Supplemental Data

Table S1, related to Figure 1.
Genomic Coordinates of Mutations Identified (and Resulting Amino Acid Substitutions).

Amino Acid	Coding DNA Sequence	GRCh37/hg19 genome location	# primary tumor samples	Novel mutation?	In germline? (# tumors evaluated if not all)
p.R1060H	c.3179G>A	2:29446388	1	Y	Y
p.T1151M	c.3452C>T	2:29445273	1	N (COSMIC)	N
p.M1166R	c.3497T>G	2:29445228	1	N (COSMIC)	N
p.I1170N	c.3509T>A	2:29445216	1	Y	N
p.I1170S	c.3509T>G	2:29445216	1	N (COSMIC)	N
p.I1171N	c.3512T>A	2:29445213	2	N (COSMIC)	N
p.F1174I	c.3520T>A	2:29443697	2	N (COSMIC)	N (1)
p.F1174V	c.3520T>G	2:29443697	2	N (COSMIC)	N
p.F1174L	c.3520T>C	2:29443697	2	N (COSMIC)	N
p.F1174C	c.3521T>G	2:29443696	4	N (COSMIC)	N (3)
p.F1174S	c.3521T>C	2:29443696	1	N (COSMIC)	Not evaluated
p.F1174L	c.3522C>A	2:29443695	26	N (COSMIC)	N (19)
p.F1174L	c.3522C>G	2:29443695	1	N (COSMIC)	N
p.I1183T	c.3548T>C	2:29443669	1	Y	Y
p.L1196M	c.3586C>A	2:29443631	1	N (COSMIC)	Not evaluated
p.A1200V	c.3599C>T	2:29443618	1	N (COSMIC)	N
p.L1204F	c.3610C>T	2:29443607	1	Y	Y
p.R1231Q	c.3692G>A	2:29436901	1	N (*dbSNP*)	Y
p.L1240V	c.3718T>G	2:29436875	1	N (COSMIC)	Not evaluated
p.F1245I	c.3733T>A	2:29436860	2	N (COSMIC)	N (1)
p.F1245V	c.3733T>G	2:29436860	6	N (COSMIC)	N (4)
p.F1245C	c.3734T>G	2:29436859	6	N (COSMIC)	N
p.F1245L	c.3735C>A	2:29436858	1	N (COSMIC)	Not evaluated
p.I1250T	c.3749T>C	2:29432739	1	N (*dbSNP*)	Y
p.D1270G	c.3809A>G	2:29432679	1	Y	Not evaluated
p.R1275L	c.3824G>T	2:29432664	3	N (COSMIC)	N (3)
p.R1275Q	c.3824G>A	2:29432664	51	N (COSMIC)	Y
p.Y1278S	c.3833A>C	2:29432655	1	N (COSMIC)	N
p.G1286R	c.3856G>C	2:29430119	1	Y	Not evaluated
p.T1343I	c.4028C>T	2:29420453	1	Y	Not evaluated
p.D1349H	c.4045G>C	2:29420436	1	N (*dbSNP*)	Not evaluated

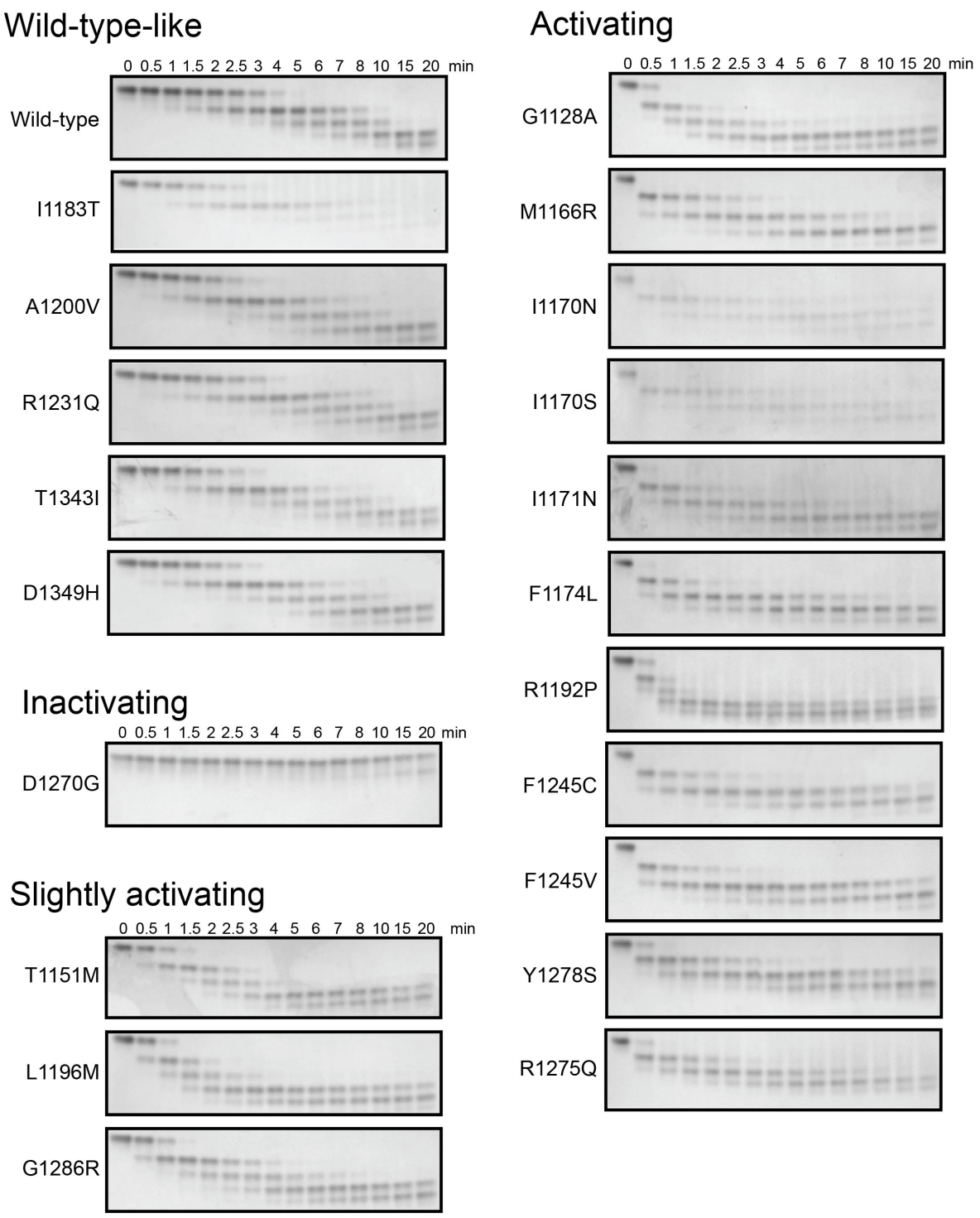

Figure S1, related to Figure 1.
Native Gel Analysis of ALK TKD Mutants.

Separation of differently autophosphorylated species of ALK TKD using native gel electrophoresis illustrates the progress of autophosphorylation when 10 μM ALK TKD (wild-type or mutated) is incubated with saturating ATP (2 mM) and 10 mM $MgCl_2$ at 37˚C. Results are representative of at least 2 independent experiments. L1204F, L1240V, and I1250T were not analyzed due to low expression levels in Sf9 cells.

Table S2, related to Figure 2.
***ALK* Mutation and Amplification Status by NB Risk Group and Age.**

	Mutation (n=126)[a] n (%)	No Mutation (n=1458)[a] n (%)	Amplification (n=24)[b] n (%)	No Amplification (n=1310)[b] n (%)
Low risk	38 (6)	586 (94)	0 (0)	454 (100)
Intermediate risk	23 (8)	263 (92)	0 (0)	241 (100)
High risk	65 (10)	595 (90)	24 (4)	604 (96)
Unknown risk group	0	14	0	11
Age <365 days	35 (6)	529 (94)	6 (1)	453 (99)
Age 1-5 years	76 (9)	755 (91)	18 (3)	689 (97)
Age 6-10 years	5 (5)	95 (95)	0 (0)	84 (100)
Age >10 years	10 (11)	79 (89)	0 (0)	84 (100)

[a]Mutation status was unknown for 12 patients.
[b]Amplification status was unknown for 262 patients.

Table S3, related to Table 2.
***ALK* Aberrations by Risk Group.**

Patient cohort	n (%)	5-year EFS[a] +/- std error	EFS[a] p value	5-year OS[b] +/- std error	OS[b] p value
LOW-RISK					
***ALK* mutation**					
At least one	38 (6%)	83 ± 7.8	0.66	97 ± 3.5	0.75
No mutation	586 (94%)	87 ± 1.8		97 ± 1.0	
Unknown	2				
Site of ALK mutation					
F1174	7 (18%)	100	0.27	100	0.79
F1245	6 (16%)	100		100	
R1275	18 (47%)	71 ± 14.4		94 ± 7.4	
Other mutation	7 (18%)	86 ± 16.2		100	
***ALK* Copy Number**					
Amplified	0 (0%)				
Gain	29 (6%)	61 ± 13.5	< 0.0001	79 ± 10	0.0001
No gain/ not amp	424 (93%)	86 ± 2.3		97 ± 1.2	
Loss	1 (1%)	1 event		100	
Unknown status	172				
INTERMEDIATE-RISK					
***ALK* mutation**					
At least one	23 (8%)	69 ± 14.5	0.02	96 ± 6.8	0.83
No mutation	263 (92%)	87 ± 2.8		94 ± 1.9	
Unknown	6				
Site of ALK mutation					
F1174	7 (30%)	86 ± 18.7	0.02	100	0.02
F1245	2 (9%)	2 events		50 ± 35.4	
R1275	13 (57%)	68 ± 19.2		100	
Other mutation	1 (4%)	100		100	
***ALK* Copy Number**					
Amplified	0 (0%)				
Gain	27 (11%)	81 ± 9.8	0.63	87 ± 8.6	0.14
No gain/ not amp	214 (89%)	85 ± 3.4		96 ± 2.0	
Loss	0 (0%)				
Unknown status	51				
HIGH-RISK					
***ALK* mutation**					
At least one	65 (10%)	29 ± 7.4	0.06	36 ± 8.7	0.06
No mutation	595 (90%)	41 ± 2.7		47 ± 2.7	
Unknown	4				
Site of ALK mutation					
F1174	24 (37%)	21 ± 13.4	0.34	31 ± 18.0	0.47
F1245	7 (11%)	17 ± 15.2		17 ± 15.2	
R1275	23 (35%)	33 ± 12.0		38 ± 13.4	
Other mutation	11 (17%)	46 ± 19.4		55 ± 21.2	
***ALK* Copy Number**					
Amplified	24 (4%)	24 ± 12.2	0.12	23 ± 11.7	0.03
Gain	138 (22%)	37 ± 5.1		46 ± 5.4	
No gain/ not amp	461 (73%)	42 ± 3.1		48 ± 3.1	
Loss	5 (1%)	50 ± 35.4		50 ± 35.4	
Unknown status	36				

[a]Event-free survival.
[b]Overall survival.

Table S4, related to Figure 3.
Kinetic Properties of Non-phosphorylated ALK TKD Variants.

Variant	Constitutively active?	Location	k_{cat} (min^{-1})[c]	$K_{M, ATP}$ (mM)[c]	$k_{cat}/K_{M, ATP}$ ($min^{-1}\cdot mM^{-1}$)	$k_{cat}/K_{M, peptide}$ ($min^{-1}\cdot mM^{-1}$)[c]
Wild-type[a]	No	--	9.32 ± 0.85	0.134 ± 0.007	69.6	3.41 ± 0.44
M1166R	Yes	αC/A-loop	127 ± 26	0.149 ± 0.004	852	29.0 ± 3.6
I1170N	Yes	αC/A-loop	200 ± 59	0.297 ± 0.015	673	65.1 ± 5.7
I1170S	Yes	αC/A-loop	200 ± 14	0.371 ± 0.013	539	71.5 ± 6.5
I1171N	Yes	αC/A-loop	188 ± 34	0.250 ± 0.013	752	46.0 ± 4.6
R1275Q[a]	Yes	αC/A-loop	119 ± 13	0.326 ± 0.033	365	46.8 ± 1.9
Y1278S	Yes	αC/A-loop	172 ± 22	0.152 ± 0.012	1132	35.8 ± 2.8
G1286R	Slightly	A-loop	16.4 ± 1.4	0.152 ± 0.006	108	3.63 ± 0.50
F1174L[a]	Yes	Phe core	365 ± 61	0.127 ± 0.011	2874	39.7 ± 2.8
F1245C	Yes	Phe core	329 ± 65	0.138 ± 0.001	2384	86.8 ± 6.8
F1245V	Yes	Phe core	341 ± 36	0.152 ± 0.009	2243	88.6 ± 10.8
G1128A	Slightly	P-loop	43.4 ± 13.8	0.152 ± 0.008	286	21.3 ± 3.0
T1151M	Slightly	N-lobe	53.4 ± 7.3	0.267 ± 0.018	200	12.4 ± 0.4
I1183T	Slightly	N-lobe	31.5 ± 5.6	0.158 ± 0.018	199	2.47 ± 1.19
R1192P	Yes	N-lobe	139 ± 33	0.192 ± 0.007	724	25.6 ± 1.5
L1196M	Slightly	Active site	45.0 ± 9.7	0.387 ± 0.033	116	13.1 ± 2.0
A1200V	No	Active site	11.1 ± 0.9	0.208 ± 0.009	53.4	1.76 ± 0.28
D1270G	No (inactivating)	Active site	0.923 ± 0.306	0.153 ± 0.019	6.03	0.267 ± 0.102
L1204F[b]	Slightly	C-lobe	27.7 ± 1.1	0.159 ± 0.002	174	5.02 ± 0.19
R1231Q	No	C-lobe	5.35 ± 1.05	0.143 ± 0.015	37.4	1.23 ± 0.23
I1250T[b]	No (inactivating)	C-lobe	2.68 ± 0.18	0.150 ± 0.008	17.9	0.443 ± 0.045
T1343I	No	C-lobe	8.57 ± 1.27	0.160 ± 0.007	53.6	2.81 ± 0.17
D1349H	No	C-lobe	11.2 ± 1.8	0.148 ± 0.014	75.7	1.65 ± 0.02

[a]Previously determined (Bresler et al., 2011).
[b]Expression levels approximately 10-fold less than wild-type.
[c]Values are the mean ± SEM of at least three experiments.

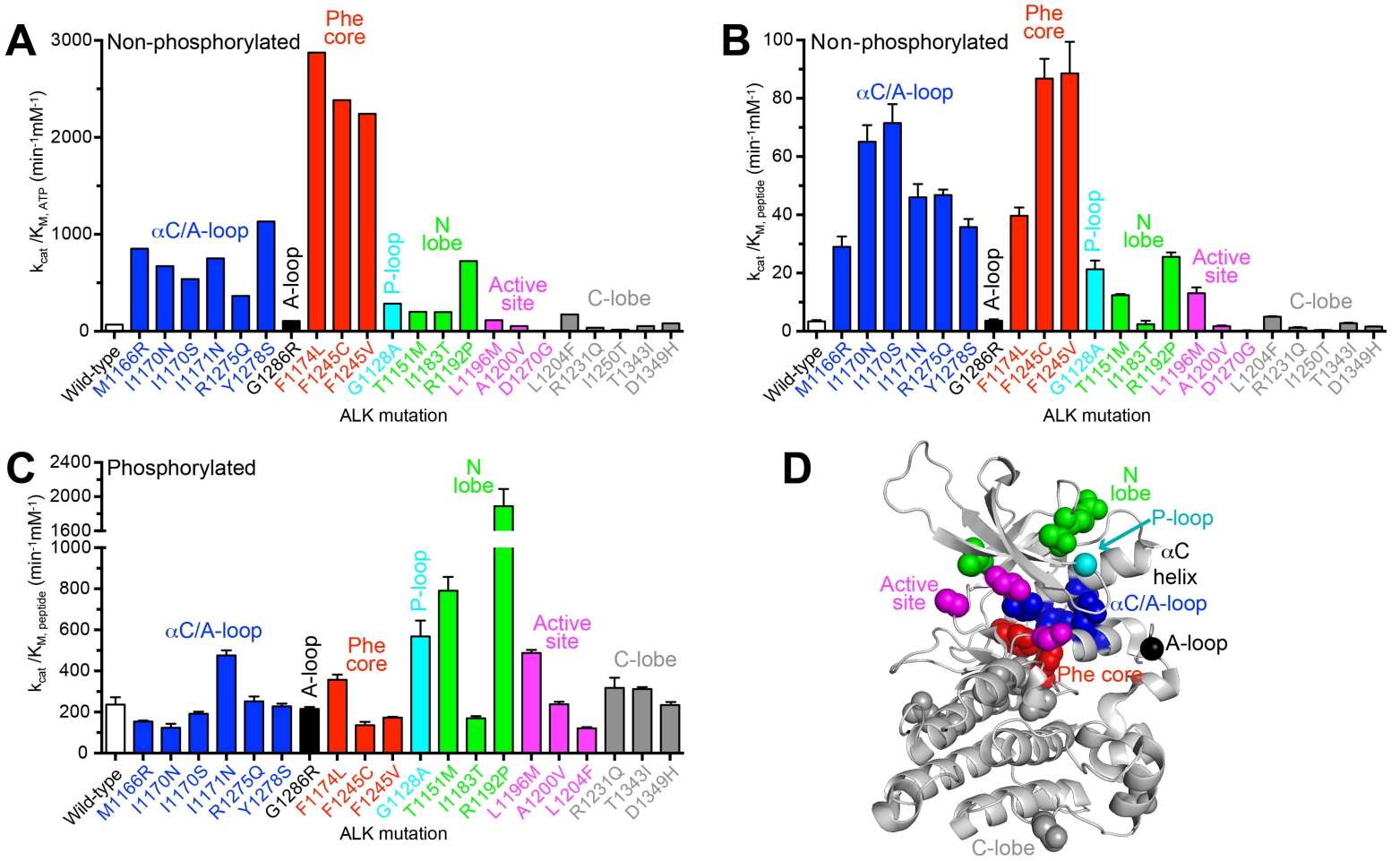

Figure S2, related to Figure 3.
Catalytic Efficiencies of ALK TKD Variants.

(**A**) $k_{cat}/K_{M,\ ATP}$ values for non-phosphorylated ALK TKD variants (N.B. error bars not present because k_{cat} and $K_{M,\ ATP}$ were determined separately, see Table S4). (**B**) $k_{cat}/K_{M,\ peptide}$ values for non-phosphorylated ALK TKD variants. (**C**) $k_{cat}/K_{M,\ peptide}$ values for phosphorylated ALK TKD variants. Data (obtained at 25˚C) are shown as the mean ± SEM for at least 3 independent experiments. (**D**) Crystal structure of ALK TKD as also shown in Figure 3D.

Table S5, related to Figure 3.
Kinetic Properties of Phosphorylated ALK TKD Variants.

Variant[c]	Location	k_{cat} (min^{-1})[d]	$k_{cat}/K_{M,\ peptide}$ ($min^{-1}\cdot mM^{-1}$)[d]
pWild-type[a]	--	424 ± 63	237 ± 35
pM1166R	αC/A-loop	461 ± 22	154 ± 5
pI1170N	αC/A-loop	150 ± 28	123 ± 20
pI1170S	αC/A-loop	287 ± 12	192 ± 10
pI1171N	αC/A-loop	542 ± 29	476 ± 24
pR1275Q[a]	αC/A-loop	347 ± 15	252 ± 24
pY1278S	αC/A-loop	477 ± 35	228 ± 13
pG1286R	A-loop	453 ± 6	215 ± 10
pF1174L[a]	Phe core	436 ± 51	357 ± 25
pF1245C	Phe core	302 ± 20	136 ± 16
pF1245V	Phe core	286 ± 24	173 ± 4
pG1128A	P-loop	455 ± 66	568 ± 77
pT1151M	N-lobe	722 ± 46	791 ± 67
pI1183T	N-lobe	362 ± 33	170 ± 10
pR1192P	N-lobe	828 ± 63	1890 ± 200
pL1196M	Active site	215 ± 38	488 ± 15
pA1200V	Active site	410 ± 23	238 ± 12
pL1204F[b]	C-lobe	251 ± 19	121 ± 5
pR1231Q	C-lobe	348 ± 67	318 ± 49
pT1343I	C-lobe	415 ± 9	312 ± 10
pD1349H	C-lobe	404 ± 10	234 ± 15

[a]Previously determined (Bresler et al., 2011).
[b]Expression level approximately 10-fold less than wild-type.
[c]D1270G and I1250T not analyzed due to limited ability to autophosphorylate.
[d]Values are the mean ± SEM of at least three experiments.

Table S6, related to Figure 4.
Quantitation of Focus Formation Assays for Intact ALK Variants in NIH 3T3 Cells.

Variant	Transforming?	Location	foci per colony (mean ± SEM)[a]	p
Wild-type	No	--	0.000263 ± 8.81e-5	--
Vector alone	No	--	0.000314 ± 0.000314	0.883
R1060H	No	JM region	0.000629 ± 0.000210	0.184
M1166R	**Yes**	αC/A-loop	0.0124 ± 0.00299	0.0156
I1170N	**Yes**	αC/A-loop	0.0398 ± 0.00592	0.0024
I1170S	**Yes**	αC/A-loop	0.0508 ± 0.00832	0.0037
I1171N	**Yes**	αC/A-loop	0.0446 ± 0.0111	0.0164
R1275Q	**Yes**	αC/A-loop	0.0427 ± 0.00509	0.0110
Y1278S	**Yes**	αC/A-loop	0.0491 ± 0.0130	0.0199
G1286R	No	A-loop	0.00107 ± 0.000509	0.195
F1174L	**Yes**	Phe core	0.0783 ± 0.0164	0.0101
L1240V	**Yes**	Phe core	0.0380 ± 0.00823	0.0117
F1245C	**Yes**	Phe core	0.0828 ± 0.0106	0.0015
F1245V	**Yes**	Phe core	0.106 ± 0.0186	0.0054
G1128A	**Yes**	P-loop	0.0271 ± 0.00839	0.0428
T1151M	No	N-lobe	0.00146 ± 0.000311	0.0207
I1183T	No	N-lobe	0.00296 ± 0.00271	0.376
R1192P	**Yes**	N-lobe	0.0533 ± 0.0124	0.0130
L1196M	**Yes**	Active site	0.0248 ± 0.00523	0.0151
A1200V	No	Active site	0.00158 ± 0.00131	0.373
D1270G	No	Active site	0.000149 ± 0.000149	0.545
L1204F	No	C-lobe	0.000117 ± 6.16e-5	0.247
R1231Q	No	C-lobe	3.31e-5 ± 3.31e-5	0.0709
I1250T	No	C-lobe	0.000529 ± 0.000125	0.157
T1343I	No	C-lobe	0.00110 ± 0.000351	0.0821
D1349H	No	C-lobe	0.00149 ± 0.000478	0.0656

[a]Values are the mean ± SEM of at least three experiments (each performed in duplicate).

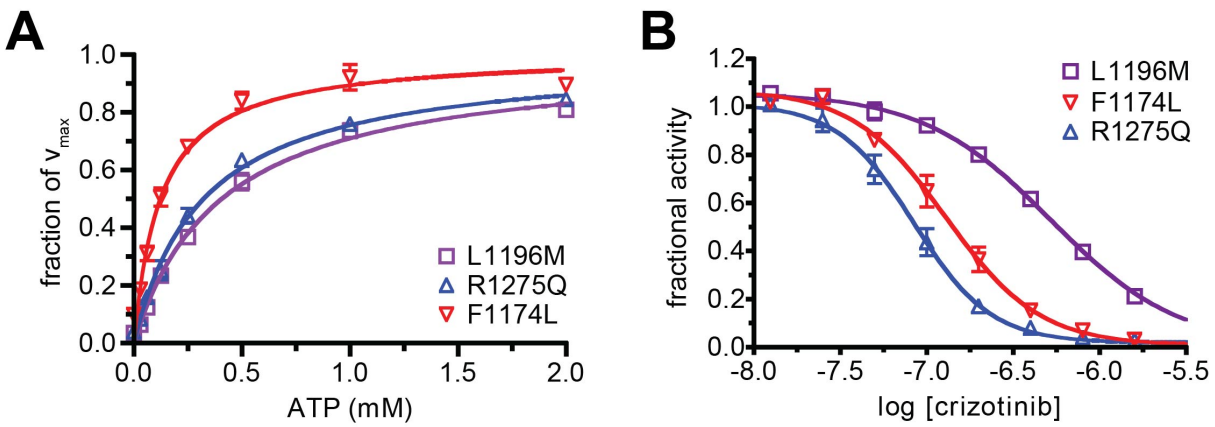

Figure S3, related to Table 3.
The L1196M Mutant is Crizotinib-Resistant Despite its high $K_{M, ATP}$.

(**A**) The L1196M mutant displays a $K_{M, ATP}$ value that approximates that of R1275Q. $K_{M, ATP}$ was measured at a constant YYY peptide concentration of 1.0 mM while varying the concentration of ATP. Enzyme concentrations were held at 50nM with an $MgCl_2$ concentration of 10 mM. (**B**) The L1196M mutation confers resistance to crizotinib (IC_{50} = 521 ± 4 nM, ~3.5-fold higher than wild-type or F1174L as reported in Table 3). Here, the concentration of crizotinib was varied from 0 to 25600 nM with an ATP concentration of 2.0 mM and a YYY peptide concentration of 1.0 mM. The enzyme concentration was held fixed at 50 nM. Results are the mean ± SEM of at least 3 independent experiments.

Supplemental Experimental Procedures

ALK copy number determination

Two datasets were used for copy number determination. Whole-genome SNP-array analyses (550k) were performed on 577 samples. Real-time quantitative PCR was performed on 1122 samples. The amplicon for *ALK* was located at 2p23.2a on intron 7 (Hs04707769_cn, Applied Biosystems). One reference amplicon, in the exonic region of *AGPS* (a unique, single-copy gene encoding alkylglycerone phosphate synthase) located at 2q31.2a, was used (Hs00927032_s1, Applied Biosystems), and was shown to have no independent copy number changes among 701 tumors analyzed. Two normal control DNA samples were used. After internal normalization, test samples were calculated for relative copy number of the *ALK* and *AGPS* amplicons. Values were calibrated to normal DNA and checked against ploidy to ensure appropriate copy number assignment. Ratios of less than 1.4 were interpreted as loss, and those greater than 2.5 were interpreted as gain, as described (D'Haene et al., 2010). A ratio of greater than 7 was interpreted as amplification. For 324 samples with overlapping SNP and real-time data, there was greater than 95% concordance of copy number evaluation. Samples with copy number data were analyzed only when there was overlapping mutation information.

Statistical considerations

For univariable analyses, event-free survival (EFS) time was calculated from the time of enrollment on the front-line or biologic study until the time of first occurrence of relapse, progressive disease, secondary malignancy, or death, or until the time of last contact if no event occurred. Overall survival (OS) time was calculated until the time of death, or until last contact if the patient was still alive. The methods of Kaplan-Meier were used to generate survival curves, and curves were compared using a log rank test (p values less than 0.05 were considered statistically significant). EFS and OS are expressed as the 5-year estimate ± the standard error,

with standard errors calculated using the methods of Peto. A chi-squared test was performed to test association between each risk factor and presence of *ALK* mutation, presence of *ALK* copy number variation, and presence of any *ALK* aberration. p values were not adjusted for multiple comparisons, with $p<0.05$ considered statistically significant. *ALK* mutations were categorized as F1174, F1245, R1275, or 'other'. Information for mutations and copy number alterations was combined into a new variable for *ALK* aberration, defined as mutation versus amplification, gain, or loss in *ALK* copy number.

For the multivariable analysis, using backwards selection model building, a Cox proportional hazards regression model was used to test for the independent statistical significance of *ALK* mutations or copy number variations after adjustment for statistically significant risk factors, and to calculate the p value and hazard ratio (HR) for each factor.

Recombinant protein expression, purification, and analysis of stability

Mutations were introduced using the QuikChange method (Stratagene). ALK TKD was treated with 1 μM YopH phosphatase, purified essentially as described (Zhang et al., 1992), for 12 hr at 4˚C to reverse spontaneous autophosphorylation that occurs within Sf9 cells during expression. Protein concentrations were determined by absorbance at 280 nm using the calculated extinction coefficient 39440 $cm^{-1}M^{-1}$. To test for a general effect of ALK TKD mutations on protein stability, we used circular dichroism to monitor unfolding as a function of temperature for wild-type protein and five different variants (M1166R, R1275Q, F1174L, R1192P, L1196M), and found that the melting temperature (47.4˚C for wild-type) was not reduced by more than 3.1˚C for any mutation (not shown).

Peptide phosphorylation and inhibition assays

Substrate phosphorylation by ALK TKD – and its inhibition by crizotinib – was analyzed exactly as described (Bresler et al., 2011), employing a peptide mimic of the ALK activation loop with sequence: biotin-ARDIYRASYYRKGGCAMLPVK referred to as 'YYY' peptide (Donella-Deana et al., 2005). Enzyme concentrations for non-phosphorylated proteins were 50nM for activating mutants (M1166R, I1170N, I1170S, I1171N, R1275Q, Y1278S, F1174L, F1245V, G1128A, R1192P, L1196M), 100 nM for T1151M, and 500 nM for wild-type-like mutants (G1286R, A1200V, L1204F, R1231Q, T1343I, and D1349H). D1270G was used at a concentration of 2 μM, and I1250T at 500 nM. Enzyme concentration for all phosphorylated proteins was fixed at 10 nM. Assays monitored incorporation of ^{32}P from γ-^{32}P ATP included at trace amounts in the reactions (~20 μCi per experiment). Peptide concentrations were determined by absorbance at 280 nm using a calculated extinction coefficient 3960 $cm^{-1}M^{-1}$. Since $K_{M,\ peptide}$ values were high (> 1 mM), and the assay did not perform well with peptide substrate concentrations greater than ~3 mM, $k_{cat}/K_{M,\ peptide}$ (catalytic efficiency) values were determined at lower peptide concentrations (where [peptide] < $K_{M,\ peptide}$) with linear fits to the equation: $V \sim (k_{cat}/K_{M,\ peptide})[Enz][peptide]$, and $K_{M,\ peptide}$ values were not directly measured.

Native gel kinase assays

Native gel electrophoresis was used to monitor autophosphorylation progress for 10 μM ALK TKD in 100 mM HEPES pH 7.4, 150 mM NaCl, 2 mM DTT, 2 mM ATP, 10 mM $MgCl_2$. A 10 μl aliquot was taken at each time point, the reaction quenched by adding EDTA to a final concentration of 20 mM EDTA, and placed on ice. Samples (10 μl) were then subjected to electrophoresis on 7.5% large-format (20 cm x 20 cm x 1 mm) native gels, with a pH 6.8 Tris-HCl stacking gel and a pH 8.8 Tris-HCl resolving gel, run using Tris-glycine running buffer (pH 8.8) at 100 V for ~14 hr Gels were stained with Coomassie brilliant blue R-250.

Focus formation assays

Low-passage NIH 3T3 cells were cotransfected with ALK variants subcloned into the MigR1 vector (Pear et al., 1998) and pSVneo DNA in a 10:1 ratio using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. The cells were allowed to recover for 2 days post-transfection, and were then divided into 2 samples and placed into 6-well plates. The first sample was allowed to reach confluence and form foci in DMEM containing 5% calf serum, while the second sample of cells was serially diluted and selected for colony formation in DMEM containing 10% calf serum and 0.5 mg/ml G418. The medium was changed every 3 days. After focus formation in the first group and colony formation in the second group, which typically took 10-14 days, cells were fixed in 3.7% formaldehyde in PBS for 5 min and then stained with 0.05% crystal violet. Foci and colonies were counted, and the number of foci counted for each ALK variant in each transfection was normalized by the number of G418-resistant colonies counted (correcting for dilution), to yield a transformation index as foci per G418-resistant colony. Each independent experiment was performed in duplicate. Final data (Figure 4B and Table S6) represent the mean and SEM of at least three independent experiments. Expression controls in parallel transfections established similar levels of expression for the different mutants by Western blotting (not shown).

Modeling methods

The inactive wild-type ALK TKD structure (residues 1096-1399) was taken from PDB entry 3LCS (Lee et al., 2010). Residues 1084-1095 and 1400-1405 were added to the model based on PDB entry 4FNW (Epstein et al., 2012) using MODELLER v9.8 (Eswar et al., 2006). Mutated structures were generated using MODELLER v9.8 by making point mutations to the modified inactive wild-type model. A homology model of active ALK TKD was generated with MODELLER v9.8, using as the primary template the active insulin receptor TKD structure (PDB entry 1IR3),

with which ALK TKD shares 46% sequence identity (Hubbard, 1997). Residues 1097-1399 were modeled from 3LCS, whereas residues 1084-1096 and 1400-1405 were again modeled from 4FNW. All structures were modeled without bound substrate.

Molecular dynamics (MD)

All structures were subjected to the same molecular dynamics (MD) protocol. Hydrogen atoms were added to the structures with Automatic PSF Generation Plugin v1.3 implemented in VMD 1.8.6 (Humphrey et al., 1996). To reflect a physiological pH of 7.0, all histidines express a +1 protonation state on the δ-nitrogen. The Solvate Plugin v1.5 and Autoionize Plugin v1.3 implemented in VMD were used to construct an electroneutral water box with 15 Å TIP3P water padding and 0.15 M Na^+/Cl^- concentration. All Na^+ and Cl^- ions were placed at least 5 Å away from protein atoms and each other. Systems contained approximately 60500 atoms.

All MD simulations were carried out with NAMD v2.8 (Phillips et al., 2005) using CHARMM27 force field parameters (MacKerell Jr et al., 1998). Periodic boundary conditions were used throughout. The particle mesh Ewald algorithm was used to treat long-range electrostatic interactions. An integration timestep of 2 fs was used. Bonds between hydrogens and heavy atoms were constrained to their equilibrium values, with the velocity correction being performed by the RATTLE algorithm (Anderson, 1983). Rigid waters were treated using the SETTLE algorithm (Miyamoto and Kollman, 1992). Long-range non-bonded van der Waals (VDW) interactions were treated by applying a smooth switching function at 10 Å with a cutoff distance of 12 Å.

To eliminate unfavorable contacts, the solvated systems underwent an energy minimization using a conjugate gradient algorithm; they were then gradually heated to 300 K. Constant temperature and pressure (NPT) simulations using a Nosé-Hoover Langevin piston (Feller et al., 1995; Martyna et al., 1994) were performed at 300 K and 1 atm to equilibrate the

volume of the solvation box. Subsequently, constant temperature and volume (NVT) simulations were run on the system. After an equilibration period, 40 ns of NVT simulation were completed on each structure.

Hydrogen-bond analysis

Hydrogen bond (H-bond) analysis was performed on the trajectory of each system using the HBonds Plugin v1.2 in VMD. Hydrogen-bond cutoff lengths of 3.2 Å and angle cutoffs of 30˚ were chosen to include H-bonds of moderate strength. The occupancies for each residue-to-residue H-bond range from 0% to 100% across the trajectory in each system. Full details of the analysis will be published separately (Huwe et al., in preparation).

A scoring function was created to analyze how 'active-like' the hydrogen bond networks were for each system, as follows:

1. For each hydrogen bond, the difference in occupancies between the active (A) and inactive (I) wild-type (wt) systems ($\Delta_{wt} = Y^{I}_{wt} - Y^{A}_{wt}$) was calculated.
2. For each bond, if $|\Delta_{wt}| > 40.0\%$, the difference in occupancies between the inactive wt and inactive mutant (mut) for each mutation ($\Delta_{mut} = Y^{I}_{wt} - Y^{I}_{mut}$) was calculated.
3. If $\Delta_{mut}/\Delta_{wt} > 0.5$, then the bond received a binary activation score of 1; otherwise, it received a score of 0.

The scores were tallied for 28 hydrogen bonds in the activation loop and the αC helix. A score of 5 or greater for a given variant was considered activating.

Hydrophobic destabilization analysis

Solvent accessible surface area (SASA) values (Connolly, 1983) were calculated in VMD using the measure SASA module, with a probe radius 1.4 Å larger than the van der Waals radius. The SASA was calculated on a per-residue basis for the residues forming the hydrophobic core

involving the activation loop, αC helix, and extended 'Phe core' (Y1096, F1098, I1170, I1171, F1174, I1179, Y1239, L1240, F1245, and F1271). The SASA values (in units of $Å^2$) were averaged over all steps of the MD trajectory, from which mean SASA values were computed for each relevant amino acid. These SASA scores were summed and compared to the summed score for the wild-type protein.

Additionally, free energy perturbation (FEP) simulations (Beveridge and DiCapua, 1989) were performed for each mutant on the inactive and active ALK TKD structures to determine computationally how each mutation affects the relative stability of the two TKD conformations. We used the dual-topology approach of FEP as implemented in NAMD (Axelsen and Li, 1998; Gao et al., 1989; Phillips et al., 2005). The potential energy function characteristic of the native residue is scaled into that representing the new residue over the course of an MD simulation. As the old residue fades out and the new residue fades in, the old and new do not interact with each other. Simulations were carried out in both the forward and reverse directions, with soft-core potentials employed to avoid "end-point catastrophes" (Beutler et al., 1994). Forward direction ΔΔG are calculated using the following equation: $\Delta\Delta G = \Delta G^{active}_{wt \rightarrow mut} - \Delta G^{inactive}_{wt \rightarrow mut}$. Forward ΔΔG values are only considered significant if the ΔΔG value is greater than the standard deviation between the forward and reverse results. Mutated systems were scored as activating in the SASA/FEP column of Table 4 if the following are true:

1. The summed SASA values for the residues that contribute to the hydrophobic core mentioned above are at least 25 $Å^2$ greater for the mutant than for wild-type ALK TKD.
2. The FEP results yielded a statistically significant negative value for ΔΔG.

Principal component analysis (PCA)

Principal Component Analysis (PCA) as implemented in Carma (Glykos, 2006) was performed on the full trajectories of each system. Principal components were obtained by diagonalizing the covariance matrix of atomic fluctuations in Cartesian space to produce eigenvalues and eigenvectors. Only α carbons of protein components were analyzed. Translations and rotations were removed by aligning all residues that were not in the activation loop, P-loop, catalytic loop, or αC helix. Eigenvalues were summed for each system and ranged from 287 to 595 $Å^2$. Every mutant with a top eigenvalue above 200 $Å^2$ was given an 'activating' score in Table 4.

Supplemental References

Anderson, H.C. (1983). Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. J. Comp. Phys. *52*, 24–34.

Axelsen, P.H., and Li, D. (1998). Improved convergence in dual-topology free energy calculations through use of harmonic restraints. J. Comput. Chem. *19*, 1278-1283.

Beutler, T.C., Mark, A.E., van Schaik, R.C., Gerber, P.R., and van Gunsteren, W.F. (1994). Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem. Phys. Lett. *222*, 529-539.

Beveridge, D.L., and DiCapua, F.M. (1989). Free energy via molecular simulation: Applications to chemical and biomolecular systems. Annu. Rev. Biophys. *18*, 431-492.

Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. Science *221*, 709-713.

D'Haene, B., Vandesompele, J., and Hellemans, J. (2010). Accurate and objective copy number profiling using real-time quantitative PCR. Methods *50*, 262-270.

Donella-Deana, A., Marin, O., Cesaro, L., Gunby, R.H., Ferrarese, A., Coluccia, A.M., Tartari, C.J., Mologni, L., Scapozza, L., Gambacorti-Passerini, C., and Pinna, L.A. (2005). Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. Biochemistry *44*, 8533-8542.

Epstein, L.F., Chen, H., Emkey, R., and Whittington, D.A. (2012). The R1275Q neuroblastoma mutant and certain ATP-competitive inhibitors stabilize alternative activation loop conformations of anaplastic lymphoma kinase. J. Biol. Chem. *287*, 37447-37457

Eswar, N., Marti-Renom, M.A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling with MODELLER. Current Protocols in Bioinformatics *Supplement 15*, 5.6.1-5.6.30.

Feller, S.E., Zhang, Y., Pastor, R.W., and Brooks, B.R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. *103*, 4613-4621.

Gao, J., Kuczera, K., Tidor, B., and Karplus, M. (1989). Hidden thermodynamics of mutant proteins: A molecular dynamics analysis. Science *244*, 1069-1072.

Glykos, N.M. (2006). Carma: a molecular dynamics analysis program. J. Comput. Chem. *27*, 1765-1768

Hubbard, S.R. (1997). Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. EMBO J. *16*, 5572-5581.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD - Visual Molecular Dynamics. J. Molec. Graphics *14*, 33-38.

MacKerell Jr, A.D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S.*, et al.* (1998). All-atom empirical potential for molecular

modeling and dynamics studies of proteins. J. Phys. Chem. B. *102*, 3586-3616.

Martyna, G.J., Tobias, D.J., and Klein, M.L. (1994). Constant pressure molecular dynamics algorithms. J. Chem. Phys. *101*, 4177-4189.

Miyamoto, S., and Kollman, P.A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comp. Chem. *13*, 952–962.

Pear, W.S., Miller, J.P., Xu, L., Pui, J.C., Soffer, B., Quackenbush, R.C., Pendergast, A.M., Bronson, R., Aster, J.C., Scott, M.L., and Baltimore, D. (1998). Efficient and rapid induction of a chronic myelogenous leukemia-like myeloproliferative disease in mice receiving P210 bcr/abl-transduced bone marrow. Blood *92*, 3780-3792.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781-1802.

Zhang, Z.Y., Clemens, J.C., Schubert, H.L., Stuckey, J.A., Fischer, M.W., Hume, D.M., Saper, M.A., and Dixon, J.E. (1992). Expression, purification, and physicochemical characterization of a recombinant Yersinia protein tyrosine phosphatase. J. Biol. Chem. *267*, 23759-23766.