in terms of the absolute number of different mutations and TKs affected (Supp. Fig. S1). The distribution could also be visualized in a central moving average plot, which gave a two-peak profile (Fig. 2D). In contrast, when we mapped the hundreds of COSMIC mutations in the TKs under study that have not been described as activating, they were almost uniformly scattered within the kinase domain (Fig. 2, Supp. Fig. S1). The distribution of these putative passenger mutations was significantly different from that of the activating mutations by both the Wilcoxon and the Kolmogorov–Smirnov tests ($P < 2.2\ 10^{-16}$) (Supp. Fig. S2, Supp. Table S2), and the clustering of the activating mutations in the HS2 and HS3 was statistically significant in a two-tailed Fisher's exact text ($P < 0.0001$) when compared with the distribution of the putative passenger mutations (Supp. Table S3).

The HS2, which clusters 76 activating mutations in 15 different kinases, comprises the C-terminal half of the subdomain SDII, the entire SDIII (αC-helix) plus SDIV (kinase hinge), and the N-terminal half of the SDV. The HS3, which clusters 71 mutations in 15 TKs, includes the subdomains VIB (catalytic loop), VII (activation loop), and two N-terminal residues of VIII (P+I loop) (Table 1 and Supp. Table S4). In the HS2 segment, a hotspot in position 776 clusters seven activating mutations in four kinases, and in the HS3 segment, a hotspot in position 861 clusters 22 activating mutations in six different TKs. In silico studies have shown that both HS2 and HS3 are regions where nonsynonymous, cancer-associated single-base changes in PKs are preferentially located and have also reported the hotspot in position 861 and a uniform distribution within the kinase domain of common, nondisease-associated single-base changes [Torkamani et al., 2008a; Dixit et al., 2009; Lee et al., 2009].

The EGFR TK alone accounts for 54 mutations in the kinase domain. When we compared the distribution of activating mutations within the HS2 including and excluding EGFR, the Spearman's coefficient showed a weak correlation ($P = 0.3$) (Supp. Table S5). While 30 activating mutations are present in the SDII of EGFR, only one is present in the remaining 20 TKs. Thus, activating mutations in the SDII, most of which are in-frame insertions/deletions, seem to be a particularity of EGFR (Fig. 2, Supp. Fig. S1). In addition, EGFR accounts for half of the activating mutations in the catalytic loop (VIB region). In consequence, if EGFR is excluded, the two HS could be redefined to comprise fewer amino acids (Table 2 and Supp. Table S4).

Activating Mutations of TKs Do Not Affect Catalytic or ATP-Binding Residues and Do Not Correlate with Conserved Positions

Of the 179 activating mutations within the kinase domain in our curated set, only five affect either the ATP binding (K745 and D855) or the key catalytic residues (D837, A839, R841, and N842, in human EGFR numbering) [Porter et al., 2004]. Four of these five mutations have been described in EGFR (p.K745_E749del, p.K745_A750del, p.A839T, and p.D855G) and one is an artificial mutation generated in NTRK1 (p.D668N). A significant number of congenital disease-associated single-nucleotide polymorphisms in PKs have also been reported to affect residues not directly involved in ATP binding or catalysis but rather buried in the catalytic core [Torkamani et al., 2008b].

To examine whether activating mutations are associated with conserved residues, we calculated conservation, as measured as the Kullback–Leibler divergence score [Cover and Thomas, 1991; modified by Marino-Buslje et al., 2010], for each position in the TK domain. Subsequently, we evaluated the predictive potential of the calculated conservation of the residues in terms of the area under the roc curve (AUC). An AUC of 1 means perfect predictive value, whereas 0.5 indicates a random process. We obtained an AUC value of 0.4, indicating that conservation scores do not correlate with positions harboring activating mutations. In fact, these positions do not have a specific pattern of conservation (Fig. 3 and Supp. Fig. S3).

An Additional Hypermutated Segment (HS1) is Apparent in RTKs, Particularly of the PDGFR Family

One hundred and twenty-one activating mutations were located in the JM region of the four RTKs analyzed belonging to the PDGFR family, plus EGFR and RET, 72 of which corresponded to KIT and 34 to FLT3. Twelve additional mutations mapped to the first nine residues of the kinase domain. This led us to define another hypermutated segment (HS1), comprising the JM region plus the N-terminus of the SDI (Tables 1 and 2). The JM region of the four PDGFR family receptors was aligned, and activating mutation positions were mapped onto the KIT structure. Most of the mutations were located in residues 551–578 (Fig. 4A, Supp. Fig. S4). EGFR and RET were excluded from the alignment due to the significant differences in the JM amino acid sequence.

Finally, two isolated short segments that cluster activating mutations were identified in the extracellular domain of the three FGF receptors analyzed, mapping in residues 248–249 and 370–373 of FGFR3. They encompassed three and six point mutations, respectively, almost all of which resulted in a new cysteine residue.

Activating Mutations in STKs Also Seem to Cluster in the Hypermutated Segments HS1–HS3

In STKs, 23 mutations in six different proteins were located within the kinase domain. In this case, we used human BRAF as a reference structure in the MSA. Due to the low number of activating mutations reported in STKs, statistical analysis was impossible and caution must be exercised when analyzing our results. However, activating mutations in STKs also seem to cluster in the HS3 segment defined for TKs (10 mutations) and, to a lesser extent, in HS1 and HS2 (five and four mutations, respectively) (Fig. 4B and Supp.

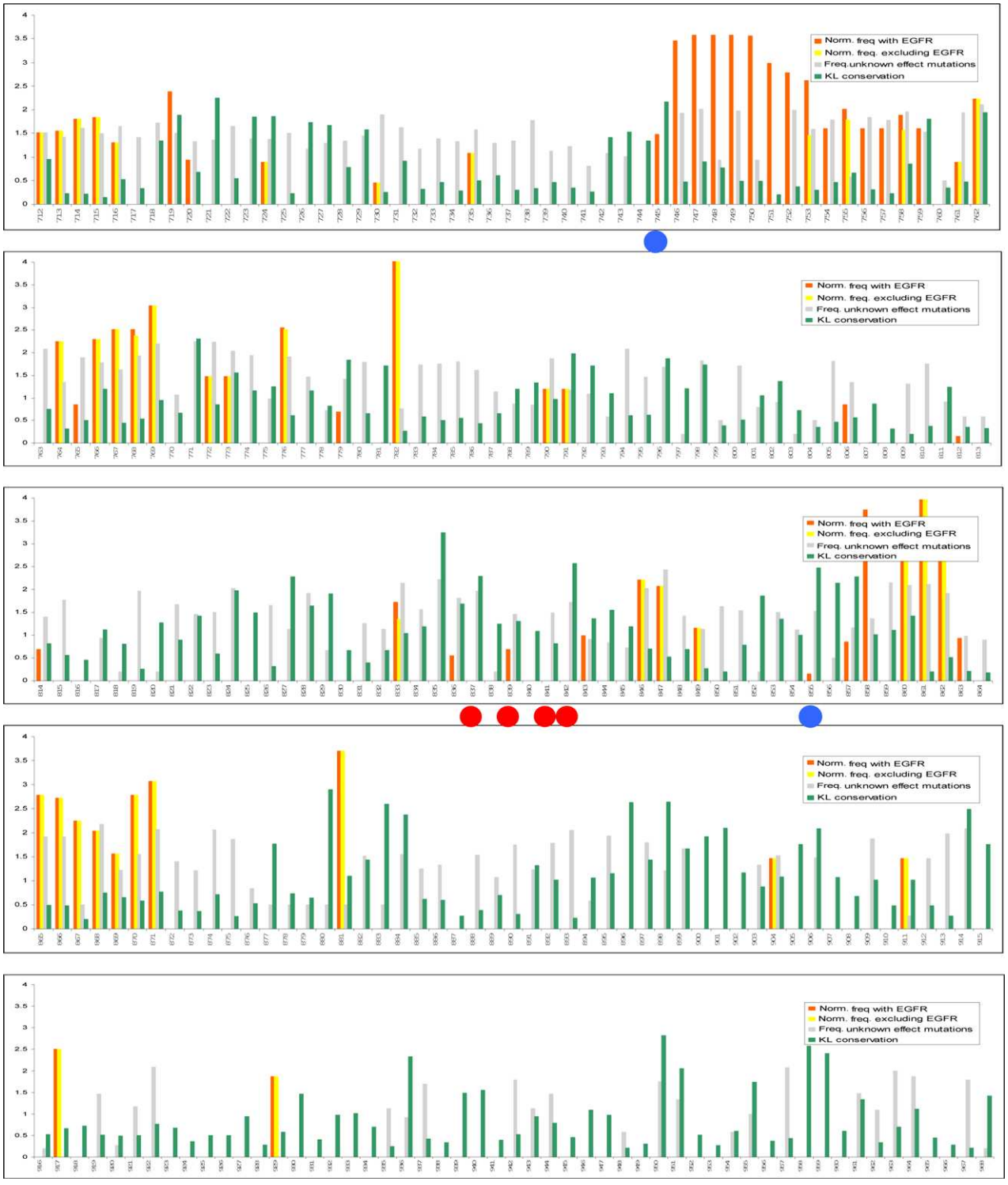

Figure 3. Activating mutations do not directly affect ATP binding or catalytic residues and do not concentrate in conserved regions. Plot showing the normalized frequency of activating mutations (with EGFR as reference) in all TKs (orange bars) and in TKs excluding EGFR (yellow bars), the KL conservation score (green bars), and the frequency of mutations of unknown effect in each TK position (gray bars). Blue dots indicate ATP-binding sites; red dots, catalytic residues. Activating mutations are present in conserved, partially conserved, and variable regions

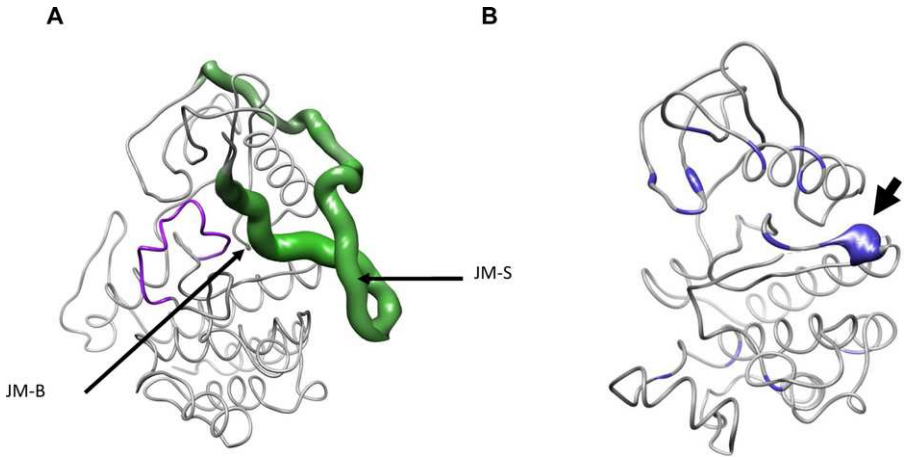

Figure 4. Activating mutations in the JM domain and in STKs. **A**: Distribution of activating mutations in the JM domain of RTKs belonging to the PDGFR family. Activating mutations are depicted in the ribbon representation of the autoinhibited human c-KIT structure (pdb code 1T45). Mutated positions are shown in green, and the diameter of the ribbon is proportional to the normalized frequency of activating mutations affecting each position, ranging from 0 to 3.33. The approximate positions of the JM binding and the JM switch motifs are shown, and the c-kit activation loop is shown in purple (see also Supp. Fig. S4). **B:** Distribution of activating mutations in STKs. Mutations are depicted on the ribbon representation of the human BRAF structure (pdb code 4E26). Mutated positions are shown in dark blue, and the diameter of the ribbon is proportional to the normalized frequency of activating mutations affecting each position, ranging from 0 to 4.12. The arrow indicates the position of the V600 hotspot in BRAF (L861 in EGFR).

Tables S4 and S6). In addition, the most frequently mutated residue in BRAF (V600) overlaps with the 861 activating mutation hotspot of TKs.

Two Newly Described Mutations of Epha7 Located in an HS Might Have a Functional Effect

Our findings seemed to indicate that mutations in PKs located within a hypermutated segment are more likely to be drivers and should thus be prioritized for validation experiments needed to prove their functional effects. We tested this hypothesis on the 66 mutations of unknown functional effects in 43 TKs described by three whole-genome sequencing reports of human tumors [Kan et al., 2010; Puente et al., 2011; Zhang et al., 2012]. Only 17 of these mutations, affecting 16 TK genes, mapped within a hypermutated segment (Supp. Table S7). We selected two mutants in *EPHA7* (p.S684 and p.D776N) for validation since the role of mutant forms of Epha7 in tumorigenesis is unclear.

We used a quantitative, fluorescence-based growth assay to determine the functional effects of expression of wild-type Epha7 and the two mutant forms. Consistent with prior data, we found that expression of HRASG12V induced growth, whereas expression of wild-type Epha7 was growth suppressive [Oricchio et al., 2011]. Two Epha7 bands were apparent by Western blotting of the cell lysates; the upper band, which corresponds to the full-length protein, and the lower band, which corresponds to a truncated receptor. Truncated forms of Epha7 lacking the kinase domain have been reported to inhibit Epha7 and Epha2 activation and to have tumor suppressor properties [Holmberg et al., 2000; Oricchio et al., 2011]. The growth-suppressive effects of Epha7 were reverted by introduction of the p.S684I or p.D776N mutation. Epha7 wild-type and mutant forms were expressed at equivalent levels in these cells (Supp. Fig. S5). Eph receptors are known to have paradoxical effects, and they can both promote and inhibit tumorigenicity by triggering a variety of cell-signaling pathways [Pasquale, 2010]. We analyzed three signaling effectors in our transfected cells, and we found that the p.D776N mutant protein was significantly more efficient than Epha7 wild type in inducing the phosphorylation of STAT-3. Taken together, these data suggest that p.S684I and p.D776N in Epha7 can be functionally relevant and deserve further experimental validation as putative activating mutations.

Discussion

In this study, we have compiled the first curated set of activating mutations in PKs, comprising 407 mutations in 41 different proteins. These data are disseminated in hundreds of articles describing one or few cases each and in several data bases. We have found that most activating mutations are missense, with a significant percentage of small insertions/deletions that affect exclusively seven TKs. We have next assessed their equivalence through MSA and structural superimposition to evaluate the frequency of mutated positions or segments through all the kinase family. Using this approach, we have discovered that activating mutations cluster in three hypermutated segments (HS1, HS2, and HS3), whereas mutations of unknown effects are almost uniformly scattered throughout the kinase domain of PKs. We have also found two hotspots harboring activating mutations, one located in the HS2 (position 776, EGFR numbering) and another in the HS3 (position 861, EGFR numbering).

PKs control key biological processes and their activity is tightly regulated. In their basal state, most PKs are autoinhibited and possess very low levels of intrinsic kinase activity but are activated in nonpathological conditions by phosphorylation and/or binding of ligands or activator proteins. A so-called molecular brake responsible for this autoinhibited state has been experimentally demonstrated in FGFR2, and structural analyses have observed it in other class III–IV RTKs [Chen et al., 2007, 2013]. Key regulatory regions of the FGFR2 kinase domain (the hinge plus αC-helix and the activation loop) have been shown to act in concert to maintain the autoinhibited state through a network of hydrogen bonds between the triad of residues E565 (in the kinase hinge), N549 (in the loop after the αC), and K641 (immediately before the activation loop) (Supp.

Fig. S6). When wild-type FGFR2 is activated after ligand binding and phosphorylation of the A loop, the network dissociates, allowing two critical changes in the kinase domain. The N-lobe of the kinase undergoes an inward rotation (6.7°) toward the C-lobe, which coincides with a major rearrangement of the phosphorylated A-loop at the local level. These coupled structural changes align the catalytic residues from different regions, including the A-loop, the catalytic loop and the αC helix, to promote peptide substrate and ATP binding and to increase catalytic efficiency. As a result, the kinase domain toggles from the autoinhibited, structurally rigid state to the active, more dynamic and conformationally heterogeneous state. Eleven pathogenic activating mutations of FGFR2 have been demonstrated to release the molecular brake, either by directly disrupting the network of hydrogen bonds (p.N549H, p.N549T, p.E565G, p.E565A, and p.K641R) or by indirectly disengaging it through allosteric communication (p.K526E and p.K659E/M/N/Q/T). This molecular brake does not seem to be exclusive to TKs; crystallographic analyses of the STK PAK1 have revealed the presence of a similar network of hydrogen bonds in the kinase domain that is disengaged in the activated PAK1 structure [Lei et al., 2000, 2005].

HS2 and HS3 coincide with the two molecular brake regions described in class III–IV RTKs. The HS2 comprises the SDIII (αC-helix), the SDIV, and the N-terminus of the SDV (hinge). The HS3 includes the SDVII (activation loop) together with the residues located in its immediate vicinity (Fig. 2D). The three residues forming the network of hydrogen bonds of the brake are all within one of the hypermutated segments (Supp. Fig. S6), similarly to the other two residues affected by the activating mutations of FGFR2 mentioned above (K526 and K659). In addition, the two mutational hotspots, 776 in HS2 and 861 in HS3, are located in residues that play a particular role in the brake: 776 (N549 in FGFR2) directly participates in the network of hydrogen bonds [Chen et al., 2007], and L861 in EGFR packs against the αC-helix, preventing the formation of an ion pair (K745/D772) associated with the active conformation [Zhang et al., 2006].

In the RTKs of the PDGFR family, the JM domain has also been demonstrated to inhibit the kinase domain, with two motifs playing a prominent role: the JM binding (JM-B) and the switch motif (JM-S) (residues 553–559 and 560–571 in c-KIT) [Chan et al., 2003; Griffith et al., 2004]. Our third hypermutated segment (HS1) coincides with the JM domain, and moreover, mutations within HS1 cluster around the JM-B and the JM-S motifs (Fig. 4A, Supp. Fig. S4).

Taken together, our findings lead us to hypothesize that an autoinhibitory mechanism mediated by structurally equivalent molecular brake regions is present in all PKs, and that the vast majority of activating mutations found in human tumors and other pathologies act by releasing this brake and consequently shifting the dynamic equilibrium of the PK toward the active form. The fact that this active form is a more conformationally dynamic and heterogeneous state can help to explain why a variety of pathogenic mutations in a wide range of residues within the molecular brake regions can result in activation of the kinase domain.

The molecular brake hypothesis can also account for the activating deletions in the SDII exclusive of EGFR. In EGFR, the last residues of the SDII form a loop that stabilizes the αC-helix in a displaced position and interacts closely with the inactive helical conformation of the A loop [Zhang et al., 2006; Shan et al., 2012]. Therefore, the SDII of EGFR acts as an additional "molecular brake" that does not seem to be present in other PKs, and disappearance of the loop due to cancer-associated deletions has an activating effect.

With the advent of high-throughput sequencing, large numbers of somatic mutations are being discovered in cancer-sequencing projects, exceeding our capacity to validate their effect through experimental functional studies and making it essential to have bioinformatics tools to predict the impact of the mutations and prioritize them for further analysis. In particular, identifying driver mutations is one of the main goals of genome resequencing, and many computational methods have been developed for this purpose. So far, these methods have only considered missense mutations and have not differentiated between inactivating/loss-of-function and activating/gain-of-function mutations, using the same criteria and algorithms for both types of drivers. In addition, due to the lack of curated sets of driver mutations, these computational methods have been validated on mutations with pronounced phenotypic effects that usually involve a loss of function of the mutated gene.

The first high-throughput methods developed to identify driver mutations relied on the detection of recurrent alterations, calculating the probability of detecting by chance the frequency of a particular mutation across the tumor samples analyzed. However, these methods have several limitations; the background mutation rate is difficult to assess correctly, and genes that are mutated in only a small fraction (<1%) of tumors can still act as drivers [Wood et al., 2007]. In addition, they are likely to favor early driver genes over those that are mutated at a later stage of tumor progression. In recent years, new methods have been developed to classify mutations based on data other than the frequency of a mutation. These methods can only be applied to missense mutations and are based on the assumption that driver mutations will preferentially affect conserved residues. Therefore, they rely more or less on sophisticated methods to calculate the conservation of individual residues. Three of these methods offer their predictions online through their Web Pages: MutationAssesor [Reva et al., 2011], transFIC [Gonzalez-Perez et al., 2012], and FATHMM [Shihab et al., 2013] (see *Material and Methods*). The MutationAssessor assigns "functional impact scores" to residue changes using evolutionary conservation patterns derived from protein family MSAs, whereas transFIC offers three scores per mutation, complementing the evaluation of violations of evolutionary constraints by other tools with a "baseline tolerance" of germline alterations. Finally, FATHMM assigns a score by interrogating sequence conservation through the underlying amino acid probabilities modeled by the internal match states of several hidden Markov models.

When we tested the 250 activating missense mutations of PKs of our curated set in MutationAssessor, only 16 were assigned a "high" functional impact, whereas mutations such as p.V600E in BRAF or p.T790M and p.L861Q in EGFR were qualified as "neutral" or "low impact." TransFITC only categorized 92 mutations as "high impact" in at least two of its three scores, whereas p.V600E and p.V600K in BRAF and p.T790M and p.L861Q in EGFR were again only assigned "medium" or "low" impact scores. With the FATHMM discriminating threshold of –3.0, 35 mutations were scored as "cancer-associated," whereas the remaining 215 were qualified as "passenger." These three tools thus seem unable to recognize well-established activating mutations, leading us to doubt their ability to screen whole-genome sequencing data in search of new gain-of-function mutations of proto-oncogenic PKs.

In the present study, we have demonstrated that these activating mutations do not correlate with conserved positions and that, in fact, positions harboring activating mutations do not have a specific pattern of conservation. Therefore, although conservation-based methods might be useful in the case of loss-of-function alterations, we cannot recommend their use for predicting gain-of-function mutations; at least in PKs. Conserved residues usually play key roles either in catalysis or folding of proteins. They have been optimized by a long evolutionary process and replacing them with another

residue will almost invariably lead to a less active protein, either through direct loss of catalytic activity or through incorrect folding. In fact, it is well known that replacement of such residues by site-directed mutagenesis is usually deleterious.

Our work demonstrates that gain-of-function mutations of PKs represent a distinct, homogeneous group of drivers and that specific computational methods should be developed to discriminate them. The new methods to predict activating mutations cannot rely on evolutionary and conservation information or on the role of residues in catalysis or ATP binding. Instead, they should be based on an analysis of the clustering of activating mutations like the one presented here. We have demonstrated the feasibility of this approach by analyzing 66 cancer-associated somatic mutations of unknown effect and identifying two in the *EPHA7* gene that can be functionally relevant and deserve further experimental validation. In addition to the clustering analysis, the new methods to identify activating mutations in PKs could incorporate the different frequencies of amino acids affected by this type of mutation (see Fig. 1D and E), taking into account their chemical properties. These new methods might eventually lead to the discovery of hitherto unknown gain-of-function mutations, which will expand our knowledge of the oncogenic process and constitute potential targets for the design of new drugs. Finally, further investigation will determine whether the same principles and models can be generalized to activating mutations of proto-oncogenes that do not codify PKs, such as *MYC* or *RAS*.

Acknowledgments

We thank Renée O'Brate and Kate Williams for editorial assistance.

Disclosure statement: The authors declare no conflict of interest.

References

Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H, Spevak W, Zhang C, Zhang Y, Habets G, Burton EA, Wong B, et al. 2010. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature 467:596–599.

Chan P, Ilangumaran S, La Rose J, Chakrabartty A, Rottapel R. 2003. Autoinhibition of the kit receptor tyrosine kinase by the cytosolic juxtamembrane region. Mol Cell Biol 23:3067–3078.

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

Chen H, Huang Z, Dutta K, Blais S, Neubert TA, Li X, Cowburn D, Traaseth NJ, Mohammadi M. 2013. Cracking the molecular origin of intrinsic tyrosine kinase activity through analysis of pathogenic gain-of-function mutations. Cell Rep 25:376–384.

Cover TM, Thomas JA. 1991. *Elements of information theory*. New York, NY: Wiley-Interscience.

Consortium TU. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40:D71–D75.

Croce CM. 2008. Oncogenes and cancer. N Engl J Med 358:502–511.

Dagher R, Cohen M, Williams G, Rothmann M, Gobburu J, Robbie G, Rahman A, Chen G, Staten A, Griebel D, Pazdur R. 2002. Approval summary: Imatinib Mesylate in the treatment of metastatic and/or unresectable malignant gastrointestinal stromal tumors. Clin Cancer Res 8:3034–3038.

Dixit A, Yi L, Gowthaman R, Torkamani A, Schork N, Verkhivker G. 2009. Sequence and structure signatures of cancer mutation hotspots in protein kinases. PLoS One 4(10):e7485–e7493.

Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell 61:759–767.

Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, et al. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in cancer. Nucleic Acids Res 39:D945–D950.

Fröhling S, Scholl C, Levine RL, Loriaux M, Boggon TJ, Bernard OA, Berger R, Döhner H, Döhner K, Ebert BL, Teckie S, Golub TR, et al. 2007. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. Cancer Cell 12:501–513.

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

Gonzalez-Perez A, Deu-Pons J, Lopez-Bigas N. 2012. Improving the prediction of the functional impact of cancer mutations by baseline tolerance transformation. Genome Med 4:89.

Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, et al. 2007. Patterns of somatic mutation in human cancer genomes. Nature 446:153–158.

Griffith J, Black J, Faerman C, Swenson L, Wynn M, Lu F, Lippke J, Saxena K. 2004. The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. Mol Cell 13:169–781.

Hanks S, Hunter T. 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J 9:576–596.

Hashimoto K, Rogozin IB, Panchenko AR. 2012. Oncogenic potential is related to activating effect of cancer single and double somatic mutations in receptor tyrosine kinases. Human Mut 33:1566–1575.

Hauschild A, Grob JJ, Demidov LV, Jouary T, Gutzmer R, Millward M, Rutkowski P, Blank CU, Miller WH Jr, Kaempgen E, Martín-Algarra S, Karaszewska B, et al. 2012. Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. Lancet 380:358–365.

Holmberg J, Clarke DL, Frisen J. 2000. Regulation of repulsion versus adhesion by different splice forms of an Eph receptor. Nature 408:203–206.

Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, et al. 2010. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 466: 869–873.

Karkkainen MJ, Ferrell RE, Lawrence EC, Kimak MA, Levinson KL, McTigue MA, Alitalo K, Finegold DN. 2000. Missense mutations interfere with VEGFR-3 signalling in primary lymphoedema. Nat Genet 25:153–159.

Lee W, Zhang Y, Mukhyala K, Lazarus R, Zhang Z. 2009. Bi-directional SIFT predicts a subset of activating mutations. PLoS One 4:e8311–e8318.

Lei M, Lu W, Meng W, Parrini MC, Eck MJ, Mayer BJ, Harrison SC. 2000. Structure of PAK1 in an autoinhibited conformation reveals a multistage activation switch. Cell 102:387–389.

Lei M, Robinson MA, Harrison SC. 2005. The active conformation of the PAK1 kinase domain. Structure 13:769–778.

Loriaux MM, Levine RL, Tyner JW, Fröhling S, Scholl C, Stoffregen EP, Wernig G, Erickson H, Eide CA, Berger R, Bernard OA, Griffin JD, et al. 2008. High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. Blood 111:4788–4796.

Manning G, Whyte D, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. Science 298:1912–1934.

Marino-Buslje C, Teppa E, Di Doménico T, Delfino JM, Nielsen M. 2010. Networks of high mutual information define the structural proximity of catalytic sites: implications for catalytic residue identification. PLoS Comput Biol 6:e1000978.

Oricchio E, Nanjangud G, Wolfe AL, Schatz JH, Mavrakis KJ, Jiang M, Liu X, Bruno J, Heguy A, Olshen AB, Socci ND, Teruya-Feldstein J, et al. 2011. The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. Cell 147:554–564.

Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, et al. 2004. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci USA 101:13306–13311.

Pasquale, E. 2010. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. Nat Rev Cancer 10:165–180.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera—a visualization system for exploratory research and analysis. J Comp Chem 25:1605–1612.

Pollock PM, Gartside MG, Dejeza LC, Powell MA, Mallon MA, Davies H, Mohammadi M, Futreal PA, Stratton MR, Trent JM, Goodfellow PJ. 2007. Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes. Oncogene 26: 7158–7162.

Porter C, Bartlett G, Thornton J. 2004. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic Acids Res 32:D129–D133.

Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, et al. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 475:101–105.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, et al. 2012. The Pfam protein families database. Nucleic Acids Res 40:D290–D301.

Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118.

Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, Majem M, Lopez-Vivanco G, Isla D, Provencio M, Insa A, Massuti B. 2009. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 361:958–967.

Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, Jumper J, Kuriyan J, Shaw DE. 2012. Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. Cell 149:860–870.

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

Siddiqui M, Scott L. 2007. Imatinib: a review of its use in the management of gastrointestinal stromal tumours. Drugs 67:805–820.

Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, Gaunt TR. 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Human Mut 34:57–65.

Taly JF, Magis C, Bussotti G, Chang JM, Di Tommaso P, Erb I, Espinosa-Carrasco J, Kemena C, Notredame C. 2011. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. Nat Protocol 6:1669–1682.

Torkamani A, Schork NJ. 2008a. Prediction of cancer driver mutations in protein kinases. Cancer Res 68:1675–1682.

Torkamani A, Kannan N, Taylor SS, Schork NJ. 2008b. Congenital disease SNPs target lineage specific structural elements in protein kinases. Proc Natl Acad Sci USA 105:9011–9016.

Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, et al. 2007. The genomic landscapes of human breast and colorectal cancers. Science 318:1108–1113.

Zhang X, Gureasko J, Shen K, Cole P, Kuriyan J. 2006. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 125: 1137–1149.

Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen SC, et al. 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 481:157–163.