AUTHORS Koivu MKA, Chakroborty D, Tamirat MZ, Johnson MS, Kurppa KJ, Elenius K

TITLE Identification of Predictive ERBB Mutations by Leveraging Publicly Available Cell Line Databases

YEAR 2020

DOI 10.1158/1535-7163.MCT-20-0590

VERSION Final draft

CITATION Identification of Predictive ERBB Mutations by Leveraging Publicly Available Cell Line Databases. Marika K. A. Koivu, Deepankar Chakroborty, Mahlet Z. Tamirat, Mark S. Johnson, Kari J. Kurppa, and Klaus Elenius. Mol Cancer Ther December 15 2020 DOI: 10.1158/1535-7163.MCT-20-0590.

Identification of Predictive *ERBB* Mutations by Leveraging Publicly Available Cell Line Databases

Marika K. A. Koivu[1,2,3], Deepankar Chakroborty[1,2,3], Mahlet Z. Tamirat[4], Mark S. Johnson[4], Kari J. Kurppa[1], and Klaus Elenius[1,5]

[1]Institute of Biomedicine, and Medicity Research Laboratories, University of Turku, Turku, 20520, Finland

[2]Turku Doctoral Programme of Molecular Medicine, Turku, 20520, Finland

[3]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, 20520, Finland

[4]Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and Engineering, Åbo Akademi University, Turku, 20500, Finland

[5]Department of Oncology, Turku University Hospital, Turku, 20521, Finland

Running title: Predictive ERBB Mutations

Keywords: ERBB, cancer, cell line databases, HER, predictive mutations, targeted drugs

Corresponding author: Klaus Elenius, MediCity Research Laboratories, University of Turku, Tykistökatu 6A, FI-20520 Turku, Finland, Tel: +358-29-4504393, E-mail: klaus.elenius@utu.fi

Conflict of interest: K.E. and K.J.K. have a research agreement with Puma Biotechnology. K.E. has ownership interest in Abomics, Novo Nordisk, Orion, and Roche.

While targeted therapies can be effective for a subgroup of patients, identification of individuals who benefit from the treatments is challenging. At the same time, the predictive significance of the vast majority of the thousands of mutations observed in the cancer tissues remains unknown. Here, we describe the identification of novel predictive biomarkers for ERBB-targeted tyrosine kinase inhibitors (TKI) by leveraging the genetic and drug screening data available in the public cell line databases: Cancer Cell Line Encyclopedia (CCLE), Genomics of Drug Sensitivity in Cancer (GDSC), and Cancer Therapeutics Response Portal (CTRP). We assessed the potential of 412 ERBB mutations in 296 cell lines to predict responses to 10 different ERBB-targeted TKIs. Seventy-six ERBB mutations were identified that were associated with ERBB TKI sensitivity comparable to non-small cell lung cancer cell lines harboring the well-established predictive EGFR L858R mutation or exon 19 deletions. Fourteen (18.4 %) of these mutations were classified as oncogenic by the cBioPortal database, whereas 62 (81.6 %) were regarded as novel potentially predictive mutations. Out of nine functionally validated novel mutations, EGFR Y1069C and ERBB2 E936K were transforming in Ba/F3 cells and demonstrated enhanced signaling activity. Mechanistically, the EGFR Y1069C mutation disrupted the binding of the ubiquitin ligase c-CBL to EGFR, whereas the ERBB2 E936K mutation selectively enhanced the activity of ERBB heterodimers. These findings indicate that integrating data from publicly available cell line databases can be used to identify novel, predictive non-hotspot mutations, potentially expanding the patient population benefiting from existing cancer therapies.

Introduction

While cancer tissues harbor thousands of mutations (1,2), only few have clinical value as predictive biomarkers guiding therapeutic decision making (3). This imbalance may be explained by several factors. A simple explanation is that the vast majority of all somatic mutations are non-functional passengers resulting from the general genetic instability of the cancer genomes (4). On the other hand, statistical data about the so-called long tail of the frequency distribution of somatically mutated genes across cancer samples indicate that 85% of all hotspots are found in less than 5% of tumors of all cancer types (5,6). These latter observations imply that a significant number of potentially clinically relevant predictive mutations still remain to be identified. This conclusion also involves that a major obstacle for identifying new predictive mutations has been their rare occurrence and relative lack of structurally defined hotspots altering specific amino acid residues. In addition, technical obstacles have limited the use of functional *in vitro* analyses to test thousands of mutations in relevant cellular and molecular contexts (7).

The ERBB family of RTKs contains four homologous receptors: EGFR/ERBB1, ERBB2/HER2, ERBB3/HER3, and ERBB4/HER4. Upon ligand binding, the ERBB receptors form either homo- or heterodimers which results in autophosphorylation, and activation of downstream signaling (8–10). In cancer tissues, the activation may become enhanced or ligand-independent by somatic oncogenic mechanisms including single missense mutations, small deletions and insertions, truncation of longer segments of the receptor ectodomain, or amplification (11–13). These events may induce oncogene addiction and sensitivity of the cancer cells to compounds blocking ERBB signaling.

A number of ERBB targeting compounds have been approved for clinical use. The best characterized examples include the anti-ERBB2 antibody trastuzumab that is used to

treat early and metastatic breast cancer with *ERBB2* amplification, and the EGFR-selective TKIs erlotinib and gefitinib that have been approved for non-small cell lung cancer (NSCLC) harboring the missense mutations or small deletions in the kinase domain of EGFR (14). More recently, the repertoire of ERBB TKIs has been expanded by osimertinib that is designed to also block signaling by cells harboring an EGFR T790M mutation conferring resistance to the first generation EGFR TKIs (15), as well as the pan-ERBB TKIs afatinib, neratinib and dacomitinib, that block the signaling of all kinase-competent ERBB receptors (16–18). Activating somatic mutations have also been reported in *ERBB2*, *ERBB3* and *ERBB4* in different cancer types (3,19–24). However, the predictive significance of these remains to be fully elucidated.

Several publicly available databases include sequencing and drug sensitivity data from thousands of cancer tissue samples and cell lines (25–27). These data allow systematic analysis of the relationship between genetic alterations and drug sensitivity in relevant cellular contexts. We hypothesized that mutations that confer sensitivity to a targeted drug, such as an ERBB-targeted TKI, can be identified as outliers in the drug sensitivity data by comparing cells lines harboring wild-type and mutant variants of the specific target gene for the drug.

By combining the drug sensitivity data from Cancer Cell Line Encyclopedia (CCLE) (25), Genomics of Drug Sensitivity in Cancer (GDSC) (28), and Cancer Therapeutics Response Portal (CTRP) (29), we were able to assess the predictive potential of 412 ERBB mutations in 296 cancer cell lines to 10 different ERBB-targeting TKI compounds. By normalizing the data with the known sensitivity of cells harboring the classical predictive EGFR mutations to EGFR-selective TKIs, putative novel predictive ERBB mutations were identified. These findings demonstrate the potential of the approach in identifying predictive biomarkers for cancer therapy with oncogene-targeted drugs.

Materials and Methods

Databases and calculation of rAUC and mEC50 values

Human cancer cell lines with available mutation data were searched from COSMIC (Catalogue of Somatic Mutations in Cancer; https://cancer.sanger.ac.uk/cosmic; v71) and CCLE (https://portals.broadinstitute.org/ccle/home; 24-Feb-2015 data) databases. CNV data were searched from cBioPortal (https://www.cbioportal.org/; based on CCLE/Broad 2019 information). For cell line nomenclature, names used by CTRP (https://ocg.cancer.gov/programs/ctd2/data-portal; v2) were primarily used, and when not available from CTRP, adopted from GDSC (http://www.cancerrxgene.org; v17.3), as indicated in Supplementary Table S1.

ERBB TKI drug sensitivity data for cell lines with known mutation status for any of the four ERBB genes were downloaded from the websites of CCLE, GDSC and CTRP. Data were processed with RStudio (R; RRID:SCR_000432) (30) using code available at the Code Ocean (DOI: 10.24433/CO.3524861.v1). To normalize AUC data, relative AUC (rAUC) values were established for each cancer cell line/ERBB TKI/database combination. To this end positive and negative reference values were generated for each TKI and database as described in Results. To normalize EC50 data, modeled EC50 (mEC50) values were determined after log-logistic modeling (code available at the Code Ocean; DOI: 10.24433/CO.3524861.v1) with Akaike Information Criterion (AIC). Potential predictive mutations were identified from the dataset by selecting cell lines with a rAUC value of ≥ 1.0 and a mEC50 value of ≤ 1.0 μM for a given drug. All the processed data are included in Supplementary Table S2.

Cell culture

Ba/F3 (DSMZ; RRID CVCL_0161) and Phoenix Ampho cells (a gift from Dr. Garry Nolan; RRID CVCL_H716) were cultured in RPMI 1640 (Lonza) supplemented with 10% FBS (Biowest), ultraglutamine 1 (Lonza), 50 U/ml penicillin and 50 U/ml streptomycin (Gibco). For maintenance of the Ba/F3 cells, the culture medium was further supplemented with interleukin-3 (IL-3) in the form of 5% conditioned WEHI cell medium. COS-7 (RRID CVCL_0224) and NIH-3T3 (ATCC; RRID CVCL_0594) cells were cultured in DMEM (Lonza) supplemented with 10% FCS, 50 U/ml penicillin, and 50 U/ml streptomycin. Cells were routinely tested for mycoplasma infection using MycoAlert (Lonza).

Plasmid constructs

pBABE-puro-gateway empty vector was a gift from Matthew Meyerson (Addgene plasmid #51070; http://n2t.net/addgene:51070; RRID:Addgene_51070). pBABE-puro-gateway-*EGFR* has been described (7). pBABE-puro-gateway-*ERBB2* was a gift from Matthew Meyerson (Addgene plasmid #40978; http://n2t.net/addgene:40978; RRID:Addgene_40978). pBABE-puro-gateway-*ERBB3*, and pBABE-puro-gateway-*ERBB4JM-aCYT-2* plasmids were constructed by Gateway cloning (Addgene; (7)) using pcDNA3.1*ERBB3-HA* (31) and pcDNA3.1*ERBB4JM-aCYT-2* (32) as the sources for inserts. Point mutations to *ERBB* inserts were introduced by site-directed mutagenesis using oligonucleotide primers listed in Supplementary Table S3. All constructs were verified by sequencing.

Generation of Ba/F3 and NIH-3T3 cells with stable ERBB expression

pBABE-based ERBB expression vectors were transfected into Phoenix Ampho cells using Fugene6 transfection reagent (Promega) for the production of infective retroviruses. Ba/F3 and NIH-3T3 cells were infected with the retroviral supernatants, as previously described (7). Cells with integrated retroviral inserts were selected with 2 µg/ml puromycin (Gibco) for 48

hours. Selection pressure was subsequently maintained by the presence of 1 µg/ml puromycin in the culture medium. To generate Ba/F3 cells with simultaneous expression of both ERBB2 and ERBB3, stable puromycin-selected cells expressing ERBB3 were transduced with supernatants from Pheonix Ampho cells expressing pBABE-puro-gateway-*ERBB2*.

Transient transfection of COS-7 cells

COS-7 cells plated on 6-well plates were transfected with the pBABE-puro-gateway-*EGFR*, pBABE-puro-gateway-*ERBB2* and/or pBABE-puro-gateway-*ERBB3* constructs or the empty pBABE-puro-gateway vector using Fugene6. Total of 2 µg of plasmid DNA diluted in Opti-MEM (Gibco) was used per transfection. Cells were used for Western analyses 48 hours after transfection.

Analysis of Ba/F3 cell growth

Ba/F3 cells washed twice with PBS were seeded at 100,000 cells/ml in RPMI-1640 + 10% FCS supplemented with or without IL-3 (5% WEHI cell conditioned medium), or 10 ng/ml EGF (R&D Systems) or 20 ng/ml NRG-1 (R&D Systems). Quadruplicate 100 µl aliquots (corresponding to 10,000 cells at the initial seeding density) of the cultures were collected daily into 96-well plate wells for analysis of cell viability with the MTT assay (CellTiter 96 nonradioactive cell proliferation assay; Promega).

Drug response assays

Quadruplicate aliquots of Ba/F3 cells (cultured in the presence or absence of IL-3 and/or ERBB ligand as indicated) were seeded at 5,000 cells/96-well plate wells in a volume of 75 µl/well and incubated overnight. The following day ERBB TKIs erlotinib, lapatinib, afatinib or neratinib (all from Santa Cruz Biotechnology) were applied in a volume of 75 µl/well to

reach a final concentration ranging from 0 to 10 μM. Cells were incubated in the presence of the drugs for another 72 hours after which the number of viable cells was analyzed with the MTT-assays. Ba/F3 cells infected with an empty vector cells growing in the presence of IL-3 served as a negative control.

Western blotting and antibodies

For short-term stimulation of the cells with ERBB ligands, NIH-3T3 and COS-7 cells were starved overnight in DMEM + 0% FCS, and stimulated for 10 min with 0 or 50 ng/ml of EGF or NRG-1, prior to suspension in the lysis buffer. Cell were lysed and analyzed by Western blotting, as previously described (33), using the following primary antibodies: anti-EGFR (Santa Cruz Biotechnology, Cat#sc-03, discontinued), anti-ERBB2 (Thermo Scientific, Cat#MA-5-14057, RRID:AB_10977723), anti-ERBB3 (Cell Signaling Technology, Cat#4754, RRID:AB_10691324), anti-ERBB4 (E200, Abcam, Cat#ab32375, RRID:AB_731579), anti-phospho-EGFR Tyr1086 (Cell Signaling Technology, Cat#2220, RRID:AB_823485), anti-phospho-ERBB2 Tyr1221/1222 (Cell Signaling Technology, Cat#2243, RRID:AB_490899), anti-phospho-ERBB3 Tyr1289 (Cell Signaling Technology, Cat#4791, RRID:AB_2099709), anti-phospho-ERBB4 Tyr1284 (Cell Signaling Technology, Cat#4757, RRID:AB_2099987), anti-AKT (Cell Signaling Technology, Cat#2920, RRID:AB_1147620), anti-phospho-AKT (Cell Signaling Technology, Cat#4060, RRID:AB_2315049), anti-ERK (Cell Signaling Technology, Cat#9102, RRID:AB_330744), and anti-phospho-ERK (Cell Signaling Technology, Cat#9101, RRID:AB_331646). Loading was controlled using β-actin (Sigma Aldrich, Cat#A5441, RRID:AB_476744) or β-tubulin (Sigma-Aldrich, Cat# T7816, RRID:AB_261770) antibodies.

Co-immunoprecipitation

NIH-3T3 cells were starved overnight in serum-free medium and stimulated with 0 or 50 ng/ml EGF for 10 minutes. The cells were lysed in a volume of 400 µl. Lysates were pre-cleared by an incubation with 40 µl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for 1 hour at 4°C. Pre-cleared lysates containing 400-600 µg of total protein were incubated with 1/100 dilution of anti-CBL (Cell Signaling Technology, Cat# #2747, RRID: AB_2275284) overnight at 4 °C. Forty µl of beads were applied to the solution and incubation was continued for another 1 hour at 4 °C. Beads were then washed four times with lysis buffer and bound antibodies and co-precipitating material eluted by adding SDS–PAGE loading buffer and heating at 100 °C for 5 minutes. EGFR co-precipitating with anti-CBL was analyzed with Western blotting.

Structure preparation

The 2.25 Å resolution X-ray structure of the ERBB2 asymmetric homodimer was obtained from the Protein Data Bank (34) (PDB code 3PP0 (35)). In order to construct the ERBB2-EGFR (activator-receiver, from here onwards) and ERBB2-ERBB4 heterodimers, the homodimers were generated for the 2.8 Å resolution X-ray structure of EGFR (PDB code 2GS2 (36)) and 2.5 Å resolution structure of ERBB4 (PDB code 3BCE (37)) using symmetry operations and the program Chimera (38). Next, the ERBB2 activator kinase domain was superimposed with the activator kinase domain of the EGFR homodimer and the ERBB4 homodimer using Chimera (38); followed by substitution with the ERBB2 activator kinase in order to produce coordinates for models of the ERBB2-EGFR and ERBB2-ERBB4 heterodimers. The ERBB2 E936K mutant forms of ERBB2-ERBB2, ERBB2-EGFR and ERBB2-ERBB4 dimers were modeled by mutating E936 to K936 in Chimera (38).

Molecular dynamics simulations

The wild-type and E936K ERBB2-EGFR structures were prepared for molecular dynamics simulations (MDS) using the protein preparation wizard in Maestro (39), in which hydrogen atoms were added, protonation states for ionizable groups were determined and the proteins were energy minimized.

The resulting wild-type and mutant ERBB2-EGFR structures were used as starting structures for all-atom MDS using the Amber program (40) (version 18) and ff14SB protein force field (41). Initially, proteins were solvated with TIP3P water molecules (42) in an octahedral box, with 10 Å between solute atoms and the box surface. The systems were neutralized by adding Na^+ ions, with additional Na^+/Cl^- ions adjusting the salt concentration to 0.15 M.

The wild-type and mutant systems were first minimized, heated and equilibrated (see (43)); production simulation was then performed in an isothermal (300 K) – isobaric (1 bar) ensemble for 350 ns. The resulting trajectories were analyzed using Cpptraj (44) and VMD (45). Hydrogen bonds were defined as a bond distance $\leq$ 3.5 Å and a bond angle $\geq$ 135°. Root-mean-square deviations (RMSD, backbone atoms) were computed with respect to the initial frame structure. Pairwise interaction energies between residues and the free energy of binding between the ERBB2 and EGFR kinase monomers in the wild-type and E936K heterodimers were carried out using the MMPBSA.py module (46) available in Amber.

Results

Generation of a collective database covering 296 unique human cancer cell lines harboring 412 different *ERBB* mutations

Data were collected from the three databases: CCLE, GDSC, and CTRP. Collectively, these databases include information about 1460 different human cancer cell lines (Fig. 1A; Supplementary Table S1) with sequence information available from 1356 unique cell lines, and drug sensitivity data for 24 (CCLE), 265 (GDSC), and 481 (CTRP) different pharmacological compounds.

Altogether 296 unique cell lines harboring one or more ERBB alteration were identified, and 997 cell lines were determined as ERBB wild-type (Fig. 1A). The total number of unique coding sequence alterations in the ERBB genes was 412 (348 missense, 25 nonsense, 32 frameshift, and 7 deletion, insertion or splice variant alterations), with 90 out of the 296 ERBB mutant cell lines (30.7%) containing more than one ERBB alteration. In addition to ERBB coding sequence mutations, CCLE data were used to annotate cell lines with copy number alterations for any of the four *ERBB* genes. The coding sequence variations for each ERBB mutant cell lines, as well as copy number information for each ERBB wild-type and mutant cell line, are listed in Supplementary Table S4.

Normalizing the ERBB TKI sensitivity data between the databases

The three databases covered data for 10 different ERBB-selective TKIs (Fig. 1A). The 10 TKIs included the EGFR selective TKIs erlotinib, gefitinib, PD153035 and pelitinib, the EGFR selective TKIs WZ4002 and WZ8040 with activity also against the resistance mutation T790M, the dual EGFR/ERBB2-selective TKI lapatinib, and the pan-ERBB inhibitors afatinib, canertinib, and neratinib.

AUC was selected as a measure of drug sensitivity as it was available in all three databases, and as it serves as a comprehensive readout covering a range of drug concentrations. To normalize the drug sensitivity data from all three databases, a relative AUC (rAUC) value was determined for each cell line/drug/database combination. This was accomplished by determining a database-specific AUC reference level for positive and negative drug responses. The reference value for a positive response was determined as the mean response (as AUC) of all cells harboring the known EGFR driver mutations – L858R or exon 19 deletions – to the EGFR-selective TKIs erlotinib, gefitinib, and PD153035 in the given database. The reference value for a negative response was determined as the mean response (as AUC) of all the ERBB wild-type cell lines to all the ERBB TKIs analyzed in the given database. The rAUC value for each cell line/drug/database combination was then generated (Fig. 1B) using the formula:

$$rAUC = \frac{AUC - reference\ value\ for\ negative\ response}{reference\ value\ for\ positive\ response - reference\ value\ for\ negative\ response}$$

Consequently, rAUC values over 1 indicated a response greater than that associated with the known EGFR driver mutations to EGFR TKIs, rAUC values between 0 and 1 indicated a response smaller than that associated with the EGFR driver mutations to EGFR TKIs but greater than that associated with ERBB wild-type cells to any ERBB TKI, and rAUC values smaller than 0 indicated a response smaller than that associated with ERBB wild-type cells to any ERBB TKI.

EC50 was used as another variable for analysis of drug response to control for high rAUC values associating with micromolar EC50s that confer a risk for off-target effects or may not be clinically relevant (Fig. 1B). In order to obtain comparable EC50 values, log-logistic modeling was applied to the drug response data, and EC50 values were determined

from the resulting fitted dose-response curves. These modeled EC50 (mEC50) values were calculated for each compound that demonstrated an absolute effect of at least 10%. The resulting normalized dataset contained 4685 unique database, cell line, gene, mutation, and drug combinations (Supplementary Table S2).

Identification of mutations potentially predictive for ERBB TKIs

Potential predictive mutations were identified from the dataset as those associated with a rAUC of ≥ 1.0 and a mEC50 value of ≤ 1.0 μM for a given drug (Fig. 1C; Data are also available in an interactive HTML format in Supplementary Fig. S1). These cut-offs resulted in identification of 26 EGFR, 14 ERBB2, 14 ERBB3, and 22 different ERBB4 mutations in a total of 43 different cell lines (Fig. 2A and B; Supplementary Fig. S2; Supplementary Table S4). Sixty-two of the 76 mutations were missense, 3 nonsense mutations, 8 frame-shift alterations and 3 deletions or insertions.

ERBB amplification as a factor associating with TKI sensitivity

To address the association of *ERBB* amplification with ERBB TKI sensitivity, cell lines with amplification of any of the four *ERBB* genes but wild-type for ERBB coding sequence mutations, were plotted against the rAUC and mEC50 values as indicators of TKI sensitivity similar to Fig. 1C (Supplementary Fig. S3; Data are also available in an interactive HTML format in Supplementary Fig. S4). The analysis indicated that few cancer cell lines harboring *ERBB3* or *ERBB4* amplification were associated with enhanced sensitivity to the drugs. Moreover, TKIs with selectivity towards EGFR were mostly ineffective regardless of the *ERBB* amplification status. However, a number of cell lines with *EGFR* or *ERBB2* amplification (n = 14 for *EGFR*; n = 18 for *ERBB2*) were sensitive to the pan-ERBB TKI afatinib. These cell lines were mostly from contexts of known sensitivity to ERBB-targeted

therapies, including *ERBB2*-amplified breast, gastric and lung cancer, as well as head and neck and esophageal cancer with amplifications in *EGFR* and/or *ERBB2* (Supplementary Fig. S3).

Of all 114 cell lines wild-type for ERBB mutations but still demonstrating sensitivity to at least one ERBB TKI with rAUC $\geq$ 1.0, 45 (39.5%) harbored copy number alterations in at least one *ERBB* gene (Supplementary Fig. S5). This suggests that factors other than *ERBB* copy number variations may explain a significant proportion of the variability in the responsiveness of cancer cell lines to different ERBB-targeting TKIs in the absence of *ERBB* mutations.

Previously reported and potentially novel activating mutations

Out of the 76 unique mutations associated with increased sensitivity to ERBB TKIs, 14 (18.4%) were defined as "oncogenic" according to cBioPortal (searched June 4, 2020) (Supplementary Table S5). Of note, these 14 mutations constituted 53.8% of the 26 "oncogenic" *ERBB* mutations annotated by cBioPortal, supporting validity of the screen. The remaining 12 oncogenic mutations annotated by cBioPortal did not consistently associate with sensitivity to any ERBB TKI when endogenously expressed in the cancer cell lines (Supplementary Fig. S6). The remaining 62 out of the 76 (81.6%) unique ERBB mutations were regarded as potentially novel activating mutations. These included 18 *EGFR*, 10 *ERBB2*, 12 *ERBB3*, and 22 unique *ERBB4* alterations (Fig. 2A).

Interestingly 17 out of 34 (50.0%) of the cell lines with putative novel activating ERBB mutations harbored more than one ERBB mutation and 8 of these cell lines had an ERBB mutation combined with chromosomal gain in an *ERBB* gene (Fig. 2B). These observations imply that the percentage of functionally relevant novel mutations could be

exaggerated in a screen simply relying on *in silico* analysis, as passenger mutation could be present in the same cell line with an oncogenic mutation.

Functional validation of single ERBB mutants

To address the functional contribution of individual ERBB mutations associated with high rAUC values, 11 ERBB mutations were selected for experimental validation *in vitro* (Fig. 2). The selected mutations included one mutation in EGFR (Y1069C), two in ERBB2 (L720P and E936K), two in ERBB3 (E928fs*16 and E952Q), and four in ERBB4 (A17V, L780P, G863E, and G936R) that were all except ERBB4 A17V identified in a context of multiple *ERBB* aberrations (Fig. 2A). Two previously established oncogenic mutations, EGFR G719S and ERBB2 S310F, served as positive controls. Sensitivity of the selected mutations and known EGFR driver mutations to individual ERBB TKIs are shown in Supplementary Fig. S7.

To analyze the activity of the receptor variants in a cellular background devoid of significant endogenous ERBB expression, the IL-3-dependent Ba/F3 cells (7,47) were transduced with retroviral vectors encoding the ERBB variants or corresponding wild-type receptors. As expected, the positive controls EGFR G719S and ERBB2 S310F promoted IL-3-independent Ba/F3 cell growth in the absence of an activating ligand, unlike the respective wild-type controls (Fig. 3A and B). In addition, EGFR Y1069C and ERBB2 E936K were capable of supporting growth in the absence of both IL-3 and an ERBB ligand, indicating transforming potential (Fig. 3A and B). However, none of the ERBB3 variants promoted IL-3-independent growth as homodimers (Fig. 3C) or enhanced the transforming potential when overexpressed together with ERBB2 (Supplementary Fig. 8A). Similarly, none of the ERBB4 variants showed enhanced transforming potential. ERBB4 L780P and G863E were actually

less potent than wild-type ERBB4 in promoting IL-3-independent Ba/F3 cell survival in the presence of the NRG-1 ligand (Fig. 3D).

EGFR Y1069C demonstrates enhanced phosphorylation and reduced association with the ubiquitin ligase c-CBL

To address the mechanisms by which the EGFR Y1069C variant promotes growth, the phosphorylation status of EGFR variants was analyzed. Western analysis of Ba/F3 cells overexpressing EGFR Y1069C or G719S indicated phosphorylation of both variants at Tyr 1086 to a greater extent when compared to the wild-type EGFR, when the cells were cultured in the absence of IL-3 but in the presence of the EGF ligand (Fig. 4A). Moreover, both of the variants were phosphorylated also when the Ba/F3 cells were cultured in the absence of both IL-3 and of EGF, a condition under which the cells expressing wild-type EGFR did not survive (Fig. 3A). Consistently, both variants stimulated downstream EGFR signaling (phospho-AKT and phospho-ERK) to a greater extent than the wild-type receptor in the presence of EGF and were able to sustain downstream signaling also in the absence of EGF (Fig. 4A). When the EGFR Y1069C and G719S variants were overexpressed in NIH-3T3 cells, the variants again demonstrated enhanced activity although the effects were less prominent as compared to the Ba/F3 cell background (Fig. 4B).

Tyrosine 1069 (Tyr 1045 if excluding the signal sequence) of EGFR has been shown to serve as a binding site for the ubiquitin ligase c-CBL (48). The binding of c-CBL to EGFR enables the c-CBL-mediated ubiquitination and degradation of EGFR negatively controlling EGFR signaling output (49). Hence, co-immunoprecipitation assays were carried out in NIH-3T3 cells in order to test whether the Y1069C mutation disrupts the binding of c-CBL to EGFR. As expected, the Y1069C mutation almost completely abolished the interaction

between EGFR and c-CBL (Fig. 4C), suggesting that the increased activity of the Y1069C mutant is associated with loss of c-CBL binding.

Phosphorylation of ERBB2, ERBB3 and ERBB4 variants

When overexpressed in Ba/F3 cells, ERBB2 E936K demonstrated enhanced tyrosine phosphorylation when compared to cells expressing wild-type ERBB2 (Fig. 5A). Enhanced phosphorylation was observed in cells maintained in the presence of the ligand NRG-1, as well as in cells emerged as clones independent of both NRG-1 and IL-3 for survival and growth, similar to cells expressing the known oncogenic ERBB2 S310F (Fig. 5A). In contrast, expression of the mutant versions of other NRG-1 receptors, ERBB3 or ERBB4, did not induce receptor phosphorylation when compared to respective wild-type controls (Supplementary Fig. S8B and S9A). Consistently with the Ba/F3 transformation assays (Fig. 3D), ERBB4 L780P and ERBB4 G863E actually demonstrated reduced or no kinase activity in phospho-Western analyses of Ba/F3 or NIH-3T3 cells (Supplementary Fig. S9A and B). Similarly, ERBB3 E928fs with truncated kinase domain and ERBB3 E952Q, that failed to promote IL-3-independent Ba/F3 growth even when expressed together with ERBB2, showed reduced or no transactivation of ERBB2 when expressed in a context of ERBB2/ERBB3 heterodimers (Supplementary Fig. S8).

While the ERBB2 S310F variant effectively promoted pAKT and pERK signaling, associated with an induction of endogenous ERBB3 in Ba/F3 cells, these downstream signaling responses were modest in cells expressing ERBB2 E936K (Fig. 5A). The E936K variant also promoted ERBB2 autophosphorylation when overexpressed in COS-7 cells but did not demonstrate any additional activity when overexpressed together with ERBB3 in the context of ERBB2-ERBB3 heterodimers (Fig. 5B).

ERBB2 E936K promotes transphosphorylation by increasing the activity of ERBB2 heterodimers

In order to understand the activation mechanism of the ERBB2 E936K mutation, we performed structural analysis and molecular dynamics simulation (MDS). Glutamate E936, located on the αG helix of the kinase domain is conserved among ERBB kinases. In the wild-type ERBB2 homodimer structure (PDB code 3PP0 (35)), E936 (activator kinase) forms part of the dimer interface, facing residues from the β4-β5 loop and the C-terminal end of the juxtamembrane-B segment of the receiver kinase (Fig. 5C). These residues are conserved in EGFR and ERBB2, but variation in the β4-β5 loop is observed for ERBB3 (Glu, Gly-Ser) and ERBB4 (Glu, Ser-Pro) in comparison to EGFR and ERBB2 (Glu, Thr-Ser) (Fig. 5C).

In the wild-type ERBB2-ERBB2 structure and ERBB2-EGFR model, prior to MDS, the side-chain polar atoms of E936 of the ERBB2 activator kinase can form an interdimer hydrogen bond with the backbone nitrogen atom and side-chain hydroxyl group of S792/S784 in the ERBB2/EGFR receiver kinase (Fig. 5D). In contrast, with the E936K mutant, the ε-amino group of K936 could form a hydrogen bond with the hydroxyl group of S792/S784 and, a potentially stronger salt bridge with E717/E709 of the ERBB2/EGFR receiver kinase (Fig. 5D), absent in the wild-type dimers. The E936K mutant would strengthen the interaction between the asymmetric dimers, likely prolonging the duration of the activated state. Similarly, ERBB2-ERBB4 heterodimer model shows that E715 and S789 from the ERBB4 receiver kinase can respectively form similar ionic and hydrogen bond interactions with E936K in the mutant activator ERBB2 kinase (Fig. 5E). The effect of the E936K mutation might be negligible in the heterodimer formed between ERBB2 and ERBB3 since the pseudokinase ERBB3 serves predominantly as an activator due to its weak kinase activity (50,51), which is consistent with our results (Fig. 5B). In contrast, since the EGFR kinase domain strongly prefers to be the receiver kinase domain among the ERBBs (51),

followed by ERBB2, the E936K mutation should have a more pronounced effect on the ERBB2-EGFR dimer.

A 350 ns simulation of the wild-type and E936K mutant ERBB2-EGFR heterodimers recorded that both wild-type E936 (4.2%) and mutant E936K (9.2%) hydrogen bond with S784 (β4-β5 loop) of the receiver EGFR domain. The β4-β5 loop in the mutant receiver kinase appears slightly more stable and less variable (RMSD 0.26 +/- 0.06 Å) compared to the wild-type complex (RMSD 0.40 +/- 0.18 Å) (Fig. 5F) and may reflect the more frequent interactions observed between S784 and E936K: the interaction energy for E936K-E707 in the mutant heterodimer represents a stronger interaction (ΔG = -0.4 +/- 0.3 kcal/mol) relative to the wild-type (ΔG = -0.03 +/- 0.15 kcal/mol). The free energy of binding also indicates a stronger interaction between the monomers of the mutant complex (ΔG of binding -77.1 +/- 8.2 kcal/mol) in contrast to the wild-type complex (-66.6 +/- 11.1 kcal/mol) (Fig. 5G). The structural analysis suggests that the E936K mutation at the ERBB2 C-lobe would lead to stronger interactions at the dimer interface in the ERBB2-EGFR heterodimer and ERBB2-ERBB2 homodimer, resulting from increased hydrogen bonding and unique ionic interactions between the monomers.

To experimentally test whether ERBB2 E936K mutant could serve as a more potent activator kinase in ERBB2/EGFR heterodimers, ERBB2 wild-type or E936K were transfected together with wild-type EGFR to COS-7 cells and stimulated or not with EGF. Indeed, ERBB2 phosphorylation was increased in cells expressing both wild-type EGFR and ERBB2 E936K as compared to cells solely expressing wild-type receptors (Fig. 5H). Quantification of four replicate experiments showed that the expression of EGFR and ERBB2 E936K heterodimers produced two times higher phospho-ERBB2 levels than the heterodimers of wild-type receptors (P = 0.026) (Fig. 5I). These findings indicate that the E936K mutant activated EGFR more effectively than the wild-type.

EGFR Y1069C and ERBB2 E936K are sensitive to ERBB TKIs

To address the sensitivity of EGFR Y1069C and ERBB2 E936K to clinically used ERBB TKIs, Ba/F3 cells expressing the mutant or wild-type receptors were cultured for 72 hours in the presence of increasing concentrations of erlotinib, lapatinib, afatinib or neratinib. Cells expressing EGFR G719S and ERBB2 S310F were included in the experiments as internal positive controls. As all the four mutant variants were able to promote IL3-independent Ba/F3 cell growth both in the presence or absence of an activating ligand (Fig. 3A and B), the drug sensitivity analyses were carried out both in the absence and presence of EGF or NRG-1. Cells infected with an empty vector and cultured in the presence of IL-3 served as a control reflecting off-target toxicity.

Both EGFR Y1069C and ERBB2 E936K demonstrated sensitivity to ERBB TKIs (Fig. 6). EGFR Y1069C was most sensitive to the irreversible ERBB inhibitors afatinib and neratinib with IC50 values at the low nanomolar range, with the cells cultured in the absence of the ligand generally sensitive to concentrations about one order of magnitude smaller as compared to cells cultured in the presence of EGF. The IC50 for afatinib was significantly smaller in cells expressing EGFR mutants as compared to cells expressing wild-type EGFR both in the absence and presence of EGF, whereas the sensitivity to neratinib was significantly enhanced only in the absence of the ligand (Fig. 6A). Cells expressing ERBB2 E936K were also sensitive at low or even subnanomolar concentrations to both afatinib and neratinib. In contrast to EGFR Y1069C, cells expressing ERBB2 E936K were significantly more sensitive than cells expressing wild-type ERBB2 only for the response to afatinib in the presence of the NRG-1 ligand (Fig. 6B). The ERBB TKI sensitivity of cell expressing ERBB4 A17V or ERBB4 G936R did not significantly differ from the sensitivity of cells expressing wild-type ERBB4 (Supplementary Fig. 9C). Taken together, these observations

validate the identification of previously unknown actionable mutations with the approach based on oncogene-targeted database screen.

Discussion

Here, we leveraged the existing genomic and drug sensitivity data in publicly available cell line databases to identify novel predictive *ERBB* mutations for ERBB TKI sensitivity across cancers. We followed a hypothesis that cell lines harboring an ERBB mutation conferring sensitivity to ERBB TKIs would stand out as outliers in the drug response data among cell lines harboring passenger ERBB mutations. By combining and normalizing drug response data from CCLE, CTRP and GDSC, we were able to identify 43 "exceptional responder" cell lines out of 296 cell lines harboring *ERBB* mutations, with 62 novel, potentially predictive *ERBB* mutations. Experimentally characterizing 11 of these mutations we discovered EGFR Y1069C and ERBB2 E936K as novel activating ERBB mutations conferring sensitivity to ERBB TKIs. In addition to the colorectal cancer cell line CCK81 (CCLE), EGFR Y1069C variant has previously been reported in a clinical sample representing cholangiocarcinoma (2). The ERBB E936K variant has only been reported from the cell line CTV-1 derived from a leukemia patient (CCLE).

Our analysis also identified 14 known oncogenic ERBB variants, which are mostly located in recurrent hot-spots. In contrast, the novel mutations associated with increased sensitivity to ERBB TKIs are clinically relatively rare and were dispersed throughout the ERBB receptors. Thus, our results strengthen the argument that the "long tail" of non-recurrent cancer-associated mutations may also be of functional relevance (5). This conclusion is in line with our recent analysis of 7,216 EGFR mutations generated by random mutagenesis that identified transforming potential in 21 EGFR single nucleotide variants at mostly non-recurrent positions (7).

A significant proportion of cell lines (50.0%) demonstrating ERBB TKI sensitivity and harboring novel *ERBB* mutations, also had one or more co-occurring mutations or

amplifications in other *ERBB* genes. This could potentially represent co-operation of oncogenic genomic alterations to promote cancer cell growth, or a single oncogenic alteration accompanied by benign passenger alterations. A recent pan-cancer analysis of multiple mutations provided evidence supporting the co-operation hypothesis. Saito et al. reported that mutations of low functional activity may co-operate with other mutations occurring *in cis* within the same oncogene to promote enhanced oncogenic signaling (52). These observations may also explain how individually rare mutation events – like most of the ERBB mutations analyzed here – participate in a relatively common co-operative phenomenon that drives tumor growth. Of note, ERBB receptors are among the oncogenes most frequently harboring multiple somatic mutations in clinical cancer samples (52).

Our experimental characterization included ERBB mutations in two cell lines where several ERBB mutations co-occurred (CCK81 and CTV-1; Fig. 2B). In these cell lines, only one *ERBB* mutation was found to be activating and transforming (EGFR Y1069C in CCK81; ERBB2 E936K in CTV-1), suggesting a mechanism where other ERBB mutations are passengers. On the other hand, two ERBB TKI-sensitive *ERBB2*-amplified breast cancer cell lines both harbored the same concomitant point mutation in ERBB3 (E952Q). This is an intriguing observation considering the importance of ERBB3-mediated signaling for the growth of ERBB2-amplified breast cancer cells (53). While in our analyses in the Ba/F3 system, the ERBB3 E952Q mutation was not more potent than wild-type ERBB3 in promoting growth or ERBB2 phosphorylation when co-expressed with ERBB2 (Supplementary Fig. S8), the variant has previously been reported to moderately enhance ERBB3 signaling in CHO cells (54). This difference may be due to the presence of different heterodimerization partners or other cell context-dependent signaling characteristics that are sensitive to a mutation at a residue potentially critical for interactions of ERBB3 at the kinase interface (55).

Another interesting observation of co-occurring genetic events was the unexpectedly high (20.9%; Fig. 2B) frequency of *ERBB4* deletions in TKI-sensitive ERBB-mutant cell lines. In addition, two of the four functionally studied ERBB4 variants, L780P and G863E, were virtually kinase dead, and completely failed to promote Ba/F3 cell growth (Fig. 3D and Supplementary Fig. 9). Structurally, these observations are consistent, in that L780 in wild-type ERBB4 is involved in hydrophobic interactions with residues from the αC and αE helices and F862 of the DFG motif, interactions key to stabilizing the active state conformation. The L780P mutation could compromise the stability and drive the transition from the active towards the inactive state of ERBB4. Similarly, the ERBB4 G863E mutation, located at the DFG motif critical for phosphotransfer in the kinase activation loop (56,57), would cause a disruption of the active state αC helix conformation due to steric clashes and charge-charge repulsion. These observations, together with previous reports of both activating (22,23) and kinase-impaired (57) ERBB4 mutations contributing to malignant growth, justify future addressing of the functional role of *ERBB4* gain- and loss-of function in the context of multiple ERBB alterations.

Putative limitations of the study include the obvious bias of missing functional ERBB variants as they were either not included in the cell line collections analyzed, or as their effects are masked by other concomitant oncogenic events in the same cell line. However, all 26 ERBB mutations defined as “oncogenic” by cBioPortal were represented by at least one cell line. Moreover, none of the 12 mutations that were listed as “oncogenic” by cBioPortal, but were not defined activating in our analysis, did show significant sensitivity to any of the tested ERBB inhibitors in a dose-response analysis (Supplementary Fig. S6). It is of note, however, that for *e.g.* EGFR L861Q the small number of cell lines available precluded any conclusions about the activity of the variant. The annotation of the cell lines as amplified for any of the *ERBB* genes was also based on crude categorization to cells harboring *ERBB*

amplification or not, and did not allow for quantitative analysis of the influence of the gain in gene copy numbers for the observed effect.

Taken together, our findings indicate that activating and potentially predictive ERBB mutation can be found by a systematic screen of mutations outside of the most prominent mutation hotspots. The data further suggests that the ERBB variants with the greatest functional significance frequently co-occur with other *ERBB* changes.

Disclosure of Potential Conflicts of Interest

K.E. and K.J.K. have a research agreement with Puma Biotechnology. K.E. has ownership interest in Abomics, Novo Nordisk, Orion, and Roche.

Authors' Contributions

Conception and design: M. K. A. Koivu, K. J. Kurppa, K. Elenius

Development of methodology: M. K. A. Koivu, D. Chakroborty

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. K. A. Koivu, D. Chakroborty, M. Z. Tamirat, M. S. Johnson, K. J. Kurppa, K. Elenius

Writing, review, and/or revision of the manuscript: M. K. A. Koivu, D. Chakroborty, M. Z. Tamirat, M. S. Johnson, K. J. Kurppa, K. Elenius

Study supervision: K. Elenius

Acknowledgements

We thank CSC IT Center for Science (project 2002814, and computing infrastructure); Dr. Jukka Lehtonen (Biocenter Finland Bioinformatics Network) for scientific IT support. The Structural Bioinformatics Laboratory is part of the university's Drug Discovery and Diagnostics strategic area. This study has been supported by Academy of Finland (274728, 316796) (to K. Elenius), the Cancer Foundation of Finland (to K. Elenius), Turku University Central Hospital (to K. Elenius), Doctoral Network in Informational and Structural Biology (Graduate School of Åbo Akademi University) (to M. Z. Tamirat), Sigrid Juselius Foundation (to K. J. Kurppa; and to M. S. Johnson), NordForsk Nordic POP (to M. S. Johnson), K. Albin Johansson Foundation (to K. J. Kurppa; and to D. Chakroborty), Juhani Aho Foundation for Medical Research (to D. Chakroborty), University of Turku Graduate School (to D.

Chakroborty), Instrumentarium Science Foundation (to K. J. Kurppa), Finnish Cultural Foundation (Varsinais-Suomi Regional Fund 85171988 (to M. K. A. Koivu); Central Fund 00200590 (to K. J. Kurppa)) and supercomputing from CSC IT Center for Science (to M. S. Johnson; and to M. Z. Tamirat).

References

1. Lawrence MS, Stojanov P, Polak P, Kryukov G V., Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;**499**:214–18.
2. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017;**23**:703–13.
3. Hyman DM, Taylor BS, Baselga J. Implementing genome-driven oncology. Cell 2017;**168**(4):584–99.
4. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;**458**(7239):719–24.
5. Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol 2016;**34**:155–63.
6. Garraway LA, Lander ES. Lessons from the cancer genome. Cell 2013;**153**(1):17–37.
7. Chakroborty D, Kurppa KJ, Paatero I, Ojala VK, Koivu M, Tamirat MZ, et al. An unbiased in vitro screen for activating epidermal growth factor receptor mutations. J Biol Chem 2019;**294**:9377–89.
8. Schlessinger J. Signal transduction by allosteric receptor oligomerization. Trends Biochem Sci 1988;**13**:443–7.
9. Lemmon MA, Schlessinger J. Regulation of signal transduction and signal diversity by receptor oligomerization. Trends Biochem Sci 1994;**19**:459–63.
10. Merilahti JAM, Elenius K. Gamma-secretase-dependent signaling of receptor tyrosine kinases. Oncogene 2019;**38**(2):151–63.

11. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, Mcguire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 1987;**235**(4785):177–82.

12. Sugawa N, Ekstrand AJ, James CD, Collins VP. Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas. Proc Natl Acad Sci USA 1990;**87**:8602–6.

13. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non–small-cell lung cancer to Gefitinib. N Engl J Med 2004;**350**:2129–39.

14. Arteaga CL, Engelman JA. ERBB receptors: From oncogene discovery to basic science to mechanism-based cancer therapeutics. Cancer Cell 2014;**25**(3):282–303.

15. Jänne PA, Chih-Hsin Yang J, Kim DW, Planchard D, Ohe Y, Ramalingam SS, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med 2015;**372**:1689–99.

16. Sequist L V., Yang JCH, Yamamoto N, O'Byrne K, Hirsh V, Mok T, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol 2013;**31**:3327–34.

17. Chan A, Delaloge S, Holmes FA, Moy B, Iwata H, Harvey VJ, et al. Neratinib after trastuzumab-based adjuvant therapy in patients with HER2-positive breast cancer (ExteNET): A multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 2016;**17**:367–77.

18. Wu YL, Cheng Y, Zhou X, Lee KH, Nakagawa K, Niho S, et al. Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): a randomised, open-label, phase 3 trial. Lancet

Oncol 2017;**18**:1454–66.

19. Stephens P, Hunter C, Bignell G, Edkins S, Davies H, Teague J, et al. Lung cancer: Intragenic ERBB2 kinase mutations in tumours. Nature 2004;**431**:525–6.
20. Shigematsu H, Takahashi T, Nomura M, Majmudar K, Suzuki M, Lee H, et al. Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. Cancer Res 2005;**65**:1642–6.
21. Jaiswal BS, Kljavin NM, Stawiski EW, Chan E, Parikh C, Durinck S, et al. Oncogenic ERBB3 mutations in human cancers. Cancer Cell 2013;**23**:603–17.
22. Prickett TD, Agrawal NS, Wei X, Yates KE, Lin JC, Wunderlich JR, et al. Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. Nat Genet 2009;**41**:1127–32.
23. Kurppa KJ, Denessiouk K, Johnson MS, Elenius K. Activating ERBB4 mutations in non-small cell lung cancer. Oncogene 2016;**35**:1283–91.
24. Hyman DM, Piha-Paul SA, Won H, Rodon J, Saura C, Shapiro GI, et al. HER kinase inhibition in patients with HER2- and HER3-mutant cancers. Nature 2018;**554**:189–94.
25. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;**483**:603–307.
26. The Cancer Genome Atlas Program - National Cancer Institute. https://www.cancer.gov/tcga.
27. International Cancer Genome Consortium. https://www.icgc.org.
28. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 2013;**41**:955–61.

29. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price E V., Coletti ME, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov 2015;**5**:1210–23.

30. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2018.

31. Merilahti JAM, Ojala VK, Knittle AM, Pulliainen AT, Elenius K. Genome-wide screen of gamma-secretase-mediated intramembrane cleavage of receptor tyrosine kinases. Mol Biol Cell 2017;**28**:3123–31.

32. Määttä JA, Sundvall M, Junttila TT, Peri L, Laine VJO, Isola J, et al. Proteolytic cleavage and phosphorylation of a tumor-associated ErbB4 isoform promote ligand-independent survival and cancer cell growth. Mol Biol Cell 2006;**17**:67–79.

33. Ojala VK, Knittle AM, Kirjalainen P, Merilahti JAM, Kortesoja M, Tvorogov D, et al. The guanine nucleotide exchange factor VAV3 participates in ERBB4-mediated cancer cell migration. J Biol Chem 2020;**295**(33):11559–71.

34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;**28**(1):235–42.

35. Aertgeerts K, Skene R, Yano J, Sang BC, Zou H, Snell G, et al. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. J Biol Chem 2011;**286**:18756–65.

36. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 2006;**125**:1137–49.

37. Qiu C, Tarrant MK, Choi SH, Sathyamurthy A, Bose R, Banjade S, et al. Mechanism of activation and inhibition of the HER4/ErbB4 kinase. Structure 2008;**16**:460–7.

38. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al.

UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem 2004;**25**:1605–12.

39. Schrödinger Release 2019–1: Maestro, version Schrödinger LLC, New York, NY, 2019.

40. Case Ross C Walker DA, Darden Junmei Wang Robert Duke TE. Amber 2018 University of California, San Francisco.

41. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 2015;**11**:3696–713.

42. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;**79**:926–35.

43. Tamirat MZ, Kurppa KJ, Elenius K, Johnson MS. Deciphering the structural effects of activating EGFR somatic mutations with molecular dynamics simulation. J Vis Exp 2020;**159**:10.3791/61125.

44. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 2013;**9**:3084–95.

45. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph 1996;**14**:33–8.

46. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An efficient program for end-state free energy calculations. J Chem Theory Comput 2012;**8**:3314–21.

47. Warmuth M, Kim S, Gu X, Xia G, Adria F, Adrián F. Ba/F3 cells and their use in kinase drug discovery. Curr Opin Oncol 2007;**19**:55–60.

48. Levkowitz G, Waterman H, Ettenberg SA, Katz M, Tsygankov AY, Alroy I, et al. Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth

factor signaling by c-Cbl/Sli-1. Mol Cell 1999;**4**:1029–40.

49. Rubin C, Gur G, Yarden Y. Negative regulation of receptor tyrosine kinases: Unexpected links to c-Cbl and receptor ubiquitylation. Cell Res 2005;**15**(1):66–71.

50. Jura N, Shan Y, Cao X, Shaw DE, Kuriyan J. Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3. Proc Natl Acad Sci USA 2009;**106**:21608–13.

51. Ward MD, Leahy DJ. Kinase activator-receiver preference in ErbB heterodimers is determined by intracellular regions and is not coupled to extracellular asymmetry. J Biol Chem 2015;**290**(3):1570–9.

52. Saito Y, Koya J, Araki M, Kogure Y, Shingaki S, Tabata M, et al. Landscape and function of multiple mutations within individual oncogenes. Nature 2020;**582**:95–9.

53. Junttila TT, Akita RW, Parsons K, Fields C, Lewis Phillips GD, Friedman LS, et al. Ligand-independent HER2/HER3/PI3K complex is disrupted by Trastuzumab and is effectively inhibited by the PI3K inhibitor GDC-0941. Cancer Cell 2009;**15**:429–40.

54. Pryor MMC, Steinkamp MP, Halasz AM, Chen Y, Yang S, Smith MS, et al. Orchestration of ErbB3 signaling through heterointeractions and homointeractions. Mol Biol Cell 2015;**26**:4109–23.

55. Monsey J, Shen W, Schlesinger P, Bose R. Her4 and Her2/neu tyrosine kinase domains dimerize and activate in a reconstituted in vitro system. J Biol Chem 2010;**285**:7035–44.

56. Hanks SK, Quinn AM, Hunter T. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. Science 1988;**241**(4861):42–52.

57. Tvorogov D, Sundvall M, Kurppa K, Hollmén M, Repo S, Johnson MS, et al. Somatic mutations of ErbB4: Selective loss-of-function phenotype affecting signal transduction pathways in cancer. J Biol Chem 2009;**284**:5582–91.

Figure Legends

Figure 1. Screen for predictive ERBB mutations.

A, Distribution of cell line and ERBB TKI data available in the indicated databases. In total, the three databases included data of 1460 different cancer cell lines (of which 296 harbored one or several ERBB coding sequence variants) and ten different ERBB targeting tyrosine kinase inhibitors (TKI). **B,** Statistical approach. The top panel shows a schematic dose-response curve (dashed red line) with area under curve (AUC) indicated in blue and effective concentration 50 (EC50) in dashed black line. The bottom panel presents an example derived from afatinib data available in the CTRP database. Relative AUC (rAUC) was calculated based on areas between the dose-response curves to determine the effect of the cell line/drug pair of interest (red dashed area) relative to the effect achieved by the positive reference controls (blue area). **C,** A dot plot presentation of the full analysis of ERBB TKI sensitivity of cell lines with coding sequence ERBB alterations. With modeled EC50 (mEC50) in x-axis and rAUC in y-axis, dots representing sensitive cell lines are located in the left upper corners of the plots. Each dot represents a cell line in one database. The color of the dot indicates the tissue of origin of the cancer cell line.

Figure 2. ERBB mutations in cancer cell lines sensitive to TKIs.

A, ERBB mutations in cell lines demonstrating sensitivity to any ERBB TKI with rAUC ≥ 1 and mEC50 ≤ 1 are presented as lollipops indicating the position of the mutation in the primary sequences of the four ERBB receptors. Height of the lollipop indicates the number of different cell lines with the mutation. Arrows point to mutations selected for further analyses.

Cell lines harboring the mutations are in parentheses. Short lines indicate positions with more than one alteration in the same residue. Black color of the lollipop indicates a previously reported oncogenic mutation (n = 14), orange a novel potentially predictive mutation (n = 62). **B,** Cell lines harboring the mutations shown in A (n = 43). Green box indicates a mutation altering the coding sequence, red lining of the box an amplification, and blue lining a deletion of the indicated ERBB receptor. Black color of the name of the cell line indicates previously reported oncogenic mutations, orange potential novel predictive mutations. Cell lines harboring both previously known as well as potentially novel mutations are marked with black text followed by an orange asterisk. The cell lines are ranked by their highest rAUC values as demonstrated by the heatmap on the right.

Figure 3. Growth promoted by ERBB variants.

The indicated ERBB expression constructs were tested for their ability to promote Ba/F3 cell survival and proliferation in the presence of interleukin-3 (IL-3) (left panels), in the absence of IL-3 but in the presence of an activating ERBB ligand (EGF or NRG; middle panels), or in the absence of both IL-3 and an ERBB ligand (right panels). Viability of the cells was monitored using MTT assays. The mean and SD are shown. The analyses were repeated three times.

Figure 4. Signaling by EGFR variants.

A, Western analysis of EGFR signaling in Ba/F3 cells expressing the indicated EGFR constructs. The cells were cultured in the presence or absence of IL-3 and/or EGF as indicated. **B,** Western analysis of EGFR signaling in NIH-3T3 cells expressing the indicated EGFR constructs. The cells were stimulated for 10 minutes with 0 or 50 ng/ml of EGF. **C,** Co-immunoprecipitation analysis of the association of the indicated EGFR constructs with

CBL in NIH-3T3 cells. The cells were stimulated for 10 minutes with 0 or 50 ng/ml of EGF. The analyses were repeated 2 to 4 times.

Figure 5. Signaling by ERBB2 variants.

A, Western analysis of ERBB2 signaling in Ba/F3 cells expressing the indicated ERBB2 constructs. The cells were cultured in the presence or absence of IL-3 and/or NRG-1 as indicated. **B,** Western analysis of basal ERBB phosphorylation in COS-7 transiently overexpressing the indicated ERBB2 and ERBB3 constructs alone or as a heterodimeric complex. Images of single gels were cropped between lanes 6 and 7. The Western analyses were repeated 2 to 3 times. **C,** Structure of the ERBB2 kinase homodimer. Residue E936 is located within the αG helix in the kinase domains (shown for the activator kinase). Residues from the receiver kinase – in the vicinity of E936 – and residing within the juxtamembrane B segment and the β4-β5 loops, are also highlighted. Equivalent residues in the ERBB kinases are listed. **D-E**, ERBB2 E936/K936 interaction with EGFR and ERBB4 kinase. Interactions of residues from the EGFR (D) and ERBB4 (E) receiver kinase domains with the wild-type E936 (left) and mutant E936K (right) of the ERBB2 activator kinase. Hydrogen bond and ionic interactions are shown as black and red dotted lines, respectively. **F-G,** ERBB2-EGFR; stability of the β4-β5 loop and free energy of binding. **F,** RMSD of the β4-β5 loop in the wild-type (blue) and E936K mutant (red) of the ERBB2-EGFR complexes during 350 ns simulations. **G,** ΔG of monomer-monomer binding for wild-type (blue) and E936K (red) ERBB2-EGFR heterodimers. **H,** Western analysis of ERBB phosphorylation in COS-7 transiently overexpressing the indicated EGFR and ERBB2 constructs. The cells were stimulated with 0 or 50 ng/ml EGF for 10 minutes. **I,** Densitometric quantification of ERBB2 phosphorylation status from Western analyses such as shown in panel H. The mean and SD are shown. n = 4.

Figure 6. Sensitivity of EGFR Y1069C and ERBB2 E936K to ERBB TKIs.

Ba/F3 cells expressing the indicated EGFR (**A**) or ERBB2 (**B**) variants were cultured for 72 hours in the presence or absence of increasing concentrations of the ERBB TKIs erlotinib, lapatinib, afatinib, or neratinib. Cells expressing ERBB receptors were analyzed in the absence of IL-3 and in the absence or presence of an activating ligand, 10 ng/ml EGF or 20 ng/ml NRG-1. Vector control cells were cultured in the presence of IL-3. Cell viability was measured using MTT assays. The mean and SD are shown for the dose-response curves. IC50 values for the drug responses were calculated from three to six independent analyses after fitting the dose-response curves with four-parameter log-logistic function (LL.4; R). Data were statistically analyzed to determine the difference in IC50 in comparison to cells expressing the respective wild-type ERBB control using unpaired two-sample t-test. *P* values < 0.05 indicating enhanced sensitivity as compared to wild-type receptor are indicated in bold.

Fig. 1

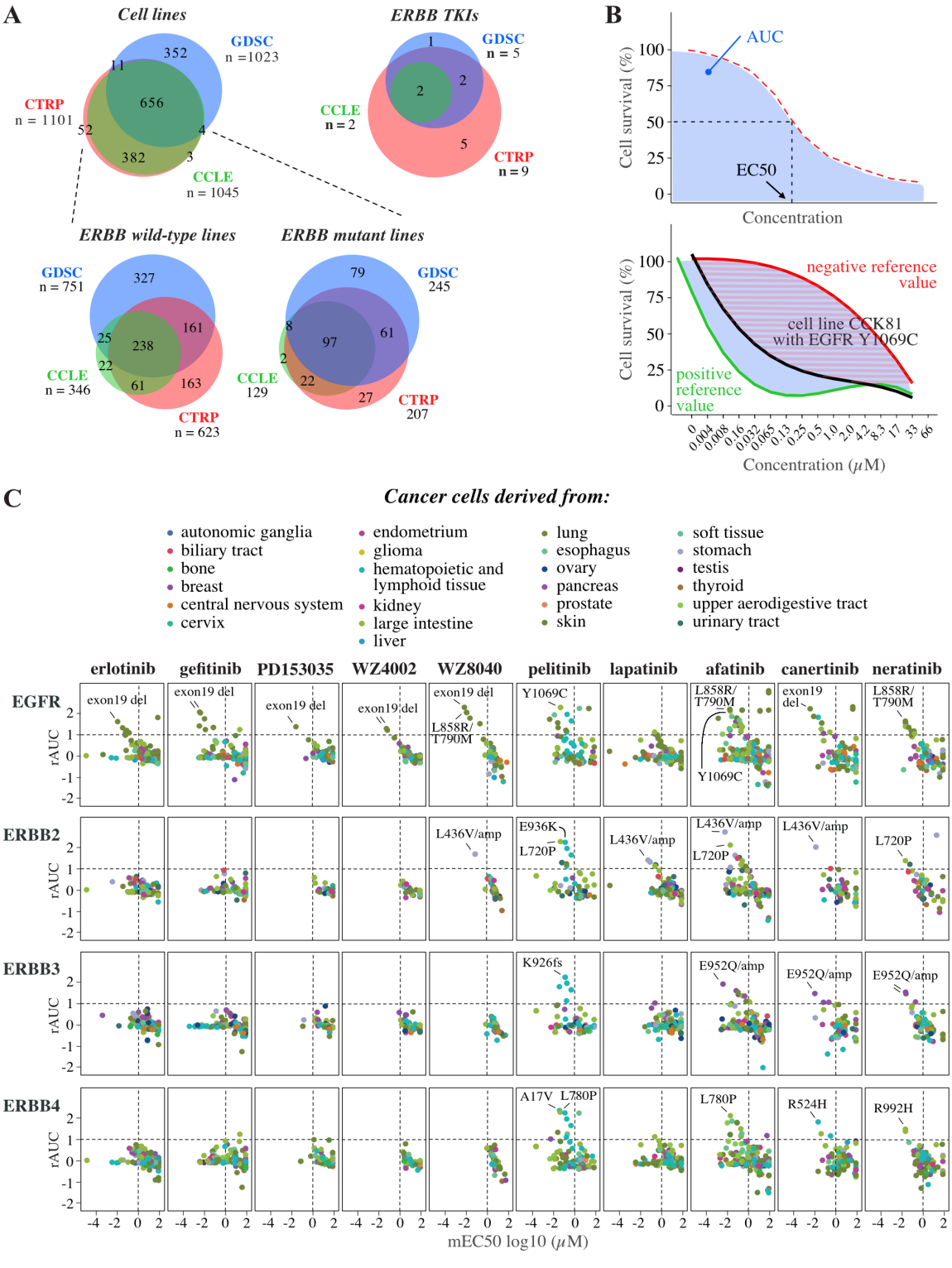


Fig. 2

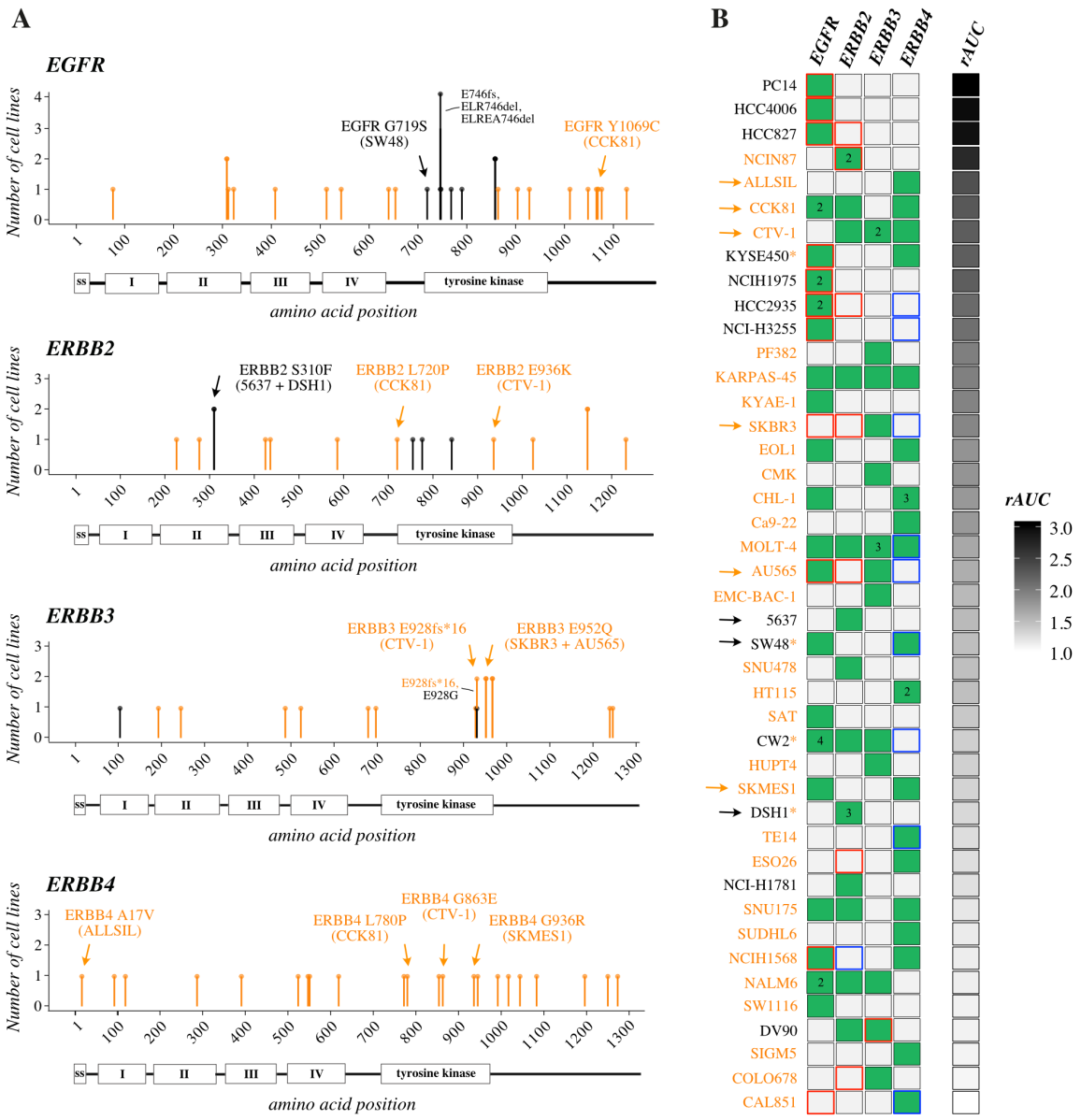


Fig. 3

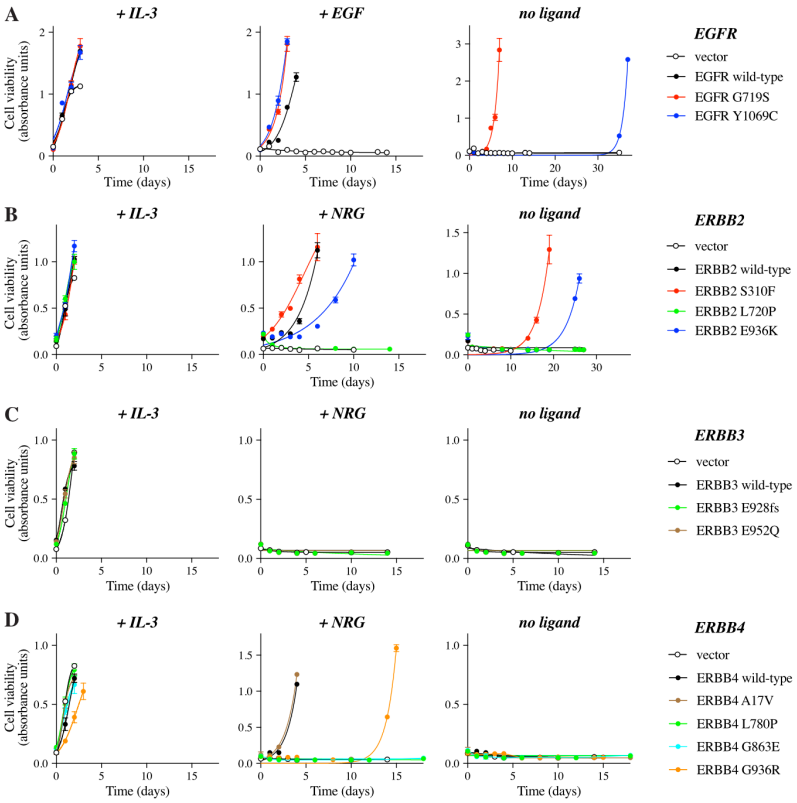


Fig. 4

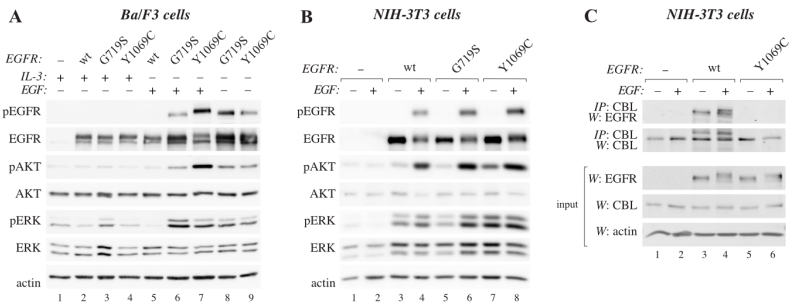


Fig. 5

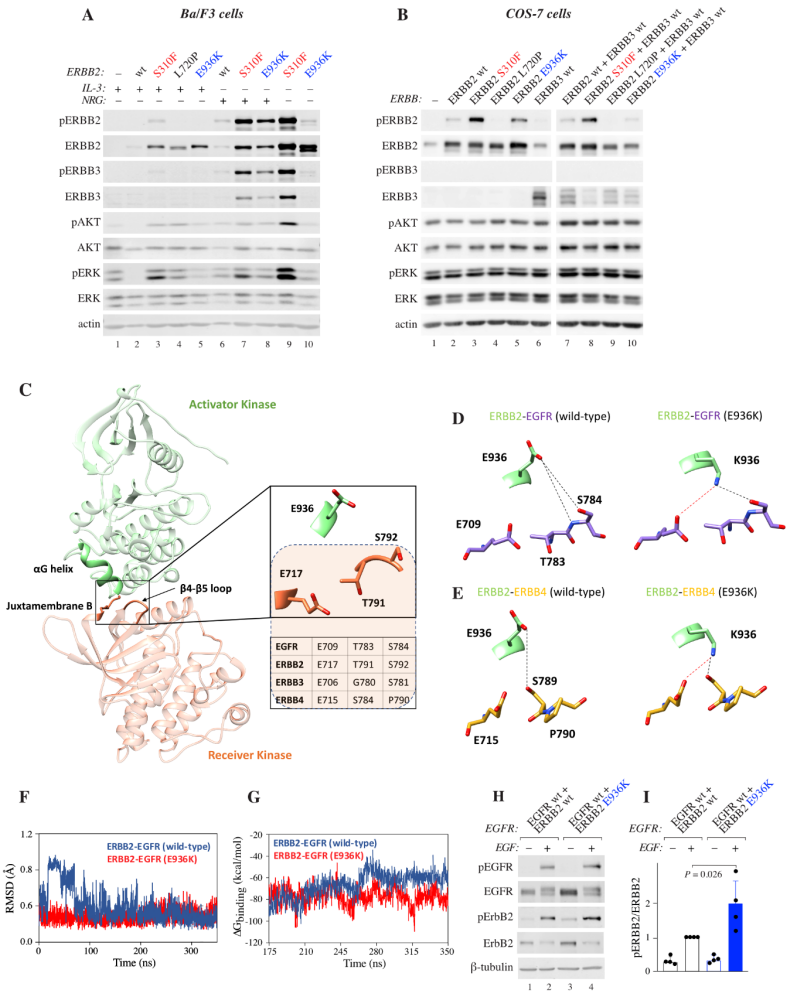
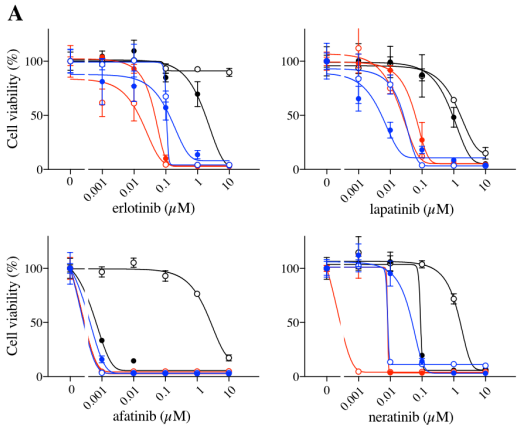


Fig. 6

A

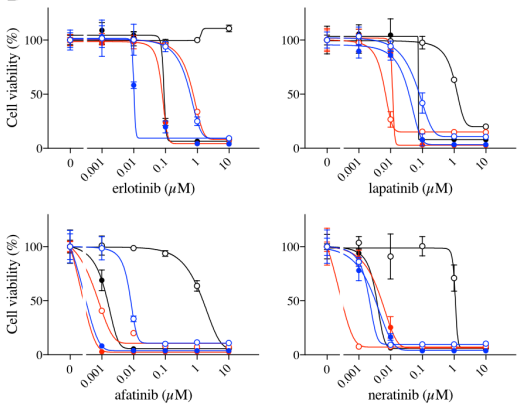


EGFR

- vector control
- wild-type + EGF
- G719S - EGF
- G719S + EGF
- Y1069C - EGF
- Y1069C + EGF

drug	variant ± ligand	IC50 ± SD (nM)	*P* (*vs.* wt)	*n*
	wild-type + EGF	2800 ± 3800		3
	G719S - EGF	17 ± 8.7	0.27	3
erlotinib	G719S + EGF	24 ± 12	0.27	3
	Y1069C - EGF	67 ± 39	0.10	6
	Y1069C + EGF	1600 ± 1400	0.58	4
	wild-type + EGF	2400 ± 1500		3
	G719S - EGF	34 ± 26	0.05	3
lapatinib	G719S + EGF	54 ± 44	0.05	3
	Y1069C - EGF	550 ± 750	0.13	3
	Y1069C + EGF	3700 ± 2700	0.53	3
	wild-type + EGF	6.0 ± 4.3		3
	G719S - EGF	0.051 ± 0.029	**0.03**	4
afatinib	G719S + EGF	0.042 ± 0.045	**0.01**	6
	Y1069C - EGF	0.12 ± 0.16	**0.01**	3
	Y1069C + EGF	1.5 ± 1.0	**0.03**	6
	wild-type + EGF	88 ± 22		3
	G719S - EGF	0.45 ± 0.96	**< 0.001**	3
neratinib	G719S + EGF	2.8 ± 1.8	**< 0.001**	6
	Y1069C - EGF	2.9 ± 1.8	**0.003**	4
	Y1069C + EGF	60 ± 27	0.17	6

B



ERBB2

- vector control
- wild-type + NRG
- S310F - NRG
- S310F + NRG
- E936K - NRG
- E936K + NRG

drug	variant ± ligand	IC50 ± SD (nM)	*P* (*vs.* wt)	*n*
	wild-type + NRG	29 ± 27		3
	S310F - NRG	3400 ± 3700	0.18	3
erlotinib	S310F + NRG	56 ± 24	0.26	3
	E936K - NRG	360 ± 99	0.005	3
	E936K + NRG	100 ± 124	0.39	3
	wild-type + NRG	31 ± 11		3
	S310F - NRG	4.5 ± 3.7	0.02	3
lapatinib	S310F + NRG	21 ± 2.3	0.19	3
	E936K - NRG	54 ± 38	0.36	3
	E936K + NRG	39 ± 13	0.44	3
	wild-type + NRG	1.0 ± 0.6		4
	S310F - NRG	1.0 ± 0.88	0.99	5
afatinib	S310F + NRG	0.045 ± 0.034	**0.05**	6
	E936K - NRG	3.6 ± 2.7	0.03	6
	E936K + NRG	0.042 ± 0.024	**0.02**	4
	wild-type + NRG	1.4 ± 1.1		6
	S310F - NRG	0.28 ± 0.50	**0.04**	6
neratinib	S310F + NRG	2.9 ± 2.4	0.20	6
	E936K - NRG	1.0 ± 0.54	0.44	6
	E936K + NRG	1.2 ± 0.96	0.71	6

AACR American Association for Cancer Research

Molecular Cancer Therapeutics

Identification of Predictive ERBB Mutations by Leveraging Publicly Available Cell Line Databases

Marika K. A. Koivu, Deepankar Chakroborty, Mahlet Z. Tamirat, et al.

Mol Cancer Ther Published OnlineFirst December 15, 2020.