SUPPLEMENTARY MATERIALS & METHODS

Cloning & site-directed mutagenesis

As described in [1]. Point mutations were introduced using site-directed mutagenesis with PfuUltra™ II Fusion HS DNA polymerase (Agilent Technologies) following the manufacturer's instructions.

FGFR3 kinase domain protein expression & purification

As described in [1].

Production of phosphorylated FGFR3 kinase domain proteins

Recombinant unphosphorylated FGFR3 kinase domain proteins at 2-10 mg/mL were incubated with 25 mM $MgCl_2$ and 10 mM ATP for 45 minutes at 22°C in order for autophosphorylation to occur. The kinase reaction was stopped with 50 mM EDTA and desalted on a HiLoad 16/60 Superdex 75 column (GE Healthcare) using an Akta Purifier (GE Healthcare) with Desalting Buffer (25 mM Tris.Cl pH 8.0, 10 mM NaCl, 1 mM TCEP). Peak fractions corresponding to FGFR3 were pooled and loaded on to a 1 mL Resource Q anion-exchange column (GE Healthcare) equilibrated in Q Buffer A (25 mM Tris.Cl pH 8.0, 1 mM TCEP). Differentially phosphorylated FGFR3 proteins were eluted with a very shallow gradient over 500 column volumes to 40% of Q Buffer B (25 mM Tris.Cl pH 8.0, 1 M NaCl, 1 mM TCEP). The resulting chromatogram showed five peaks, corresponding to the 0P, 1P, 2P, 3P and 4P phosphorylated forms of FGFR3. This was confirmed by native PAGE analysis of pooled fractions from each peak. The 4P peak fractions were pooled and concentrated to 1-2 mg/mL using Vivaspin concentration units (Vivaproducts). Protein was stored until required at -80°C after snap freezing in liquid nitrogen.

Crystallisation and crystallography

$FGFR1^{R675G}$

Crystals of $FGFR1^{R675G}$ KD were grown by sitting drop vapour diffusion method on MRC 2-well plates. The protein (at 10 mg/ml concentration) was equilibrated at a ratio of 1:1 protein to mother liquor (20 % PEG 400, 0.75 – 2.0 M ammonium sulphate, 0.1 M magnesium chloride, 0.1 M Hepes pH 7.5) and incubated at 16 °C. The crystals grew in about 5 – 7 weeks. X-ray diffraction data were recorded on beamline I04 at Diamond Light Source. The data were processed, integrated and scaled to a high resolution of 2.58 Å using XDS [2] and aimless [3] suite respectively (see Supplementary Table S2) in primitive orthorhombic space group. Initial phases of $FGFR1^{R675}$ KD were calculated using molecular replacement [4] method in Phaser software suite [5] after trimming of all the loop regions in apo FGFR1 KD (PDB: 4UWY [1]). There were a total of 5 molecules per asymmetric unit with a Matthews coefficient of 2.7 and solvent content of 54.6 %.

Model building and refinement were performed using Phenix suite [6]. A subset of 5 % of reflections were kept aside for Rfree calculation. The structure was refined to a final Rcryst/Rfree value of 0.1937/0.2545 and rmsd in bond lengths and angles were 0.003 Å and 0.752 Å respectively. There were a total of 94.5 and 5.5 % of residues in the favoured and additionally favoured regions of Ramachnadran plot. The refined structures were validated using validation tools in Phenix [6] and online server Molprobity [7].

FGFR1 complex with JNJ42756493

FGFR1 wild-type protein was produced as described in Norman, R. *et al* (2012) [8]. Protein was crystallised using the hanging-drop vapour diffusion method at 277K. Drops of equal volumes of protein and reservoir solutions were suspended over a reservoir of 18-20% (w/v) PEG8000, 200mM ammonium sulphate, 100mM PCTP pH 6.75 and 20% (v/v) ethylene glycol. For soaking experiments, a solution of 22%(w/v) PEG8000, 200mM ammonium sulphate, 100mM PCTP pH 6.75, 20% (v/v) ethylene glycol and 1mM AZ13793811 was used and soaks were left to incubate for 24 hours. All work was carried out at 277K. Crystals from soaking experiments could be flash frozen in a stream of nitrogen gas at 100K directly from the drop. Diffraction data were collected at 100 K at I04-1 beamline at the Diamond synchrotron (Oxford, UK) on a Pilatus 6M detector. The data was integrated using the program XDS [2] and scaled with Aimless [3] as implemented in autoPROC [9] to a final resolution of 1.69Å. The FGFR1-JNJ crystals belong to the space group C1 2 1 and contain two complexes per asymmetric unit. The structure was solved by molecular replacement using the program PHASER [5] and an in-house FGFR1 structure as a search model. Subsequent model building and refinement were conducted using COOT [10] and BUSTER [11]. Quality checks were carried out using validation tools in COOT and Molprobity [7], while the compound stereochemistry was checked against the Cambridge Structure Database (CSD) [12] using Mogul [13].

Supplementary Table S2 gives a summary of the key data collection and refinement statistics.

PyMOL software suite (www.pymol.org) was used to prepare all the figures.

In vitro kinase assays

Kinase assays were carried out *in vitro* using the ADP-Glo™ Kinase Assay (Promega). Kinase reactions were carried out in Kinase Reaction Buffer (50 mM HEPES.Cl pH 7.5, 20 mM NaCl, 20 mM $MgCl_2$, 100 μM Na_3VO_4 and 100 μM TCEP) in 15 μL triplicates. All reactions in which the final kinase concentration was below 0.5 μM also contained 2 mM $MnCl_2$ and 0.1 mg/mL BSA. A poly-Glu-Tyr ($PolyE_4Y_1$) peptide (Sigma) was used as a synthetic kinase substrate. All reaction components were serially diluted in Kinase Reaction Buffer. Reactions were started by the addition of either ATP or kinase, depending on the type of assay. After the required duration, kinase reactions were stopped by the addition of 15 μL ADP-Glo™ Reagent for 40 minutes, followed by the addition of 30 μL Kinase Detection Reagent (KDR) for 30-60 minutes (depending on the concentration of ATP in the kinase reactions). Luminescence was measured at 520 nm on a FLUOstar Optima microplate reader (BMG Labtech). Data were analysed using Prism software (GraphPad). For Ki determination assays the method outlined in [1] was followed.

Cell culture

NIH 3T3 cells were cultured at 37°C and 10% CO_2 in medium consisting of high-glucose Dulbecco's Modified Eagle's Medium (DMEM) (Sigma) supplemented with 10% Fetal Bovine Serum (FBS) (Biosera or Gibco) and 1% GlutaMAX™ (Life Technologies). Stable FGFR3b NIH 3T3 cell lines, generated by retroviral transduction, were cultured similarly with medium supplemented with hygromycin B (Invitrogen) at a maintenance dose of 100 μg/mL.

Phoenix™ cell lines were cultured at 37°C and 5% CO_2 in the same medium as above supplemented with hygromycin B at a maintenance dose of 300 μg/mL.

Production of stable NIH 3T3 cell lines

Full length FGFR3 constructs were generated in the pFB retroviral expression vector (Stratagene) modified to contain a hygromycin b resistance cassette [14]. Plasmids were prepared using an alkaline lysis large-scale plasmid preparation method. Two point five μg of purified plasmid was then mixed with TransIT-293 Transfection Reagent (Mirus) in the presence of serum-free medium for 20 minutes at room temperature. Complexes were transfected into Phoenix™ cell lines and maintained daily as described above. After 72 hours the retroviral supernatant was harvested and mixed with 8 μg/mL polybrene (Sigma). NIH 3T3 cells were incubated in the retroviral supernatant for 4 hours, and then maintained in medium until they were considered 'stable' (two passages post-transduction). Successfully transduced cells were propagated by selection with 200 μg/mL hygromycin B before freezing the stable cell lines. Upon re-seeding from stocks, cell lines were subjected to selection with medium containing 200 μg/mL hygromycin b for a further three passages before they were fully selected. Cells were then maintained in medium containing maintenance dose hygromycin b at 100 μg/mL.

Protein extractions & immunodetection

Cultured cells were starved for 2 hours at 37°C and 10% CO_2 in FBS-free medium and then lysed *in situ* in Lysis Buffer (20 mM Tris.Cl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% [v/v] Triton, 20 mM sodium pyrophosphate, 25 mM sodium fluoride, 1 mM β-glycerophosphate, 3 mM Na_3VO_4, 1 μg/ml leupeptin) containing cOmplete™ protease inhibitor cocktail (Roche). Lysates were clarified by centrifugation at 16,100 x *g* for 15 min at 4°C. Total protein concentration was determined using the Pierce™ BCA Protein Assay Kit (Thermo Scientific) and 15 μg protein was run on 10% SDS-PAGE gels (produced in-house). Proteins were transferred to PVDF or nitrocellulose membranes using the Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) and blocked in 5% [w/v] NFDM or 3% [w/v] BSA respectively. Membranes were incubated with primary antibodies for 14 hours at 4°C. Primary antibodies used were anti-β-actin (1:2000 in 3% [w/v] NFDM) (Abcam), anti-FGFR3 clone B9 (1:1000 in 3% [w/v] NFDM) (Santa Cruz Biotechnology), anti-ERK (1:1000 in 3%[w/v] NFDM) (Millipore) and anti-phosphoERK (1:2000 in 3 % [w/v] BSA) (Cell Signaling Technology). These were all used with HRP-conjugated secondary antibodies and ECL Prime (GE Healthcare) and imaged using the Odyssey Fc System (Li-Cor Biosciences) or Super RX Fuji medical x-ray film (Kodak). For immmunoprecipitation, protein samples (normalized for FGFR3 amount using a separate aliquot) were incubated with antibodies to FGFR3 extracellular domain (F3922, Sigma) for 16 hours at 4°C, followed by the addition of protein A sepharose beads (Amersham Biosciences) for 4 hours. The beads were washed four times in PBS, resuspended in 2xSDS loading and subjected to SDSPAGE and Western blotting.

Anchorage independent growth assays

Stable NIH 3T3 wildtype and mutant cell lines were seeded in triplicate at 5 x 10^3 cells per well of a 6-well plates in medium containing 0.4% agarose (Gibco), on a base of medium containing 0.8% agarose. Cells were fed weekly with medium containing 0.4% agarose. After two

weeks, cells were stained for 24 hours with 0.3% [w/v] p-iodonitrotetrazolium violet (Sigma) and each well was imaged using a Hamamatsu ORCA-285 camera coupled to a Nikon SMZ1000 microscope. Thresholding parameters for area and perimeter of colonies with a mean diameter greater than 100 μm were set on ImageJ. Colonies were counted only if they satisfied both the area and perimeter criteria above the threshold values.

Computational methods

Sequences

Sequences were downloaded from UniProt in FASTA format for FGFR1 (P11362-1), FGFR2 IIIc (P21802-1), FGFR3 IIIc (P22607-1) & FGFR4 (P22455-1). Alignments used default settings of ClustalOmega [15] via Jalview 2.8.2 [16]. FGFR3 canonical amino acid numbering refers to FGFR3c (UniProt: P22607-1).

Structures & preparation

For structural studies we used "F3-Active", the active-like conformation of an FGFR3 K650E mutant (PDB ID: 4k33 [17]) and "F3-Apo" a model of FGFR3 inactive/*apo* kinase generated with Modeller [18] based on FGFR1 kinase-domain structure in an inactive/*apo* conformation (PDB ID: 4uwy [1]). Model quality was checked using MolProbity [7], ProSA-web [19], ModEval [20] and DOPE [21]. In addition, the model has a very high SSAP score [22] (>90) when compared to F3-Active, with a corresponding RMSD of 2.45Å.

For FOLDX analysis using F3-Active, three mutations found in PDB structure 4k33 were back-mutated, changing S482 to C, S582 to C and E650 to K to match the canonical FGFR3c sequence. UCSF Chimera "Dock Prep" tool [23] was used on both F3-Active and F3-Apo to remove ions, ligands and solvent and replace missing side-chains, add hydrogen atoms and AMBERff14SB [24]-calculated charges, followed by energy minimisation.

Mutations

The majority of somatic mis-sense mutations were obtained using COSMIC v71 (cancer.sanger.ac.uk) [25]. To ensure a more comprehensive assessment of rare mutations, we also included data from TCGA (TCGA Research Network: http://cancergenome.nih.gov/), ICGC (International Cancer Genome Consortium https://icgc.org/) and BioMuta [26]. Downloaded COSMIC mutations were reported across all cancer types for entries annotated as 'Substitution - Mis-sense' and having a specific amino acid mutation call. Simple Somatic Mutations were downloaded from ICGC (v17) and BioMuta (1st December 2014) for FGFRs1, 2, 3 & 4. TCGA data in MAF (Mutation Annotation File) format was obtained for 19 cancer types (ACC, BLCA, BRCA, CESC, COAD, COADREAD, GBM, GBMLGG, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, OV & PAAD) using Broad Institute GDAC Firehose [27] on 6th December 2014. Annotations of variants at the amino acid level were obtained for each cancer type using Ensembl Variant Effect Predictor (VEP) [28]. Care was taken to avoid duplicating mutation counts, for example, by not counting ICGC mutations that originated from TCGA twice. For FGFRs1-4 we combined all data sources, aggregating mutation counts both by the total number at each amino acid position and by the number of specific mutations found (i.e. K650E, K650M, etc).

Germline mutations were downloaded on 17th June 2015 from both UniProt Human Variants ('http://www.uniprot.org/docs/humsavar.txt) and Clinvar (http://www.ncbi.nlm.nih.gov/clinvar/). In addition, we referred to a review [29] to determine which amino acid positions on FGFR3 had reported Dysplasia mutations coincident with cancer mutations in either FGFR1, 2 or 3.

FOLDX stability calculations

We used FOLDX version 3b6 [30] to assess effects of mutations on protein stability of all possible single nucleotide variant, mis-sense mutations in FGFR3 kinase domain. The FOLDX "Position Scan" function was used in script mode for F3-Active and F3-Apo, using five runs per mutation and enabling the "RepairPDB" option. Multiple runs allow averaging of energy differences arising from side-chain rotamer conformations. For each mutation on each structure we obtained the difference in free energy of protein unfolding caused by the change to mutant amino acid in *kcalmol*$^{-1}$ as:

$$\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$$

We compared differential effects of mutations between F3-Active and F3-Apo using:

$$\Delta\Delta\Delta G = \Delta\Delta G_{Active} - \Delta\Delta G_{Apo}$$

Here, negative values of $\Delta\Delta\Delta G$ indicate stabilisation of F3-Active with respect to F3-Apo, indicating a shift in equilibrium that favours activated kinase. We categorised $\Delta\Delta\Delta G$ for each mutant using percentiles calculated from the total distribution of values for all possible mutants (5%: Stabilising - very high, 10%: Stabilising- high, 20%: Stabilising - medium and 30%: Stabilising - low; destabilising mutants were categorised equivalently). The middle 40% of $\Delta\Delta\Delta G$ values were not annotated as either stabilising or destabilising.

Clustering of mutations on structure

In any given tumour sample most mutations are extremely rare, thus lack of reported mutations of specific residues may reflect insufficient coverage by

cancer genomics projects. In identifying potential novel driver mutations, residues that are part of a cluster of observed mutations on the protein structure are reasonable candidates. We used an in-house method "MutClust" to test whether the density of all observed mutations within a spherical volume centred on each amino acid in turn is greater than would be expected by chance, using the F3-Active structure.

Briefly, all known mutations across FGFRs1-4 were mapped to F3-Active. For each amino acid, we tallied mutation-harbouring residues within spheres of radii 4, 5, 6 & 7Å. The observed mutation counts in each sphere may reflect both random variation in underlying mutation rates and different local protein structure densities (for example, on average we would expect more neighbouring mutations of residues in the densely packed protein core compared to a loop region). To account for these, we developed a random model based on a permutation test. We generated random mutant structures by assigning each amino acid a discrete value as either mutated (1) or not (0) using a probability estimated from the data as:

$$P(r_i = 1) = \frac{1}{len(s)} \sum_{j=1}^{len(s)} s_j \mid r_i, s_j \in \{0,1\} \simeq 0.33$$

Where (for a given random mutant) r_i is the ith amino acid of randomised sequence r and s is the F3-Active sequence, where $s_j = 1$ if a mutation is observed at amino acid j in any of the databases outlined for any of FGFRs1-4, or 0 otherwise.

We used 1000 permutation tests for each sphere radius (4, 5, 6 & 7Å) allowing calculation of means & standard deviations of the number of neighbouring random mutations found for each amino acid at each sphere radius. Hotspots were defined as residues with more observed mutated neighbouring residues than by chance using 99% significance. Residues were assigned to one of two identified hotspot clusters by identifying the least overlapping sets of residues within the significant cluster spheres (see Supplementary Figure S2).

Pathogenicity predictions

Condel [31] (**con**sensus**del**eterious) assessed multiple methods for predicting deleteriousness of non-synonymous SNVs: SIFT [32], Polyphen2 [33], MutationAssessor [34] and FATHMM [35]. Condel v2 data (downloaded via FannsDB at http://bg.upf.edu/fannsdb/ on 28th Nov 2014) uses a consensus of MutationAssessor and FATHMM. Condel scores were converted to z–scores using their values over the F3-Active kinase-domain. We defined 'highly deleterious' mutations as those having Condel $z > Q_3$ (Q=Quartile).

F3-Active mutations occurring in cancer or germline were submitted to SAAPdb [36]. SAAP predicts structural effects (e.g. disruption of H-bonding, salt-bridges, clashes, burying charge) and reports highly conserved sites and known functional sites from UniProt (e.g. binding sites, active sites).

Mutations submitted for SAAPdb analysis were also submitted to SAAPpred [37], a machine learning method that classifies SAAP features, outputting a prediction of neutral (SNP) or pathogenic (PD) and a confidence score.

Technical implementation

We used SQL scripts to import, combine and filter data, using a custom-built database. Further processing of data for structural clustering and FOLDX calculations used R scripts, with final presentation using R [38] & Microsoft Excel.

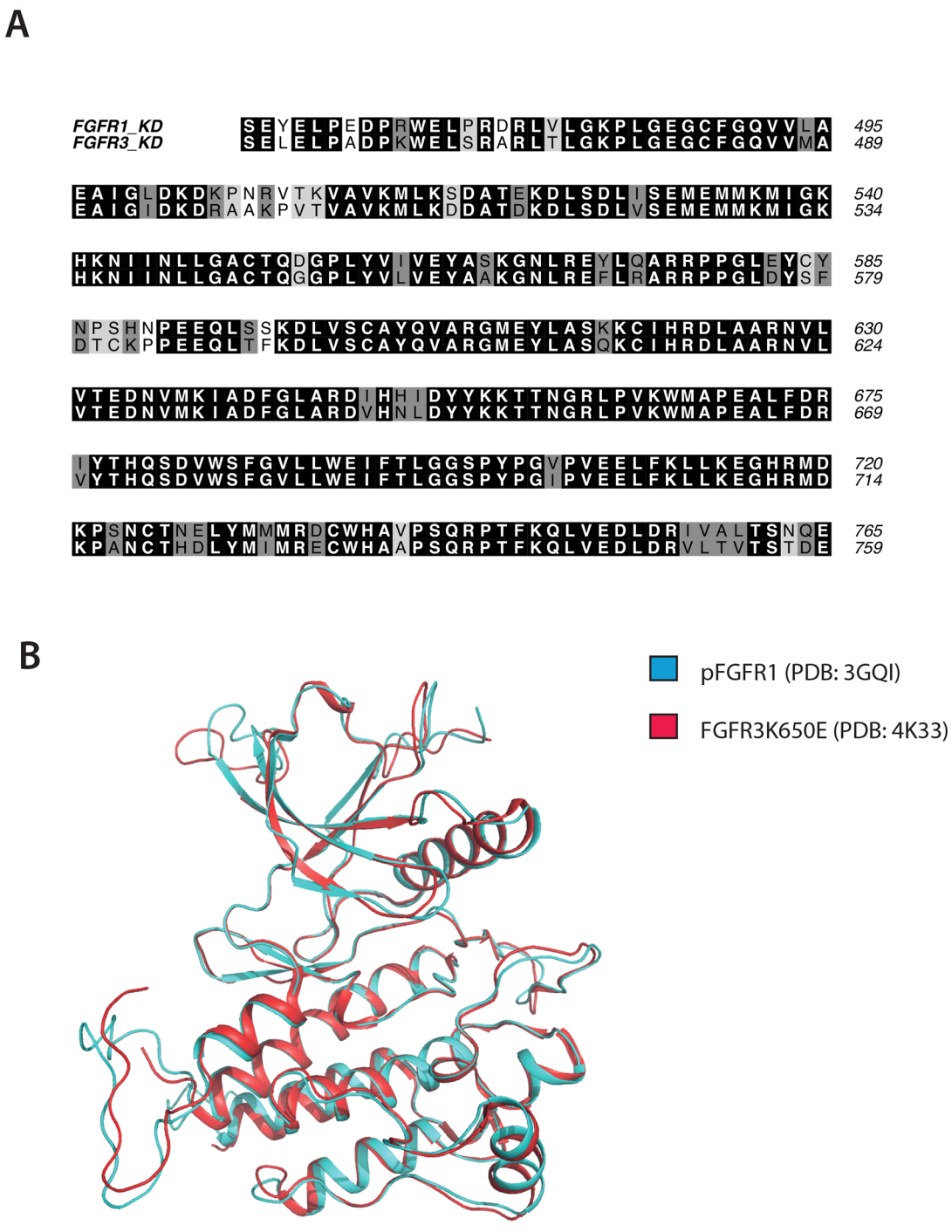
REFERENCES

1. Bunney TD, Wan S, Thiyagarajan N, Sutto L, Williams SV, Ashford P, Koss H, Knowles MA, Gervasio FL, Coveney PV and Katan M. The Effect of Mutations on Drug Sensitivity and Kinase Activity of Fibroblast Growth Factor Receptors: A Combined Experimental and Theoretical Study. EBioMedicine. 2015; 2:194-204.
2. Kabsch W. Processing of X-ray snapshots from crystals in random orientations. Acta crystallographica Section D, Biological crystallography. 2014; 70:2204-2216.
3. Evans PR and Murshudov GN. How good are my data and what is the resolution? Acta crystallographica Section D, Biological crystallography. 2013; 69:1204-1214.
4. Rossmann MG. The molecular replacement method. Acta crystallographica Section A, Foundations of crystallography. 1990; 46:73-82.
5. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ. Phaser crystallographic software. Journal of applied crystallography. 2007; 40:658-674.
6. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography. 2010; 66:213-221.
7. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica Section D, Biological crystallography. 2010; 66:12-21.

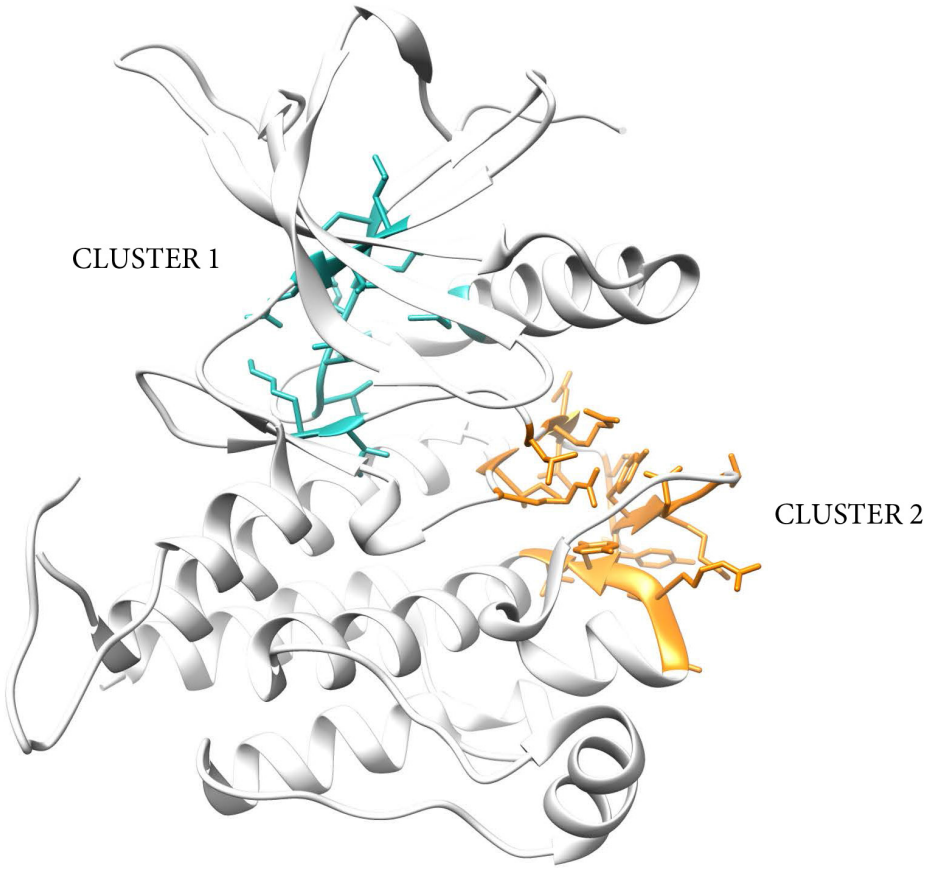
8. Norman RA, Schott AK, Andrews DM, Breed J, Foote KM, Garner AP, Ogg D, Orme JP, Pink JH, Roberts K, Rudge DA, Thomas AP and Leach AG. Protein-ligand crystal structures can guide the design of selective inhibitors of the FGFR tyrosine kinase. Journal of medicinal chemistry. 2012; 55:5003-5012.
9. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T and Bricogne G. Data processing and analysis with the autoPROC toolbox. Acta crystallographica Section D, Biological crystallography. 2011; 67:293-302.
10. Emsley P and Cowtan K. Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography. 2004; 60:2126-2132.
11. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek P, Roversi P, Sharff A, Smart O, Vonrhein C and Womack T. (2010). BUSTER version 2.11.1. Global Phasing Ltd.).
12. Allen FH. The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta crystallographica Section B, Structural science. 2002; 58:380-388.
13. Bruno IJ, Cole JC, Kessler M, Luo J, Motherwell WD, Purkis LH, Smith BR, Taylor R, Cooper RI, Harris SE and Orpen AG. Retrieval of crystallographically-derived molecular geometry information. Journal of chemical information and computer sciences. 2004; 44:2133-2144.
14. Tomlinson DC, L'Hote CG, Kennedy W, Pitt E and Knowles MA. Alternative splicing of fibroblast growth factor receptor 3 produces a secreted isoform that inhibits fibroblast growth factor-induced proliferation and is repressed in urothelial carcinoma cell lines. Cancer research. 2005; 65:10441-10449.
15. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD and Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011; 7:539.
16. Waterhouse AM, Procter JB, Martin DM, Clamp M and Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009; 25:1189-1191.
17. Huang Z, Chen H, Blais S, Neubert TA, Li X and Mohammadi M. Structural mimicry of a-loop tyrosine phosphorylation by a pathogenic FGF receptor 3 mutation. Structure. 2013; 21:1889-1896.
18. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U and Sali A. Comparative protein structure modeling using MODELLER. Current protocols in protein science / editorial board, John E Coligan [et al]. 2007; Chapter 2:Unit 2 9.
19. Wiederstein M and Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007; 35:W407-410.
20. Sali A and Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology. 1993; 234:779-815.
21. Shen MY and Sali A. Statistical potential for assessment and prediction of protein structures. Protein science : a publication of the Protein Society. 2006; 15:2507-2524.
22. Orengo CA and Taylor WR. SSAP: sequential structure alignment program for protein structure comparison. Methods in enzymology. 1996; 266:617-635.
23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605-1612.
24. Case DA, Darden T, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM and al. e. AMBER 12. University of California, San Francisco. 2012; 1:3.
25. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015; 43:D805-811.
26. Wu TJ, Shamsaddini A, Pan Y, Smith K, Crichton DJ, Simonyan V and Mazumder R. A framework for organizing cancer-related variations from existing databases, publications and NGS data using a High-performance Integrated Virtual Environment (HIVE). Database : the journal of biological databases and curation. 2014; 2014:bau022.
27. (2015). Analysis-ready standardized TCGA data from Broad GDAC Firehose stddata 2014 12 06 run. In: Center BIoMaHBITGDA, ed.
28. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P and Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010; 26:2069-2070.
29. Wilkie AO. Bad bones, absent smell, selfish testes: the pleiotropic consequences of human FGF receptor mutations. Cytokine & growth factor reviews. 2005; 16:187-203.
30. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F and Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005; 33:W382-388.
31. Gonzalez-Perez A and Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. American journal of human genetics. 2011; 88:440-449.

32. Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009; 4:1073-1081.
33. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS and Sunyaev SR. A method and server for predicting damaging missense mutations. Nature methods. 2010; 7:248-249.
34. Reva B, Antipin Y and Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011; 39:e118.
35. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN and Gaunt TR. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Human mutation. 2013; 34:57-65.
36. Hurst JM, McMillan LE, Porter CT, Allen J, Fakorede A and Martin AC. The SAAPdb web resource: a large-scale structural analysis of mutant proteins. Human mutation. 2009; 30:616-624.
37. Al-Numair NS and Martin AC. The SAAP pipeline and database: tools to analyze the impact and predict the pathogenicity of mutations. BMC genomics. 2013; 14:S4.
38. Team RC. (2014). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).

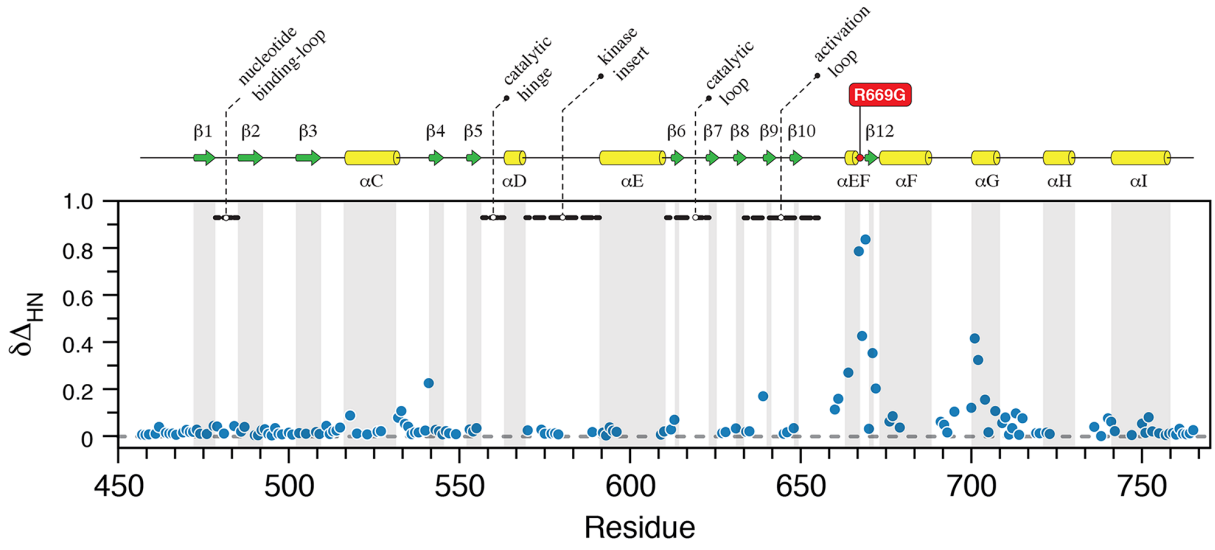
SUPPLEMENTARY FIGURES AND TABLES



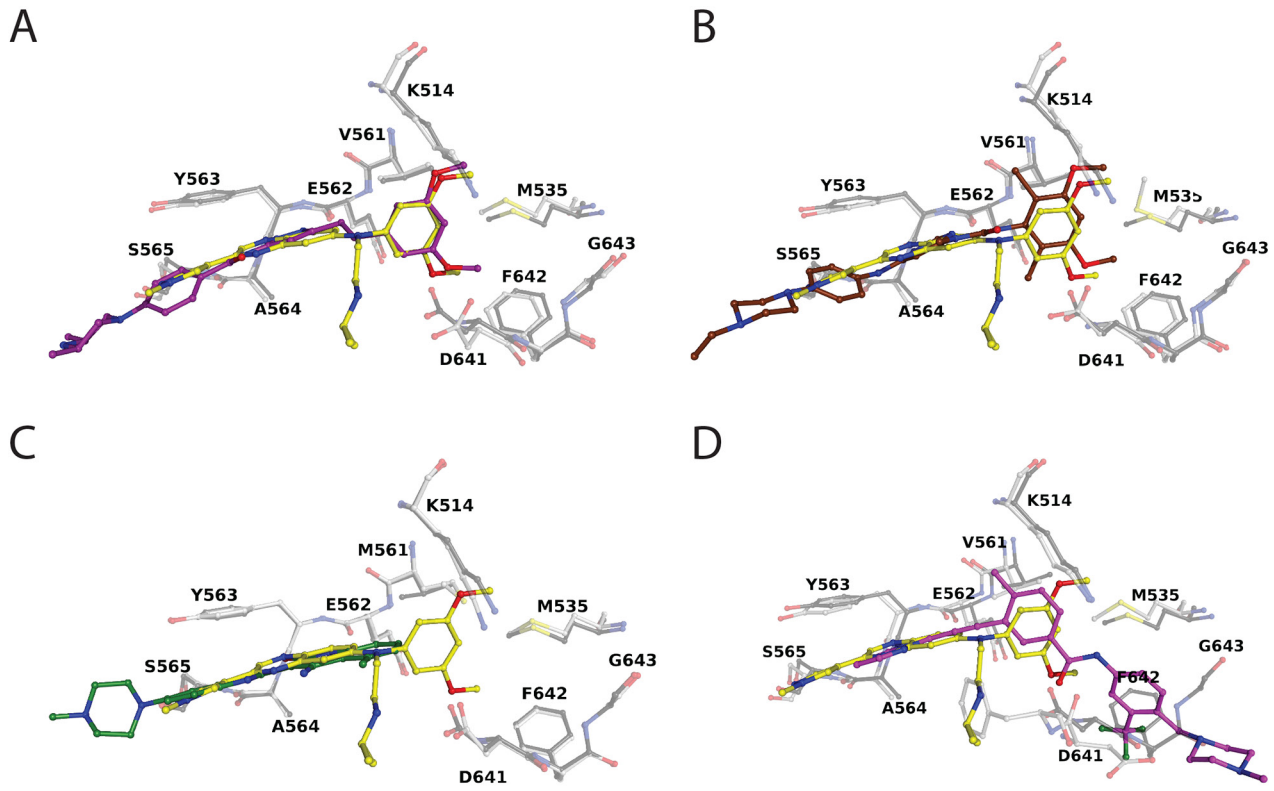
Supplementary Figure S1: Comparison of FGFR1 and FGFR3 kinase domains. A. Sequence alignment of kinase domains of FGFR1 and FGFR3. Identical residues are highlighted in black background while amino acids with similar properties (conservative substitutions) are highlighted in dark and light grey. Highlighted in dark grey are those amino acids that share similar charge. The proteins share 83% sequence identity and 92% sequence similarity. **B.** Cartoon representation and overlay of kinase domain of phosphorylated FGFR1 (PDB:3GQI) (blue) and activating mutant of FGFR3 (PDB: 4K33) (red). The secondary structure elements and overall structures are highly conserved between these two proteins.



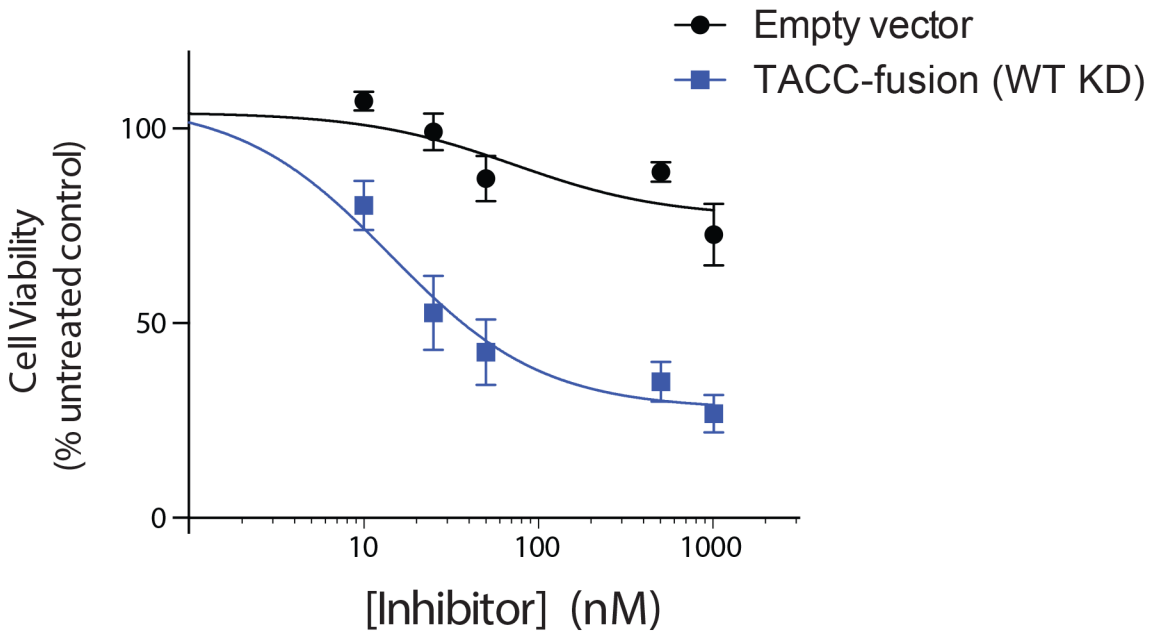
Supplementary Figure S2: "MutClust" mutation clusters on FGFR3. Two clusters of residues obtained from the MutClust algorithm indicate regions of high mutation density around the molecular brake (in cyan) and A-loop (in orange). Clusters are shown on PDB ID:4k33.



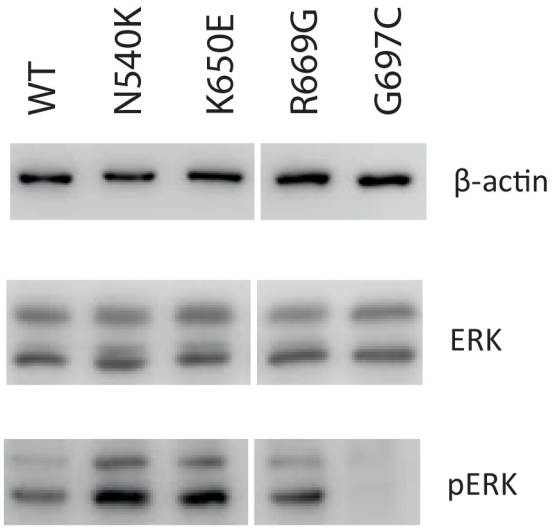
Supplementary Figure S3: HSQC Chemical shift perturbations in a comparison of the kinase domains of FGFR3^{WT} and FGFR3^{R669G}. Corresponding FGFR3 secondary structure elements and noteworthy functional elements are highlighted above the chemical shift plot as in main Figure 1D.



Supplementary Figure S4: Comparison of JNJ42756493 binding to FGFR1 with other inhibitors. A. Superposition of JNJ42756493 (in yellow) with AZD4547 (PDB: 4V05) (in purple). Nearby contact residues are shown as ball-and-stick model for JNJ42756493 (dark grey) and AZD4547 (light grey). **B.** Superposition of JNJ42756493 (in yellow) with BGJ-398 (PDB: 3TT0) (in chocolate). Nearby contact residues are shown as ball-and-stick model for JNJ42756493 (dark grey) and BGJ-398 (light grey). **C.** Superposition of JNJ42756493 (in yellow) with TKI258 (PDB: 5AM7) (in forest). Nearby contact residues are shown as ball-and-stick model for JNJ42756493 (dark grey) and TKI258 (light grey). **D.** Superposition of JNJ42756493 (in yellow) with AP24534 (PDB: 4V01) (in purple). Nearby contact residues are shown as ball-and-stick model for JNJ42756493 (dark grey) and AP24534 (light grey).



Supplementary Figure S5: Efficacy of AZD4547 in NIH3T3 stable cell lines. The efficacy towards NIH3T3 cell line expressing FGFR3 with the WT KD (FGFR3TACC3-fusion) was compared to NIH3T3 cell line stably transfected with the corresponding empty vector. The inhibitor concentrations used were 0, 10, 25, 50, 500 and 1000 nM.



Supplementary Figure S6: Levels of ERK expression and phosphorylation in NIH3T3 stable cell lines. Cell lysates from NIH3T3 cells expressing FGFR3 WT, N540K, K650E, R669G and G697C variants were subjected to Western blotting using indicated antibodies. The white separation between the lanes shows that an intervening lane from the same blot was not included.

Supplementary Table S1a: Computational analysis of FGFR KD mutations.

See Supplementary File 1

Supplementary Table S1b: Summary of bioinformatics predictions of SNV effects.

See Supplementary File 2

Supplementary Table S1c: Summary of Overlap between Experimental and Computational Predictions.

See Supplementary File 3

Supplementary Table S2: X-ray data collection and refinement statistics.

See Supplementary File 4

Supplementary Table S3: Measurements of Ki for selected FGFR3 variants and inhibitors.

See Supplementary File 5

Supplemental Table S1a: Computational Analysis of FGFR3 Kinase Domain Mutations.

ALIGNMENT	ANNOTATION	ALL MUTATIONS AT THIS AMINO ACID POSITION			SPECIFIC FGFR3 MUTATIONS			FOLDX STABILITY				CONDEL	SAAP			ALIGNMENT	
FGFR3_AA		FGFR1-4 total cancer mutations	Skeletal Dysplasia	3D cluster (MutClust)	WT	MUT	FGFR3 mutations	Apo ΔΔG (kcal/mol)	FGFR3 (4k33/WT) ΔΔG (kcal/mol)	ΔΔΔG = ΔΔG(FGFR3/WT) - ΔΔG(apo) (kcal/mol)	ΔΔΔG class	z-score	Effect	Pred	Pred Conf	FGFR3_AA	PANEL
466		7			GLU	LYS	2	1.3569	1.5962	0.2393		1.05	No structural effects identified	SNP	0.7	466	MUT_PANEL
469		4			ARG	GLN	1	0.8761	3.2061	2.33	Destabilising - low	1.12	No structural effects identified	SNP	0.76	469	
490		2			GLU	GLY	2	1.4404	1.5994	0.159		-0.25	No structural effects identified	SNP	0.73	490	
500		1			ALA	THR	1	0.1759	-0.2716	-0.4475	Stabilising - low	-0.66	No structural effects identified	SNP	0.85	500	MUT_PANEL
505		1			VAL	ILE	1	1.1085	-0.2999	-1.4084	Stabilising - high	0.63	Conserved site	SNP	0.59	505	
507		3			VAL	MET	2	0.5384	5.6024	5.064	Destabilising - low	0.83	Clash|Conserved site	PD	0.8	507	
538	Molecular brake	4	Dysplasia & cancer mutation		ILE	PHE	1	11.33951	3.8555	-7.48401	Stabilising - very high	0.63	No structural effects identified	SNP	0.64	538	MUT_PANEL
					ILE	VAL	0	0.6259	1.0697	0.4438	Destabilising - low	-1.01	No structural effects identified	SNP	0	538	MUT_PANEL
540	Molecular brake	61	Dysplasia & cancer mutation		ASN	LYS	0	2.1699	-0.3467	-2.5166	Stabilising - very high	-1.36	HBonds|Conserved site	0	0	540	MUT_PANEL
					ASN	SER	1	1.8546	1.5002	-0.3544	Stabilising - low	-1.09	Conserved site	SNP	0.78	540	MUT_PANEL
555	Gatekeeper	2			VAL	MET	1	-0.3313	-0.7755	-0.4442	Stabilising - low	-1.83	Conserved site	SNP	0.07	555	MUT_PANEL
569		1			ALA	VAL	1	1.1687	1.3753	0.2066		-0.32	No structural effects identified	SNP	0.87	569	
572		1			PRO	ALA	0	1.9	1.1313	-0.7687	Stabilising - medium	-0.57	No structural effects identified	SNP	0	572	MUT_PANEL
576		2			ASP	ASN	1	0.1427	-0.1713	-0.314		-0.32	No structural effects identified	SNP	0.82	576	
582		4			CYS	PHE	2	1.6788	-0.1992	-1.878	Stabilising - high	-0.61	Surface Phobic	SNP	0.76	582	MUT_PANEL
603		6			ARG	GLN	3	0.8708	0.6045	-0.2663		-1.25	Conserved site	SNP	0.78	603	
608		2			LEU	MET	1	0.7697	0.0215	-0.7482	Stabilising - medium	0.75	Conserved site		*0.04*	608	
614	Catalytic loop	1			ILE	ASN	1	2.463	3.82	1.357	Destabilising - low	1.90	Core Philic|Conserved site	PD	0.6	614	
616	Catalytic loop,HRD motif	4	Dysplasia & cancer mutation		ARG	GLY	1	0.6658	-0.4531	-1.1189	Stabilising - medium	0.58	Buried Charge|HBonds|Conserved site	PD	0.78	616	
617	Catalytic loop,HRD motif	1			ASP	GLY	1	2.5016	-0.2262	-2.7278	Stabilising - very high	2.32	Buried Charge|HBonds|SProtFT	PD	0.65	617	MUT_PANEL
621	Catalytic loop	3	Dysplasia & cancer mutation		ARG	HIS	2	-0.414	1.4743	1.8883	Destabilising - low	0.74	HBonds|Conserved site	SNP	0.02	621	
627		6			GLU	ASP	1	0.4188	0.4545	0.0357		-0.57	No structural effects identified	SNP	0.86	627	MUT_PANEL
					GLU	GLY	1	0.7871	0.9393	0.1522		-0.32	No structural effects identified	SNP	0.72	627	
					GLU	LYS	2	0.143	0.3906	0.2476		-0.77	No structural effects identified	SNP	0.79	627	
					GLU	VAL	1	0.6774	0.871	0.1936		0.45	Surface Phobic	SNP	0.73	627	
630		2			VAL	ALA	1	1.3942	1.4217	0.0275		-0.38	Conserved site	SNP	0.68	630	
					VAL	MET	1	0.2983	0.5522	0.2539		-0.29	Conserved site	SNP	0.7	630	MUT_PANEL
636	Activation loop,DFG motif,Regulatory spine	1			PHE	LEU	1	1.5435	2.679	1.1355	Destabilising - low	0.69	Conserved site	SNP	0.61	636	
637	Activation loop,DFG motif	2			GLY	TRP	2	14.60959	-1.777	-16.38659	Stabilising - very high	2.34	Conserved site	SNP	0.67	637	MUT_PANEL
640	Activation loop	3			ARG	TRP	2	-0.1865	2.9882	3.1747	Destabilising - low	2.48	Conserved site|Surface Phobic	PD	0.55	640	
641	Activation loop	5			ASP	ASN	2	0.264	0.3708	0.1068		-0.33	No structural effects identified	SNP	0.75	641	MUT_PANEL
					ASP	GLY	1	0.6006	1.6588	1.0582	Destabilising - low	-0.28	HBonds	SNP	0.55	641	MUT_PANEL
643	Activation loop	4			HIS	ARG	3	-0.2915	-1.3046	-1.0131	Stabilising - medium	-1.07	No structural effects identified	SNP	0.82	643	
					HIS	ASP	1	-0.2355	0.5107	0.7462	Destabilising - low	-1.78	No structural effects identified	SNP	0.81	643	MUT_PANEL
646	Activation loop	7			ASP	ASN	1	0.8063	0.3807	-0.4256	Stabilising - low	-0.91	Conserved site	SNP	0.82	646	
					ASP	GLY	1	2.2881	0.5843	-1.7038	Stabilising - high	-1.00	Conserved site	SNP	0.79	646	
					ASP	TYR	1	1.3938	-0.0788	-1.4726	Stabilising - high	-0.28	Conserved site	SNP	0.81	646	MUT_PANEL
647	Activation loop,Phosphorylated tyrosines	0			TYR	CYS	0	2.0028	2.2561	0.2533		0.13	Conserved site	0	0	647	MUT_PANEL
650	Activation loop	210	Dysplasia & cancer mutation		LYS	ASN	4	0.2716	-2.504	-2.7756	Stabilising - very high	-0.28	Buried Charge	SNP	0.05	650	MUT_PANEL
					LYS	GLN	6	0.3504	-2.3932	-2.7436	Stabilising - very high	-0.65	Buried Charge	SNP	0.01	650	
					LYS	GLU	83	-0.0104	-4.2792	-4.2688	Stabilising - very high	-0.38	No structural effects identified	SNP	0.14	650	MUT_PANEL
					LYS	MET	81	-0.1739	-4.0829	-3.909	Stabilising - very high	-0.39	Buried Charge|HBonds	PD	0.6	650	
					LYS	THR	4	-0.1266	-1.2372	-1.1106	Stabilising - medium	-0.40	Buried Charge		*0.02*	650	
653	Activation loop	2			ASN	HIS	1	0.7217	0.1799	-0.5418	Stabilising - low	-1.05	Conserved site	SNP	0.84	653	MUT_PANEL
					ASN	SER	1	0.2161	0.0724	-0.1437		-1.63	Conserved site	SNP	0.84	653	
669		5	Dysplasia & cancer mutation		ARG	GLN	1	-0.7112	-0.5591	0.1521		-1.10	No structural effects identified	SNP	0.79	669	MUT_PANEL
					ARG	GLY	0	-0.319	-0.164	0.155		-1.78	No structural effects identified	SNP	0	669	MUT_PANEL
677		4			VAL	ILE	1	-0.3179	-0.4012	-0.0833		-0.44	Conserved site	SNP	0.6	677	MUT_PANEL
679		2			SER	PHE	1	4.4703	2.753	-1.7173	Stabilising - high	2.10	HBonds|Conserved site	PD	0.72	679	
686		4			GLU	LYS	2	5.2432	4.6084	-0.6348	Stabilising - low	2.05	Buried Charge|HBonds|Conserved site	PD	0.73	686	
689		1			THR	MET	1	-1.5705	-0.47	1.1005	Destabilising - low	0.79	HBonds	SNP	0.58	689	
696		1			PRO	LEU	1	1.3335	1.5784	0.2449		0.44	Surface Phobic	SNP	0.83	696	
697		47			GLY	CYS	44	2.0335	3.5217	1.4882	Destabilising - low	0.53	Conserved site	SNP	0.65	697	MUT_PANEL
700		1			VAL	ALA	0	0.7684	0.7455	-0.0229		-1.45	No structural effects identified	SNP	0.84	700	
715		2			LYS	MET	2	0.6027	0.0987	-0.504	Stabilising - low	-0.29	HBonds|Conserved site	SNP	0.58	715	
716		4			PRO	HIS	3	37.7316	27.0001	-10.7315	Stabilising - very high	1.36	Buried Charge|Core Philic	PD	0.73	716	
725		1			MET	ILE	1	0.3884	0.3385	-0.0499		-0.64	No structural effects identified	SNP	0.8	725	

Summary of bioinformatics analyses for FGFR3 kinase-domain for all residues having at least 1 cancer mutation (plus panel mutant Y647C)

Molecular Brake residues (538 & 540) are part of a cluster of mutations (highlighted in light blue) that encompass the pharmacologically important gatekeeper (V555M). Mutations I538F, D540K/S & V555M all show evidence for shifting equilibrium to favour active kinase conformation according to FOLDX.

Observed oncogenic mutations in A-loop residues are predicted to preferentially stabilise active kinase conformation in 60% of cases (where they don't also involve pathogenic mutations according to Condel or SAAP) and show pronounced stabilisation effects for the K650 hotspot.

Activating mutations R669G/Q are in a mutation cluster with A-loop residues.

"Hotspots" are highlighted red where 10 or more cancer mutations are observed. Condel scores: red (>Q(uartile)4), mid-red (>Q3)

FOLDX ΔΔΔG classified for each mutant using percentiles calculated from the total distribution of values for all possible mutants (5%: Stabilising - very high, 10%: Stabilising - high, 20%: Stabilising - medium and 30%: Stabilising - low; destabilising mutants were categorised equivalently)

FOLDX equilibrium predictions (excluding pathogenic mutations E466, D617 & G637 and based on ~7 fold activating mutations in Figure 2A including I538V; using FOLDX very high and high only):

Sensitivity (True Positive Rate) = TP/(TP+FN) = 0.43; Positive Predictive Value = TP/(TP+FP) = 0.5

Specificity (True Negative Rate) = TN/(TN+FP) = 0.63; Negative Predictive Value = TN/(TN+FN) = 0.75

Supplemental Table S1b: Summary of Bioinformatics Predictions of SNV Effects.

Panel mutations			Bioinformatics methods indicating SNV effect						
			Sequence based			Seq & struct	Structure-based		
Position	Native	Mutant	Observed mutations *Cancer & Dysplasia*	*F1-4 high freq cancer*	Predicted *Condel*	consequence *SAAP/pred*	Clustering *MutClust*	Stabilising? *FOLDX*	Total effects
466	GLU	*LYS*			■				1
500	ALA	*THR*							0
538	ILE	*PHE*	■				■	■	**3**
538	ILE	*VAL*	■				■		2
540	ASN	*LYS*	■	■		■	■	■	5
540	ASN	*SER*	■			■	■		**3**
555	VAL	*MET*				■	■		**2**
572	PRO	*ALA*							0
582	CYS	*PHE*				■		■	2
617	ASP	*GLY*			■	■		■	**3**
627	GLU	*ASP*							0
630	VAL	*MET*				■			1
637	GLY	*TRP*			■	■		■	**3**
641	ASP	*ASN*					■		1
641	ASP	*GLY*				■	■		2
643	HIS	*ASP*							0
646	ASP	*TYR*				■	■	■	**3**
647	TYR	*CYS*				■			1
650	LYS	*ASN*	■			■	■	■	**4**
650	LYS	*GLU*	■	■			■	■	**4**
653	ASN	*HIS*				■			1
669	ARG	*GLN*	■				■		2
669	ARG	*GLY*	■				■		2
677	VAL	*ILE*				■			1
697	GLY	*CYS*		■		■			2

FOLDX stabilising: Very High and High only (top 10% most stabilising)

Supplemental Table S1c: Summary of Overlap between Experimental and Computational Predictions.

Panel mutations			Evidence of activating mutation			
			Experimental	Computational		Other effects
Position	Native	Mutant		Postive pred	Negative pred	
466	GLU	*LYS*	N	Condel/PD		
500	ALA	*THR*	N			
538	ILE	*PHE*	N		FOLDX + MutClust	
538	ILE	*VAL*	**Y**	MutClust		
540	ASN	*LYS*	**Y**	FOLDX + MutClust		SAAP:H-Bonds/conserved
540	ASN	*SER*	**Y**	MutClust		SAAP:Conserved
555	VAL	*MET*	**Part**	MutClust		SAAP:Conserved
572	PRO	*ALA*	N			
582	CYS	*PHE*	N		FOLDX	SAAP:surface-phobic
617	ASP	*GLY*	N	Condel/PD		SAAP/PD
627	GLU	*ASP*	N			
630	VAL	*MET*	N			SAAP:Conserved
637	GLY	*TRP*	N	Condel/PD		
641	ASP	*ASN*	**Part**	MutClust		
641	ASP	*GLY*	**Part**	MutClust		SAAP:H-Bonds
643	HIS	*ASP*	N			
646	ASP	*TYR*	N		FOLDX + MutClust	SAAP:Conserved
647	TYR	*CYS*	N			SAAP:Conserved
650	LYS	*ASN*	**Y**	FOLDX + MutClust		SAAP:buried charge
650	LYS	*GLU*	**Y**	FOLDX + MutClust		
653	ASN	*HIS*	N			SAAP:Conserved
669	ARG	*GLN*	**Y**	MutClust		
669	ARG	*GLY*	**Y**	MutClust		
677	VAL	*ILE*	N			SAAP:Conserved
697	GLY	*CYS*	N			SAAP:Conserved

Experimental: "Y"- activating according to Figure 2A (7 fold); "Part" ~5 fold
FOLDX stabilising: Very High and High only (top 10% most stabilising)
Condel/PD: Condel pathogenic deleterious

Supplemental Table S2: X-ray data collection and refinement statistics.

	FGFR1 R675G	**FGFR1 bound to JNJ42756493**
Space group	P 2_1 2_1 2_1	C2
Cell dimensions	a = 76.3 Å; b = 152.3 Å; c = 195.9 Å; $\alpha = \beta = \gamma = 90^o$	a = 209.1 Å; b = 57.9 Å; c = 65.3 Å; α = 90 ^{o}; β = 107.5 ^{o}; $\gamma = 90^o$
Resolution range (Å)	76.26 – 2.58	62.24 – 1.67
R_{symm}[a] (outer shell)	0.097 (0.603)	0.045 (1.027)
$I/\sigma I$ (outer shell)	10.4 (1.5)	20.3 (1.5)
Completeness (outer shell) %	98.2 (97.2)	96.7 (75.5)
Total number of reflections	288,249	577657
Number of unique reflections	70,919	89791
Redundancy (outer shell)	4.1 (3.1)	6.4 (4.8)
Wilson B-factor (Å^2)	56.0	29.90
R_{cryst}[b]/R_{free}[c]	0.194/0.255	0.184/0.209
Average B-factor (Å^2)		
Overall	52.8	37.32
Protein (chain A, B, C, D, E)	53.8, 45.0, 50.5, 45.1, 71.7	33.12, 40.10
Solvent	47.2	45.8
Ligands (SO_4^{2-}, PEG, Cl^-, $CH_3CO_2^-$)	88.1, 65.1, 60.3, 68.5	28.16, 46.60
RMS deviation		
Bond length (Å)	0.003	0.010
Bond angle (°)	0.752	0.97
Ramachandran plot statistics		
Favoured (%)	94.5	98.6
Less favoured (%)	5.5	1.4
PDB ID	**5FLF**	**5EW8**

[a] $R_{symm} = \Sigma_h \Sigma_i |I(h) - I_i(h)|/\Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ and $I(h)$ are the ith and the mean measurements of the intensity of reflection h, respectively.

[b] $R_{cryst} = \Sigma_h |F_o - F_c|/\Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively.

[c] R_{free} is equal to R_{cryst} for a randomly selected 5.0% subset of reflections not used in the refinement.

Supplemental Table S3. Measurements of Ki for selected FGFR3 variants and inhibitors

FGFR3	K_i of FGFR Inhibitor / nM				
	AZD4547	**BGJ-398**	**TKI258**	**JNJ42756493**	**AP24534**
WT	4.4 ± 2.4	69.3 ± 7.1	65.4 ± 14.6	1.87 ± 0.24	94.5 ± 12.2
V555M	82.5 ± 10.9	511 ± 40	12.9 ± 3.8	330 ± 81	184.5 ± 18.9
K650E	14.3 ± 1.3	14.0 ± 2.7	41.7 ± 2.5	3.93 ± 1.38	127 ± 15
R669G	26.3 ± 7.0	12.9 ± 3.5	341 ± 59	6.7 ± 1.4	69.9 ± 14.0
N540S	18.7 ± 2.2	45 ± 14	294 ± 42	35.2 ± 11.0	41.9 ± 4.0
N540K	99.7 ± 9.8	13.6 ± 1.7	124 ± 17	9.06 ± 0.8	75.9 ± 14.3
I538V	152.6 ± 1.7	236.8 ± 78.1	1604 ± 6.5	72.0 ± 7.8	674.2 ± 99.7