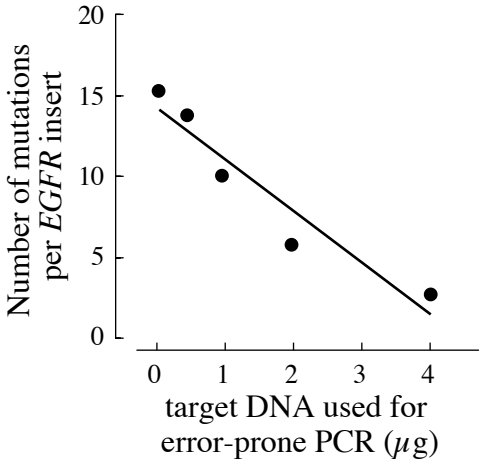
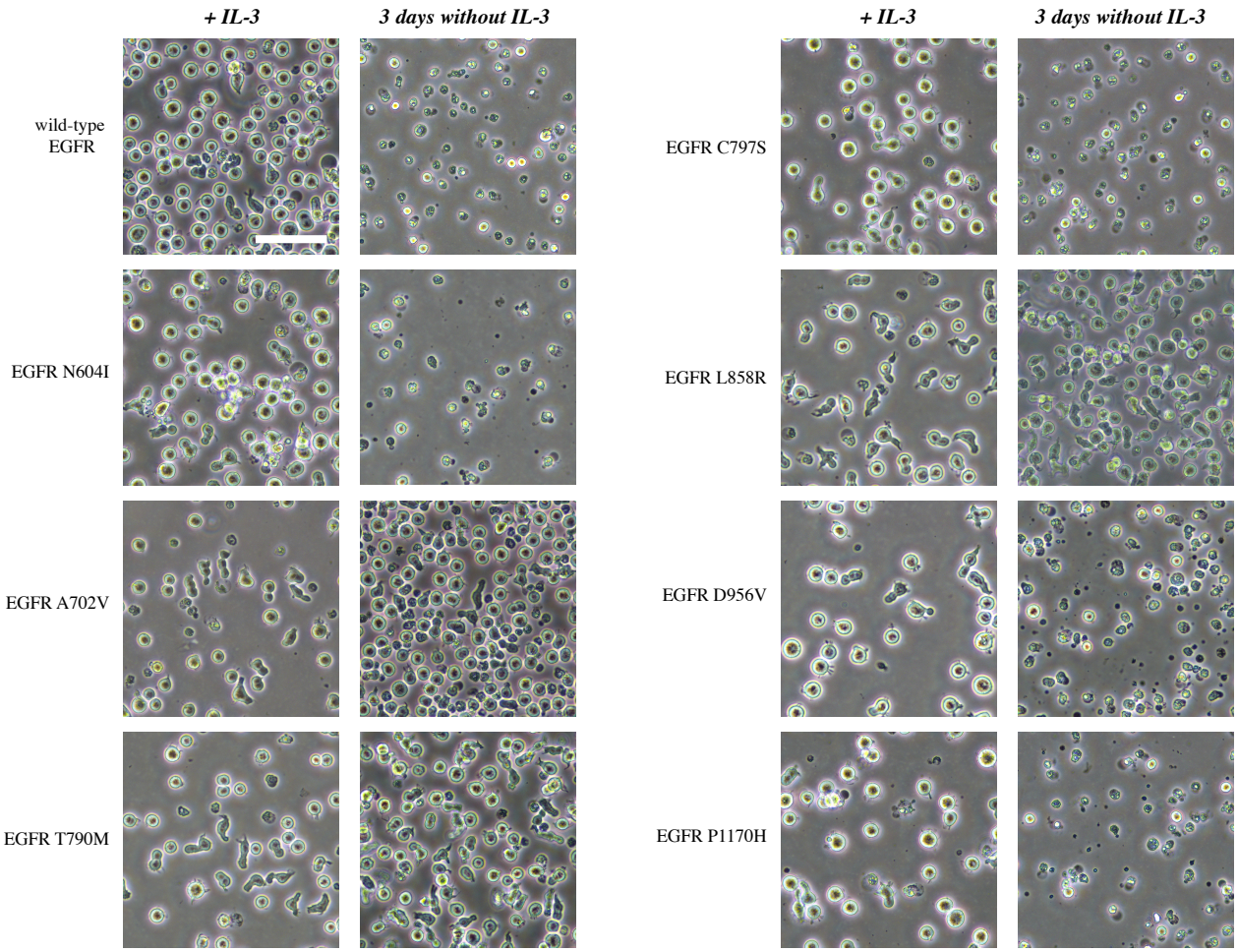
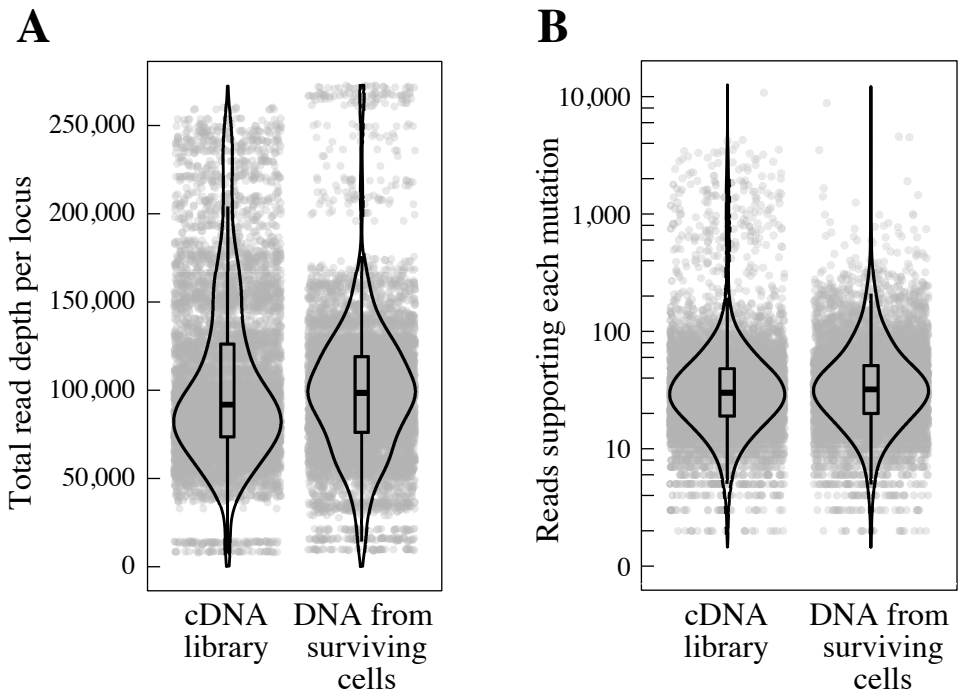


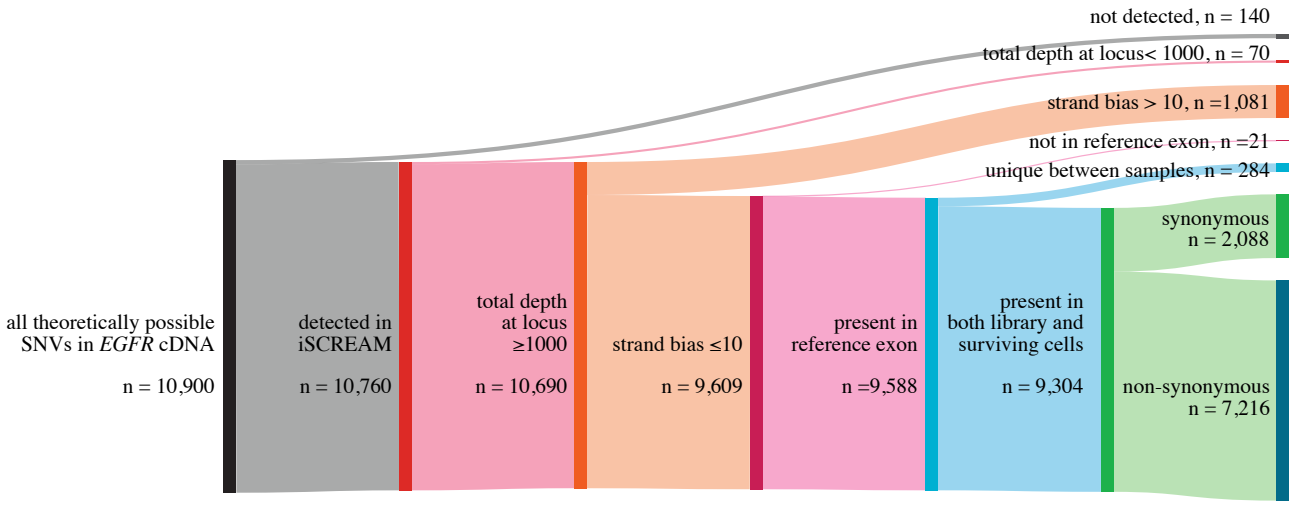
Supplementary Figure S1. Mutation frequency after error-prone PCR. Detected mutation frequency in different randomly mutated *EGFR* cDNA libraries, shown as a function of varying amounts of wild-type *EGFR* template DNA used for the PCR reaction. The solid line shows a linear fit to the observed mutation frequency.

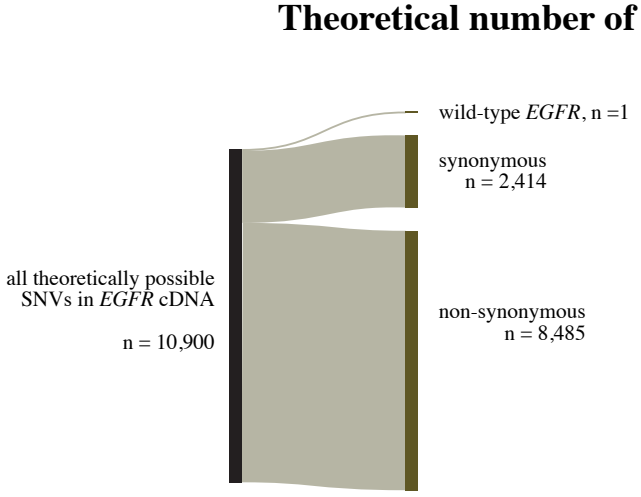


Supplementary Figure S2. Survival of Ba/F3 cells expressing EGFR variants. Microscopic images of Ba/F3 cells expressing the indicated EGFR variants. The cells were cultured for 3 days in the presence or absence of IL-3. Expression of the EGFR variants A702V, T790M and L858R supported IL-3-independent Ba/F3 cells survival. Expression of wild-type EGFR or the other indicated EGFR variants did not. Scale bar in top left panel, 50 μm.



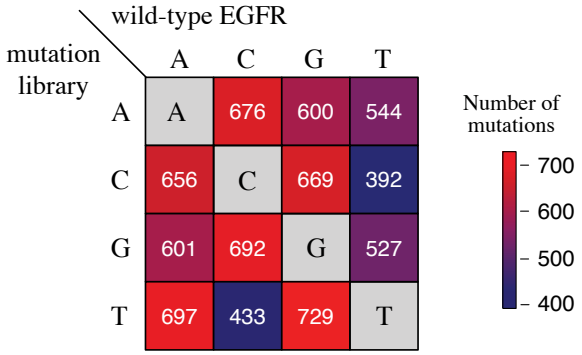
Supplementary Figure S3. Depth of targeted next-generation sequencing. Scatter plots indicate **A**) total read depth and **B**) reads supporting a particular variant (log10 scale) for each of the 7,216 mutations detected in the NGS analysis of the *EGFR* cDNA expression library and the *EGFR* insert amplified from the surviving cells. The distribution of the data points is summarized using violin and boxplots.



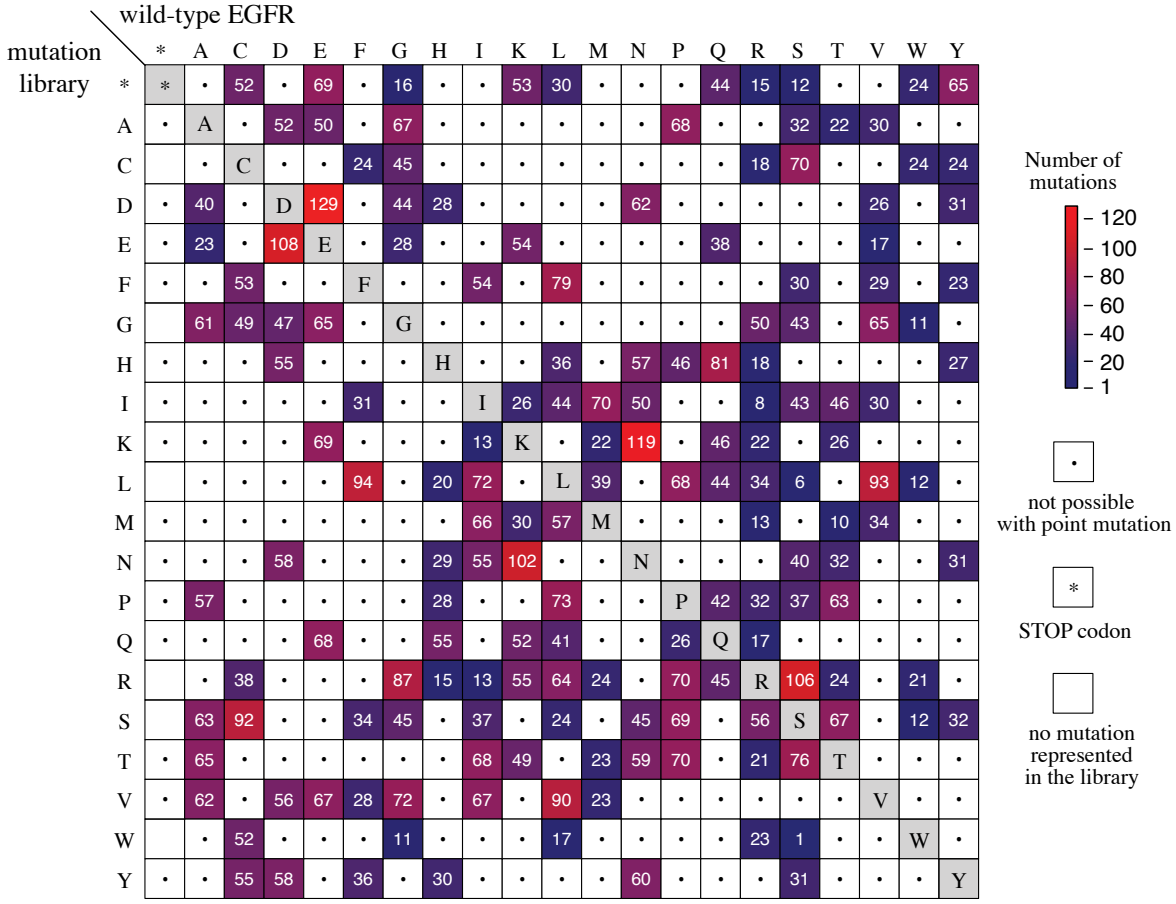


Supplementary Figure S4. Number of *EGFR* variants analyzed. Sankey diagram indicates the filtering workflow used during data analysis.

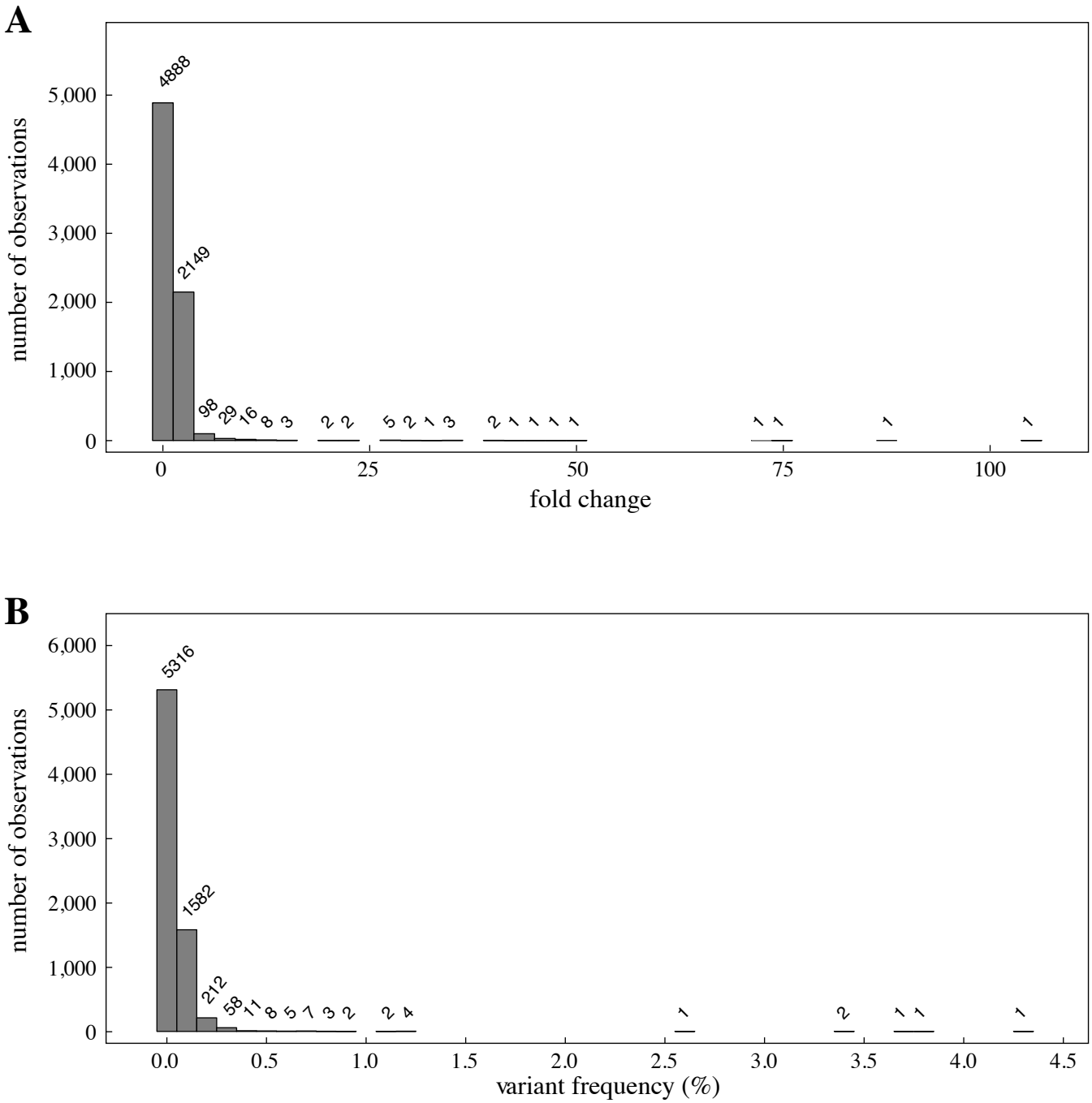
A **Mutation distribution at nucleotide level**



B **Mutation distribution at amino acid level**



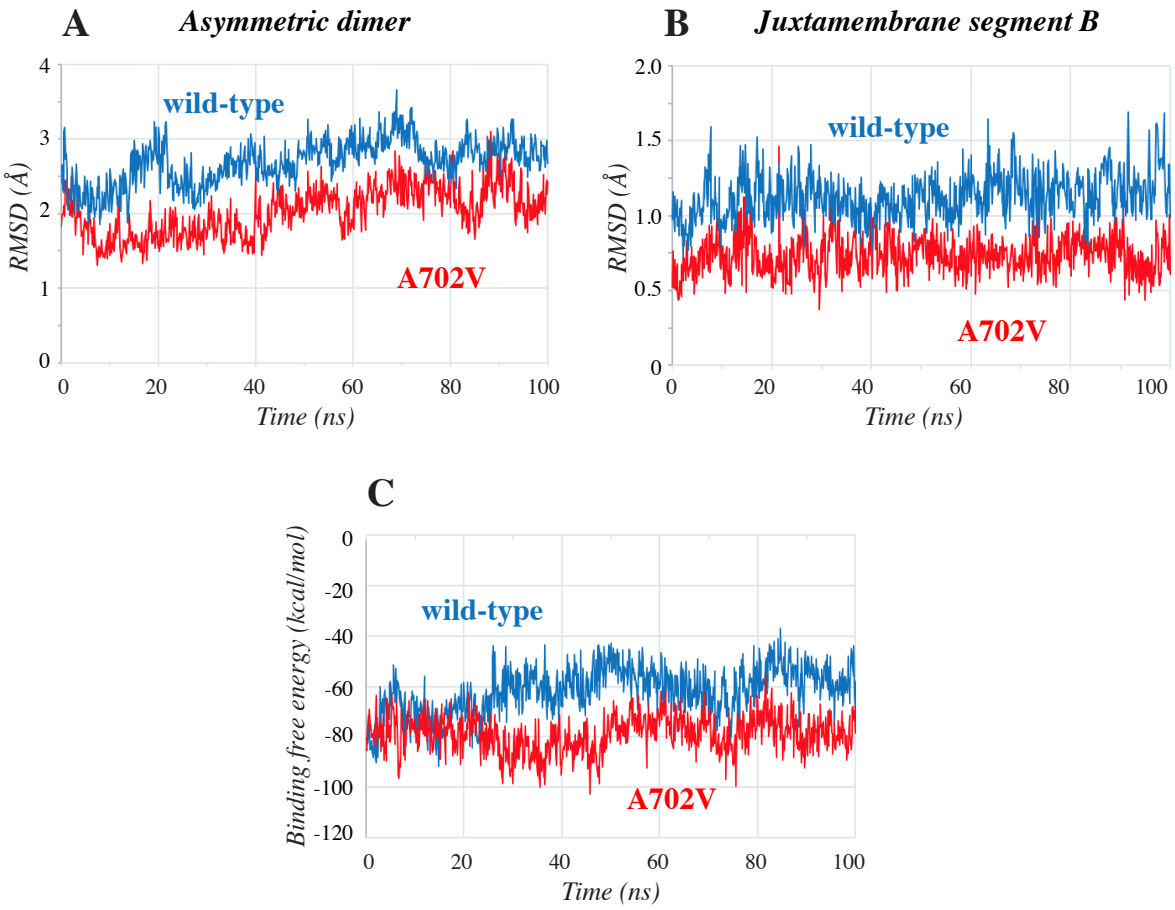
Supplementary Figure S5. Frequency of mutations observed in the *EGFR* random mutation library. **A**) Matrix showing distribution of observed nucleotide changes in the mutation library. The column labels represent the nucleotide in the wild-type *EGFR* cDNA reference sequence, and the row labels represent the mutations observed for that nucleotide. The observed frequency of a particular nucleotide change in the mutation library is shown in the cell at the intersection of the reference and mutant nucleotide pair. **B**) Matrix showing distribution of observed amino acid changes in the mutation library. Column labels represent wild-type EGFR primary sequence, and the row labels represent the amino acid changes observed for the reference amino acid. The observed frequency of a particular amino acid change is shown in the cells at the intersection of reference and mutant amino acid pair. The cells marked with a dot are amino acids changes that are not possible in the EGFR primary sequence by single point mutations.



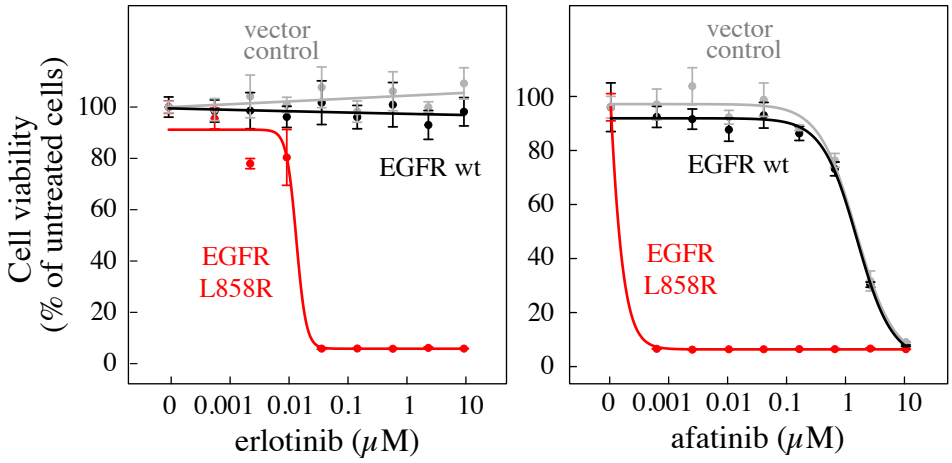
Supplementary Figure S6. Relative numbers of enriched EGFR mutations. A) Histogram showing the distribution of the 7,216 *EGFR* mutations in different categories of fold change between the surviving cell pool and the cDNA expression library. **B**) Histogram showing the distribution of the 7,216 *EGFR* mutations in different categories of variant frequency in the surviving cell pool.

Supplementary Figure S7 is available as a separate file: Suppl. Fig. S7.html

Supplementary Figure S7. An interactive map of the screen for activating EGFR mutations. A web-browser friendly version of Figure 3 showing the scatter plot depicting the fold change for the 7,216 EGFR mutations analyzed. The y-axis indicates the observed fold change in variant frequency of a particular mutation when the mutational signature of the surviving cells was compared to that of the mutation library. The x-axis shows the position of the mutated residue on the EGFR primary sequence. The size of the dot as well as the intensity of its red color indicate the variant frequency of the mutation in the surviving cell pool. In addition to listing the mutation and the observed fold change, on hovering the mouse cursor over a particular mutation, the tooltip lists the variant frequency and the number of reads supporting the mutation in the surviving cell pool and in the mutation library.



Supplementary Figure S8. Effect of EGFR A702V on molecular dynamics of the kinase interface. **A**) RMSD values (backbone atoms) for the wild-type EGFR (blue) and A702V mutant (red) asymmetric dimers in reference to their initial structures as a function of the simulation time. **B**) RMSD values (backbone atoms) of the juxtamembrane segment B of the wild-type (blue) and A702V mutant EGFR (red) kinase domains from their starting structures plotted against the simulation time. **C**) Binding free energy of the wild-type EGFR (blue) and A702V mutant (red) kinase asymmetric dimers plotted against the simulation time.



Supplementary Figure S9. Sensitivity of controls cells expressing wild-type EGFR to erlotinib and afatinib. Drug response curves of Ba/F3 cells expressing wild-type (wt) EGFR, EGFR L858R or EGFP (vector control). Cells were cultured for 3 days in the presence of the indicated concentrations of the drugs. Cells expressing EGFR L858R were cultured in the absence of IL-3, cells expressing wt EGFR and the vector control cells in the presence of IL-3.

MUTATION	ORIENTATION	DNA SEQUENCE (5'-3')
N604I	FW	aacatcaccctggtctgg
	REV	ttctcccatgactcctgc
A702V	FW	caagttctcttgaggatctt
	REV	gttgggagcttctccact
T790M	FW	atcatgcagctcatgccc
	REV	gagttgcacggtggaggtg
C797S	FW	ggcagcctcctggactatg
	REV	gaagggcatgagctgcgt
L858R	FW	ggcgggccaaactgctg
	REV	caaaatctgtgatcttgacatgctg
D956V	FW	gcagttagtcgcccaaag
	REV	gtctatcatccagcacttgacc
P1170H	FW	aaccatgactaccagcaggac
	REV	gtccaggctaatttggtggc

Supplementary Figure S10. Primers. Forward (FW) and reverse (REV) primers used for site-directed mutagenesis to generate the indicated point mutations in EGFR are listed.