EGFR Rearrangements as Oncogenic Drivers and Therapeutic Targets in Lung Cancer

By

Jean-Nicolas Gallant

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

May, 2017

Nashville, TN

Approved:

Christopher Williams, M.D., Ph.D.

Robert Coffey, M.D.

Rebecca Cook, Ph.D.

Sally York, M.D., Ph.D.

Christine Lovly, M.D., Ph.D.

To Jessica, without whom none of this would be possible

ACKNOWLEDGEMENTS

Science is a human endeavor, and the research described herein would not have been possible without the support of an incredible team of people. I am delivering this work today thanks to the help of a small town; if it takes a village to raise a child, then, surely, it must have taken a medium-sized township to get me here.

Logistically, none of this would have happened without the generosity of my sponsors and institution—thank you to Thermo Scientific, the AACR, the V-Foundation, the NIH / NCI, and Vanderbilt. I particularly need to thank the Vanderbilt MSTP for taking a chance on an everyday-normal-guy (me) with a knack for cancer biology. Thank you to the leadership team, past and present; to Drs. C. Williams, Grundy, Estrada, York, Winder, Dermody, Bills, Swift, M. Williams and, last but definitely not least, Melissa for your unending kindness, understanding, support, and flexibility over the years. This is the best MST program in the country—thanks to you. I'd be remiss not to thank Vanderbilt, as a whole, for creating a collegial atmosphere chock-full of resources, cores, and outsource providers that have made my time in lab all that much more productive and enjoyable.

The highly translational nature of this research would not have been possible without an outstanding team of collaborators with diverse expertise. Here, at Vanderbilt, I have to thank the groups with whom we held regular lab meetings and those individuals who shaped me and the science in this dissertation: thank you to the Pietenpol lab (especially Jennifer, Brian Lehmann, Tim Shaver, and Scott Beeler); the Quaranta lab (Vito and Darren Tyson); and the Structural Biology–Precision Medicine working group—the Arteaga lab (Carlos, Valerie Jansen, Monica Red-Brewer, Ariella Hanker, and Sarah Croessmann) and the Center for Structural Biology (Jens Meiler, Johnathan Sheehan, Tony Capra et. al). Thanks to all those with whom I otherwise interacted on the 6th floor of PRB (the best floor of the VICC); thank you for your time, input, and friendship. No thanks to the Balko lab and Mellissa Nixon for their constant chicaneries. The foundation [sic] of this work was data provided by the awesome folks at Foundation Medicine; special to Vincent Miller, Siraj Ali, Doron Lipson, Phil Stephens, Jeff Ross, Deborah Morosini, Sohail Balasubramanian, and Kyle Gowen for working with, supporting, and advocating for Dr. Lovly and I—it has been a pleasure and I hope to see you in Cambridge sometime soon. Finally, thank you to our clinical collaborators at Memorial Sloan Kettering Cancer Center (Marc Ladanyi, Mark Kris, and Raghu Chandramohan), Baylor Sammons Cancer Center (Kartik Konduri and Andrew Whiteley), Robert H. Lurie Comprehensive Cancer Center of Northwestern University (Young Kwang Chae and Francis Giles), the Addario Foundation (Barbara Gitlitz [USC] and Tiziana Vavalà [Torino]), Winship Cancer Institute of Emory (Taofeek Owinoko, Suresh Ramalingam, and Satyanarayan Reddy), and Abramson Cancer Center of the University of Pennsylvania (Vijay Peddareddigari and Beth Eaby-Sandy) for being willing to pool resources to maximize the impact of this research. I hope to have the pleasure of meeting you all in person down the road.

To my dissertation committee: thank you for your patience, understanding, and guidance. More so than a better scientist, you have made me a better person. I leaned on each of you during tough times, and you pushed back—helping me realize that I had the strength and will to keep going when I did not think I did. Special thanks to Dr. Williams for introducing me to the YAMCs, to Dr. Coffey for his bottomless experience with EGFR, to Dr. Cook for reminding me to strive for breadth vs. depth of scientific knowledge, to Dr. York for embodying the ideal warm/caring physician and cold/sharp scientist duality, and to Dr. Lovly for, well, everything. I very much appreciate the time each of you set aside for my scientific formation and can only wish to have spent more time with you.

I have been fortunate to have spent the past few years in a friendly lab environment interacting with some awesome people. Thank you to the remnants of the Pao lab; to Katie Hutchinson for normalizing late night lab, to Caroline Nebhan for keeping me happy and well fed, to Eiki Ichihara for always smiling (even if it may have been a misunderstanding), to Abudi Nashabi for staying one step ahead of every lab need, and to Catherine Meador for serving as a shining example of all that is right with the world. Thank you to Merrida Childress, the work wife, the ying to my yang, for keeping me honest, questioning my science, and making the past few years truly enjoyable. Thank you to Yingjun Yan, lab technician extraordinaire, for keeping everything in order, asking all the right questions, and always smiling—you are an inspiration. Thank you to Huan Qiao for teaching me everything that I know about molecular biology. To Karinna Almodovar: thanks for lending an ear, chatting, and keeping me sane. To David Westover: thanks for destroying any/all archetypes I had in my head by way of your hip-hop loving, socially conscious, OU-educated, sharp, small-scale self. To Zhenfang Du: thank you for picking up this work where I have left off and apologies for not spending more time over beverages these past few months. To Vincent Huang: I only ever make fun of you because your future is so

incredibly bright; keep going! To Gabrielle Hampton: keep working this hard and it will pay off. To all others that have come in and out of lab: thank you for brightening my life.

I could not have spent all this time in lab without the support, perspective, and refreshment provided by my family and friends. Thanks to the CBSA, MSTPals, and Handleton crew for pulling me out of lab at all the right times. Thank you to Mom, Dad, and Aunt Wil for your understanding and for taking care of Jessica and Brigitte during all the best and the worst times. Thank you to Mamour et Grandpapa for your blind support; to Mamie for the prières; and Josée for encouragement. PaPa: thanks for leading the way and showing me what is important and what is possible. Etienne: thanks for putting the heat on from behind and making sure I do my best; no doubt you'll fly ahead in the coming years. Maman: thanks for hanging in there during the tough times. Wilbur & Sam: I'm sorry, and I miss you. Jessica: you know I could not have done any of this without you; thank you for coming along for the journey. Brigitte: allô!

Finally, while most would thank their mentor first and foremost, I am thanking mine last, but definitely not least. Dr. Lovly belongs here, at the back end of a long list of names; a senior author coordinating all of the above. Even with the help from all the aforementioned people, I'm not so sure I could have put together this dissertation without Dr. Lovly's invisible hand. I can only hope to one day emulate your focus, selflessness, dedication, and attention to detail. Thank you for taking a chance on me, for spending way too much time making sure I get things right, for your patience in the face of my deadline-driven timing, for creating a warm and caring lab atmosphere—the list goes on and on. You have been an inspiration, a mother, a boss, and a mentor. You've managed to make me a better communicator, scientist, and person—by way of shared laughter, troubleshooting, editing, crying, writing, cursing, and winning. Thank you for everything that you have done and are doing to aid in my development. I will continue to lean on you as I navigate the medical and research years ahead. I am fortunate to have trained with you and am excited to see where you go.

The work in this dissertation is based on a handful of nameless lung cancer patients that were willing to jump into the unknown for hope and for science. Thanks to you, your families, and your sacrifices. Godspeed.

PREFACE

This dissertation is being submitted to the Faculty of the Graduate School of Vanderbilt University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research herein—detailing the discovery and the initial characterization of epidermal growth factor receptor (EGFR) rearrangements in patients with lung cancer—was conducted by Jean-Nicolas Gallant (me), under the supervision of Dr. Christine M. Lovly, M.D., Ph.D. between 2014 and 2016. This work is, to the best of my knowledge, original, except where acknowledgements and references are made to previous work. For example, part of this work already has been published:

Gallant JN, Sheehan JH, Shaver TM, Bailey M, Lipson D, Chandramohan R, Brewer MR, York SJ, Kris MG, Pietenpol JA, Ladanyi M, Miller VA, Ali SM, Meiler J, and Lovly CM. EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib. Cancer Discovery. 2015 Nov 1;5(11):1155–63.

Konduri K*, Gallant JN*, Chae YK, Giles FJ, Gitlitz BJ, Gowen K, Ichihara E, Owonikoko TK, Peddareddigari V, Ramalingam SS, Reddy SK, Eaby-Sandy B, Vavala T, Whiteley A, Chen H, Yan Y, Sheehan JH, Meiler J, Morosini D, Ross JS, Stephens PJ, Miller VA, Ali SM, and Lovly CM. EGFR Fusions as Novel Therapeutic Targets in Lung Cancer. Cancer Discovery. 2016 Jun 2;6(6):601–11.
*co-first authors

Some final logistical points: references follow each chapter; endnotes (also found at the end of each chapter) are indicated as superscripted miniscule letters; and tables in landscape orientation have no pagination.

Enjoy.

TABLE OF CONTENTS

Page

LIST OF TABLES

Main Tables Page

Appendix Tables

LIST OF FIGURES

Main Figures Page

Appendix Figures

CHAPTER I: INTRODUCTION

Overview

This introductory chapter will supply background information necessary to situate and understand the original research contained in this dissertation. The chapter is divided into two sections: the first is focused on clinical background, namely, lung adenocarcinoma; the second is focused on the biology of the epidermal growth factor receptor (EGFR)[a]. The first section will detail the cellular and molecular pathology of lung adenocarcinoma, explain the transition from the historical histological characterization of the disease to the modern molecular characterization, and, finally, describe EGFR-mutant lung cancer, a molecular subtype of lung adenocarcinoma. The second section will concentrate on the biology of EGFR, including the receptor's structure, function, and aberrant activation in lung cancer.

Clinical Background

Lung Adenocarcinoma

Lung cancer is the leading cause of cancer-related deaths worldwide, accounting for over 160,000 deaths per year in the United States (2). Despite advances in the molecular characterization of tumors, the histopathology of lung cancer remains the basis for diagnosis and treatment of the disease. As such, lung cancer remains broadly divided into histologically-defined non-small cell lung cancer (NSCLC; 85% of lung cancers) and small cell lung cancer (SCLC; 15% of lung cancers) **(Figure 1.1A)** (3). NSCLC can further be divided into several histological subtypes, of which adenocarcinoma is the most common—comprising nearly half of all lung cancers **(Figure 1.1B)** (4).

Although lung adenocarcinoma accounts for under 5% of cancer diagnoses every year, it accounts for over 10% of all cancer-related deaths (5). The prognosis remains poor for patients with lung adenocarcinoma because the cancer most commonly is detected at an advanced stage[b]. The majority of symptomatic patients with lung adenocarcinoma have advanced disease. In fact, more patients with lung adenocarcinoma present with vague symptoms of metastatic disease (such as headache, visual changes, or seizures) or nonspecific constitutional symptoms (such as anorexia, fatigue, and weight loss) than with symptoms related to the primary tumor (such as cough or dyspnea) (8-12).

Treatment of lung adenocarcinoma is commensurate with the stage of the disease. Surgery provides the best chance for cure in the setting of early stage disease (13). Multimodal treatments, which can include combinations of surgery, radiotherapy, and/or chemotherapy, are used successfully in intermediate stage disease (14). However, because patients commonly present with advanced or unresectable disease, chemotherapy is the mainstay of treatment for lung adenocarcinoma. Until recently, two agent mixes of cisplatin or carboplatin and paclitaxel, docetaxel, gemcitabine, or pemetrexed—called platinum doublets—were used to treat the majority of patients with lung adenocarcinoma (15)[c]. Irrespective of the mix of agents used, objective response rates (ORR[d]) of lung adenocarcinoma treated with platinum doublets ranged from 20–35% and overall survival (OS[e]) plateaued at 8–10 months (21,22). The advent of immunotherapy, emerging as the standard of care for most patients with NSCLC (23), is likely to change these dire outcomes.

From Phenotype to Oncogene Addiction

Nearly all lung cancers exhibit features of epithelial cells and are accordingly classified as carcinomas. The histological features that distinguish adenocarcinomas from other lung cancers are glandular differentiation and mucin production (24). The glandular elements and mucin vacuoles of adenocarcinomas may be arranged in several aberrant patterns, and these various arrangements form the basis for the division of adenocarcinomas into several different subtypes **(Figure 1.1C)** (4). Historically, the classification of lung adenocarcinoma relied on the expertise of individual pathologists to identify these patterns—leading to differing diagnoses, treatments, and prognoses. To rectify this issue, the World Health Organization (WHO) created, and has updated and maintained (3), a standard nomenclature for the classification lung cancer based on precise histology and/or cytology.

Although histological subtypes still influence tumor classification and treatment, the recognition that lung cancer is a genetic disease has prompted a push to reconfigure cancer taxonomies according to molecular criteria (3,25). While there are, on average, more than 300 non-synonymous mutations per lung adenocarcinoma (26), mutations in only 140 or so of these genes can promote or 'drive' lung tumorigenesis (27). The products of these mutated cancer driving genes—drivers (28)—can be mapped onto a limited set of cell signaling pathways (29), which themselves can be grouped into a couple of cellular processes that define

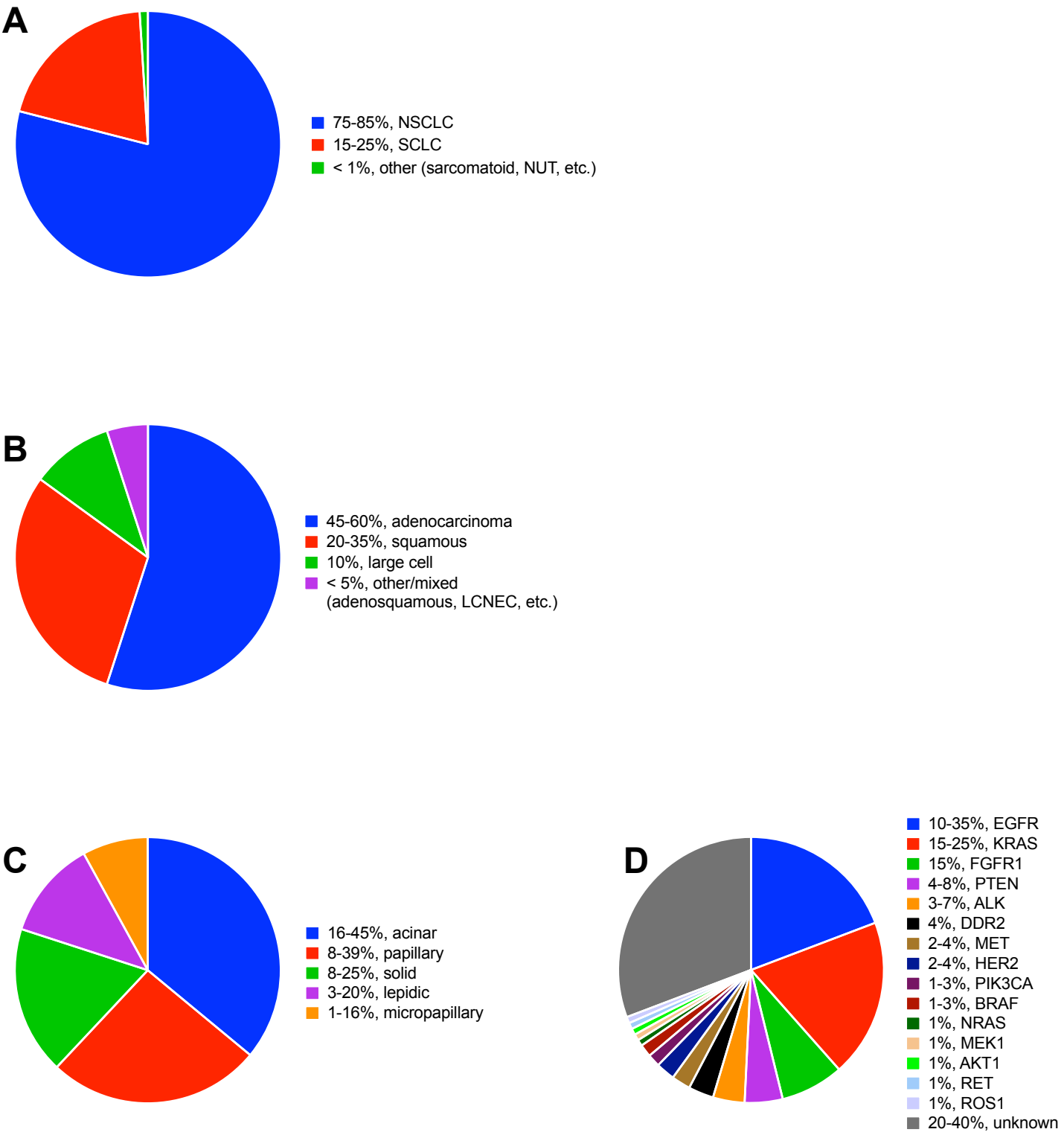


Figure 1.1: Progressive subtyping of lung cancer
A, classical view of tumors arising in the lung including non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) (3). Rare tumor types include NUT carcinoma, defined by the presence of BRD-NUT fusion protein. **B,** histological subtypes of NSCLC (5). Rarer tumor types include large cell neuroendocrine carcinoma (LCNEC). **C,** histological subtypes of lung adenocarcinoma (4,30). **D,** molecular subtypes of lung adenocarcinoma (31). EGFR, KRAS, FGFR1, etc. correspond to the dominant driver (gene) alteration.

the selective advantage of cancer cells: altered cell fate and survival (27). Though there is still some debate, just a few key driver mutations are sufficient to transform normal human lung cells to a tumorigenic state: P53 or RB inactivation and MYC, RAS, and/or TERT activation (32-36). Inhibition of any of these key drivers can lead to loss of the malignant phenotype (37), hinting that these drivers may not only be sufficient, but also necessary for lung tumor formation. Indeed, many data accumulated over the past decades show that cancers are 'addicted to' (that is, physiologically dependent) on the continued activity of a few specific activated or overexpressed drivers for their malignant phenotype (38,39). Nearly all lung adenocarcinomas contain genetic alterations in—at least some of— the arbiters of cell fate (P53, RB, MYC, TERT) leading to a selective growth advantage by favoring cell division over differentiation (26). Where lung adenocarcinomas seem to differ from one another is in the way that they activate cell survival signaling (typified by RAS activation) (40,41). Therefore, despite the enormous histological variation (over 10 subtypes (42)) and genetic complexity (over 10 mutations per megabase (43)) present in lung adenocarcinoma, one can 'simply' classify tumors by the way in which their cell survival signaling is driven.

As such, lung adenocarcinoma can now be considered a cluster of discrete molecular diseases, with most being defined by a single driver alteration that effects cell survival **(Figure 1.1D)** (44). The defining molecular alterations used to subtype adenocarcinomas are all involved in cell survival for two reasons: firstly, lung adenocarcinomas otherwise share a similar cell fate (highlighted by some combination of P53, RB, MYC, TERT alterations); and, secondly, the alterations involved in cell survival, as opposed to cell fate, are currently actionable—that is, potentially responsive to an available targeted therapy (45). Such known and/or putative (cell survival) drivers in lung adenocarcinoma include mutated AKT1, BRAF, DDR, EGFR, HER2, KRAS, MEK1, MET, NRAS, PIK3CA, PTEN; re-arranged ALK, EGFR, NTRK1, RET, ROS1; and amplified FGFR1 and MET[f] (31). Alterations in any of these driver genes effects cell survival—mainly via modulation of the MAPK and/or PI3K/AKT signaling pathways—and are, or may be, amenable to available targeted therapy (50). Rather than empiric treatment based on histological classification (an approach whose benefits have plateaued; ORR 20–35% (21)), molecular classification of lung adenocarcinoma by driver alterations has revolutionized the prognosis and treatment of the disease. It is now standard for tumors to be genotyped and for patients to be given a driver-targeted therapy, to which they will likely respond (ORR ≥ 60% (51)). Such advancements have made lung adenocarcinoma the poster child for precision medicine, which is most apparent in EGFR-mutant (EGFR-driven) lung adenocarcinoma.

EGFR-mutant lung cancer

As soon as EGFR, a receptor tyrosine kinase (RTK), was connected with lung cancer (52), much effort was spent in attempts to therapeutically block the protein's aberrant signaling, first with antibodies[g], and then with small molecule EGFR tyrosine kinase inhibitors (EGFR-TKIs). Ever since phase III studies demonstrated their efficacy (57,58), the first EGFR-TKIs, erlotinib (59) and gefitinib (60), have been in clinical use. However, during initial clinical studies against NSCLC, the ORR of EGFR-TKIs varied greatly, from 7 to 70%, and averaged ≤ 25%—comparable to standard chemotherapy (24). Retrospective analyses led to the determination that response to an EGFR-TKI was dependent on specific clinical factors including smoking history (never-smokers responded 38% of the time vs. former/current smokers 10%), gender (men 13% vs. women 36%), histological subtype (adenocarcinoma 27% vs. 7% for all others), and ethnicity (East Asian 29% vs. 8% for all others) (61). These observations led to the design of new clinical trials which attempted to evaluate the efficacy of EGFR-TKIs in never-smoking East Asian patients with lung adenocarcinoma, leading to more predictable and durable responses (62).

In seeking to understand why some patients with lung adenocarcinoma had dramatic responses to first-generation EGFR-TKIs, while others derived little clinical benefit, somatic mutations in EGFR were discovered in a subset of NSCLC patients (63-65). These mutations, primarily found as a series of deletions in exon 19 (ex19del) or as leucine to arginine substitutions at codon 858 (L858R), enable constitutive activation of the EGFR tyrosine kinase domain (TKD) and sensitize cancer cells to EGFR-TKIs (66). This information was incorporated into clinical trial design, leading to trials comparing EGFR-TKIs to platinum-based chemotherapy in select patients with EGFR-mutant (defined as EGFR ex19del or L858R) tumors. Groundbreaking initial results demonstrated that patients with EGFR-mutant tumors achieved significantly longer progression-free survival (PFS[h]) with gefitinib as opposed to those receiving platinum doublets (hazard ratio[i] of 0.48 (62)**.** Moreover, ORRs in these patients were over 60% (62). Similar results subsequently were observed in populations of differing ethnicities and with different EGFR-TKIs (68-74). As a result, erlotinib, gefitinib and, most recently, afatinib (a second generation EGFR-TKI (75)) received approval for first-line treatment of EGFR-mutant NSCLC. Furthermore, several multivariate analyses have revealed that EGFR mutation status is

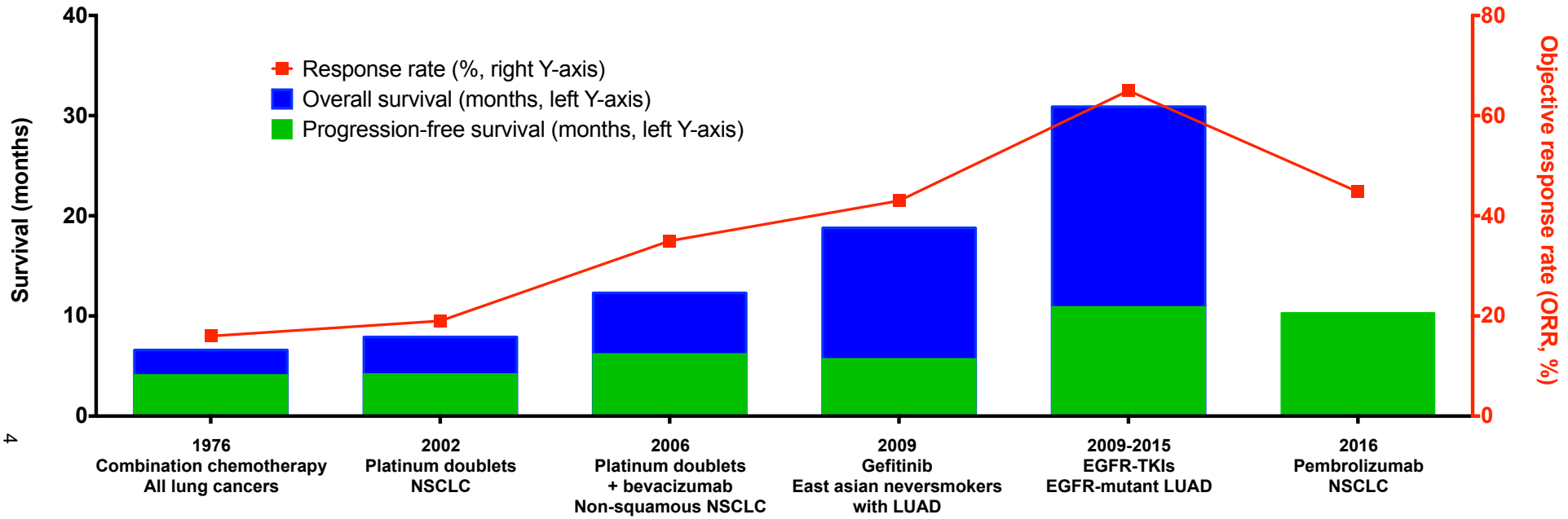


Figure 1.2: Refinement in the treatment of lung cancer
Evolution in the progression free survival (PFS, green), overall survival (OS, blue) and objective response rates (ORR, red) of treatment naïve lung cancer. The trend towards treating molecularly-defined cohorts of lung adenocarcinoma (LUAD) to achieve better response and survival rates was recently bucked by pembrolizumab, an anti-PD1 antibody (17,21,23,62,76,77). This immunotherapy agent holds promise in treating broad swaths of NSCLC without a driver mutation. The ORR of pembrolizumab is significantly better than platinum-based chemotherapy against NSCLC, as is its survival benefit—so much so that median OS data has yet to be reached and is not shown here.

the only significant prognostic factor (as opposed to previously identified clinical factors [smoking status, gender, histology, and ethnicity]) in predicting response to EGFR-TKIs (78-80). Together, this has led to a revolution in the treatment of NSCLC with upfront molecular testing and treatment using an EGFR-TKI for patients with EGFR-mutant lung cancer **(Figure 1.2)** (4,81,82).

Despite advances in the identification and treatment of patients with EGFR-mutant lung cancer, there remain large clinical hurdles. For one, patients with EGFR-mutant lung adenocarcinoma treated with an EGFR-TKI have a similar OS as patients treated with platinum doublets (83-85)[j]. Whilst these patients' PFS and quality of life are measurably better (86), increasing OS remains a major goal of cancer therapy[k]. Secondly, almost all patients treated with an EGFR-TKI develop acquired resistance after 10–16 months of treatment (88,89). The most common mechanism of resistance to erlotinib, gefitinib, and afatinib is a secondary mutation in the EGFR TKD that results in a change from threonine to methionine in codon 790 (T790M) (90,91). Other mechanisms of resistance, including EGFR gene amplification, activation of bypass signaling, and histological transformation have all been described in the setting of acquired resistance to EGFR-TKIs (92-94). Options are available for patients who progress on EGFR-TKIs—such as osimertinib (95); an FDA-approved T790M-specific EGFR-TKI—but there is still a long way to go in order to make EGFR-mutant lung adenocarcinoma a managed chronic disease.

Molecular Background

The HER Family of Receptors

EGFR is a 1210 amino acid (aa) residue nascent polypeptide, and, after processing of the N-terminal signal peptide sequence, a 1186 aa, 170 kD, N-linked-glycosylated, membrane-bound RTK (96). The processing event has led to two widely used but divergent numbering systems for EGFR's amino acid residues: one for the nascent polypeptide and one for the mature protein. Residues 1–1186 of the mature protein numbering system correspond to aa 25–1210 of the nascent peptide number system. Although the use of the mature protein numbering system is ingrained in the biochemistry literature, the nascent peptide numbering system (1–1210) will be used throughout as it is easier to reference when interconverting between nucleic and amino acids.

While EGFR was the first RTK discovered (97) and used as a model to work out much of the protein family's biochemistry (98), it is not a typical RTK (99). EGFR has been shown to have unique biochemistry amongst RTKs[l] (for one, it has no need for activation loop phosphorylation (102,103)) and its biology has been proven to be nuanced: the effects of knocking out the EGFR gene in mice range from embryonic lethality in one genetic background to peri-partum mortality in another[m] (105). Partly owing to EGFR's complexity is its wide array of ligands and familial dimerization partners. EGFR was the first of four subsequently discovered human epidermal growth factor receptor (HER) family members, which also include HER2 (106), HER3 (107), and HER4 (108). Aside from HER2, each HER family member binds ligand, forms 2:2 receptor:ligand complexes[n], and homo- or hetero- dimerizes with another HER to propagate downstream signal (110,111). EGFR binds at least seven different activating ligands in humans: EGF itself, TGFα, BTC, HB-EGF, ARG, EPR, and EGN (112). Including all known ligands, receptors, and isoforms, there are 611 possible combinations of ligand-activated HER complexes (113). Due to subtle structural differences in the receptors, each of these ligand-receptor combinations may elicit distinct but overlapping sets of responses in cells (114).

Architecture, Activation, and Multimerization of EGFR

The major structural features of EGFR are its extracellular domain (ECD), transmembrane helix (TM), tyrosine kinase domain (TKD), and carboxyl terminus region (C-term) **(Figure 1.3A-B)**. The extracellular domain itself is divided into four sub-domains: domains I and III, which are related leucine-rich segments[o] that participate in ligand binding, and domains II and IV, which contain numerous cysteine residues that participate in disulfide bond formation and dimerization (116). The extracellular domain is separated from the EGFR TKD by the TM; a single membrane-spanning α-helix. The EGFR TKD catalyzes the transfer of γ-phosphate groups from bound ATP onto specific tyrosyl moieties—mainly onto tyrosines located within its own C-term (117)[p]. In addition to the previously mentioned TKD and C-term, the intracellular portion of EGFR contains a juxtamembrane segment (JM) that helps promote receptor activation (119,120). While each of these structural features has been studied in detail, due to its involvement in cancer, the EGFR TKD has been of particular interest to (cancer) researchers.

The structure of the EGFR TKD resembles that of typical TKDs (98,121-123). The active site is located between the two lobes (the smaller N-(terminal) lobe and the larger C-(terminal) lobe) of the TKD. Upon ATP

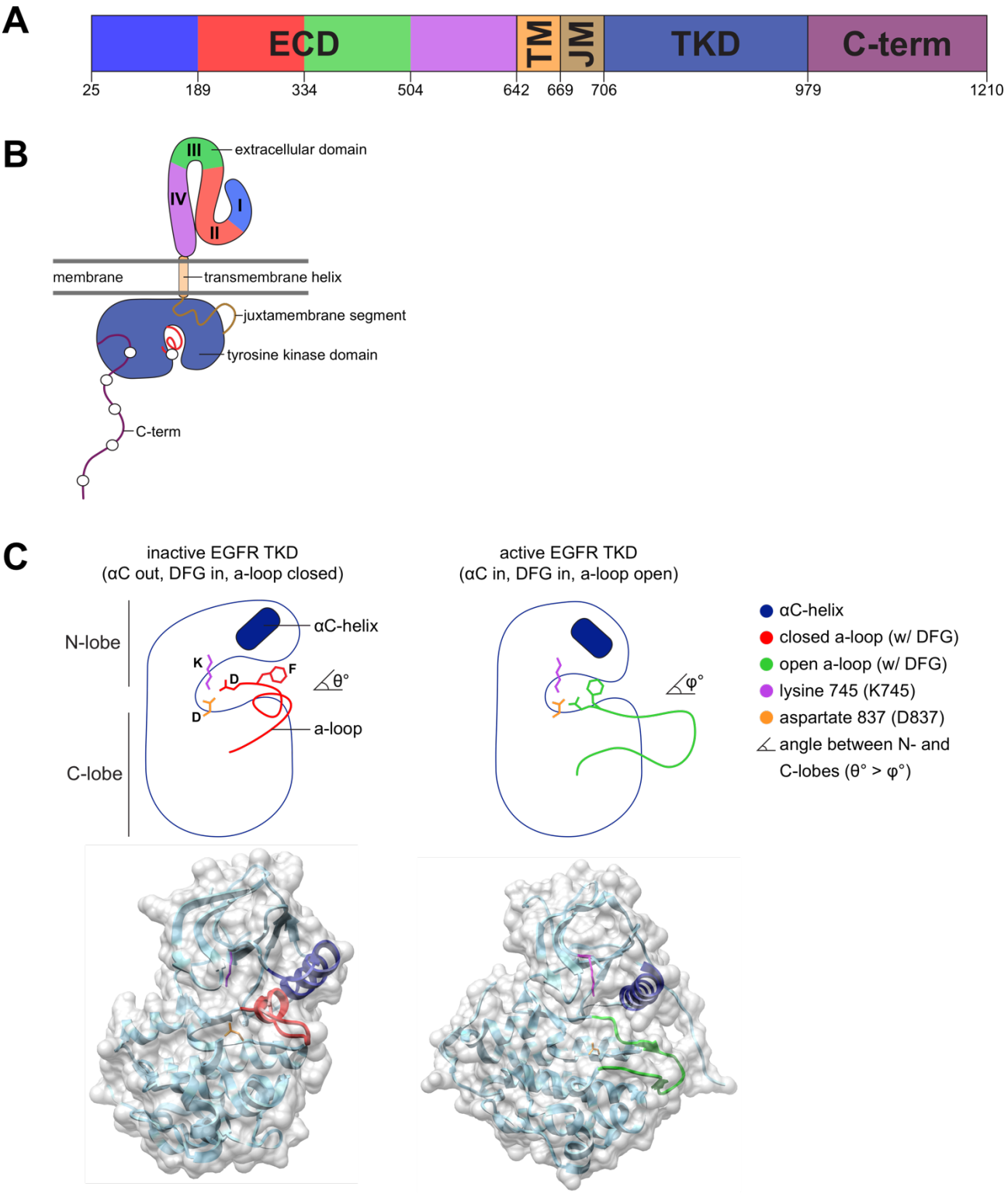


Figure 1.3: Overview of EGFR architecture
A, schematic representation of EGFR protein domain structure. Numbers correspond to amino acid residues. ECD, extracellular domain; TM, transmembrane helix; JM, juxtamembrane segment; TKD, tyrosine kinase domain; C-term, carboxyl terminus region. **B,** general overview of EGFR protein relative to the cell membrane; adapted from (116). I-IV indicate ECD sub-domains. **C,** top: Detailed view of the inactive (left) and active (right) EGFR TKDs; adapted from (116). Bottom: ribbon diagram and space filling model of the in/active EGFR TKDs generated using Chimera (124) (active EGFR PDB=2GS6; inactive=2GS7 (125)). Throughout: the αC-helix is colored navy, the ATP-coordinating lysine is colored purple, the catalytic aspartate is colored orange, and the Asp–Phe–Gly (DFG) motif / activation loop colored red if inactive and green if active. D, F, and K stand for the amino acid residues aspartate, phenylalanine, and lysine, respectively.

binding, a hinge motion changes the relative orientation of these two lobes from a more open state (in the inactive conformation) to a more closed state (in the active conformation) **(Figure 1.3C)**. Three secondary structures within the kinase domain—the N-lobe αC-helix (αC in=active EGFR; αC out=inactive), the C-lobe DFG motif (DFG in=active; DFG in=inactive[q]), and the C-lobe activation loop (a-loop open=active; a-loop closed=inactive)—control this hinge motion (113). In three dimensional space, the αC, DFG, and a-loop combine to form two highly conserved tertiary structures or 'spines' (126,127). Residues in the spines that are oriented towards the active site are key to the catalytic properties of EGFR (and other TKDs) (128). Of particular importance for EGFR are a lysine at position 745 (K745, which forms salt bridges with the α- and β-phosphates of ATP) and a catalytic aspartate at position 837 (D837, which accepts a proton from the tyrosine substrate's hydroxide group). Along with several other conserved residues on the spines (123), K745 and D837 coordinate the transfer of the γ-phosphate of ATP onto exogenous tyrosyl substrates and/or the C-term of EGFR, the latter in a trans manner. While these structural elements of EGFR resemble that of typical TKDs, the protein's mechanism of activation diverges from canonical RTKs (99).

Because EGFR does not require activation loop phosphorylation[r], the classical view of EGFR activation involves the monomeric receptor binding to a single ligand (110), the ligand-EGFR complex unfolding from the 'tethered' into the 'extended' form (129), and the C-term being trans-autophosphorylated while EGFR is in the dimeric form (111)**(Figure 1.4A)**. As opposed to all other RTKs, dimerization and activation of EGFR are driven by receptor-receptor interactions (98), with a 'dimerization arm' in domain II providing the majority of contacts across the ECD interface (130-132). Without bound ligand (in the monomeric receptor) this dimerization arm is buried in a tight intramolecular tether via an interaction with domain IV (133,134), and dimerization of the receptor is mostly (auto)inhibited[s] (138). Activating ligands, such as EGF, bind to domains I and III[t] of the ECD (130,131) and promote a hinge motion in the ECD of the receptor—leading to the extended form of EGFR and the exposure of the dimerization arm (domain II). This model of EGFR activation holds that dimerization arms from individual ligand-bound HER family members need to interact in order for the receptors to form active dimers. However, receptor dimerization is not sufficient for EGFR signaling. The introduction of inter-receptor disulfide bonds (leading to constant dimerization) produces both active and inactive dimers, demonstrating that downstream signaling require stereospecific ICD interactions (143)[u].

Indeed, activation of the EGFR kinase involves the specific formation of an asymmetric dimer of two EGFR TKDs, in which one kinase serves as an allosteric activator of the other (125). This allosteric interaction involves the C-lobe of one EGFR TKD (the 'activator') forming a tight interface with the N-lobe of another (the 'receiver'), which stabilizes the receiver in the active conformation **(Figure 1.4B)**. This mode of activation does not require the tyrosine in the activation loop, as with all other RTKs (98); instead, it requires the intact asymmetric TKD dimerization interface (125). EGFR variants containing mutations that disrupt the asymmetric dimer interface—such as the C-lobe activator-impaired V948R mutant and the N-love receiver-impaired I706Q mutant—do not respond to EGF binding. However, together, the activator-impaired and receiver-impaired mutants can form an intact asymmetric interface and autophosphorylate EGFR (116). Thus, while ligand binding and ECD interactions are important in bringing two EGFR molecules together, specific TKD interactions, namely asymmetric dimerization, are needed to activate EGFR and enable trans-autophosphorylation.

Additional molecular interactions stabilize the activated form of EGFR and make it possible for EGFR to multimerize **(Figure 1.4C)**. While research on EGFR has focused on the inactive monomer to active dimer transition, it is known that tetramers (145-147) and clustering (148) are important for signaling[v]. Inspection of the EGFR ECD dimer interface shows that, although the dimerization arm provides the majority of interactions, several secondary receptor-receptor contacts also can be made by other parts of domains II and IV and EGFR (109,130-132). The transmembrane helix, which connects the ECD and ICD, also dimerizes (154,155), and this dimerization is coupled to the asymmetric dimerization of the TKDs (156-158). There also is a key role for the juxtamembrane domain in the activation EGFR, whereby this small structure allosterically regulates EGFR by locking (latching) the TKD in the active confirmation (119,120,159). Taken together these, and other, interactions enable head to tail (C- to N-lobe) EGFR multimerization that is able to increase C-term phosphorylation (109,160)[w]. Whether the highly-active multimeric (as opposed to dimeric) EGFR is the main signaling unit remains an active area of investigation (162).

Signaling Downstream of EGFR

EGFR propagates intracellular signal by recruiting and activating proteins via the phosphorylated tyrosine (pY) residues on its C-term **(Figure 1.5A)**. Binding of effector proteins via their SH2 and PTB (pY-

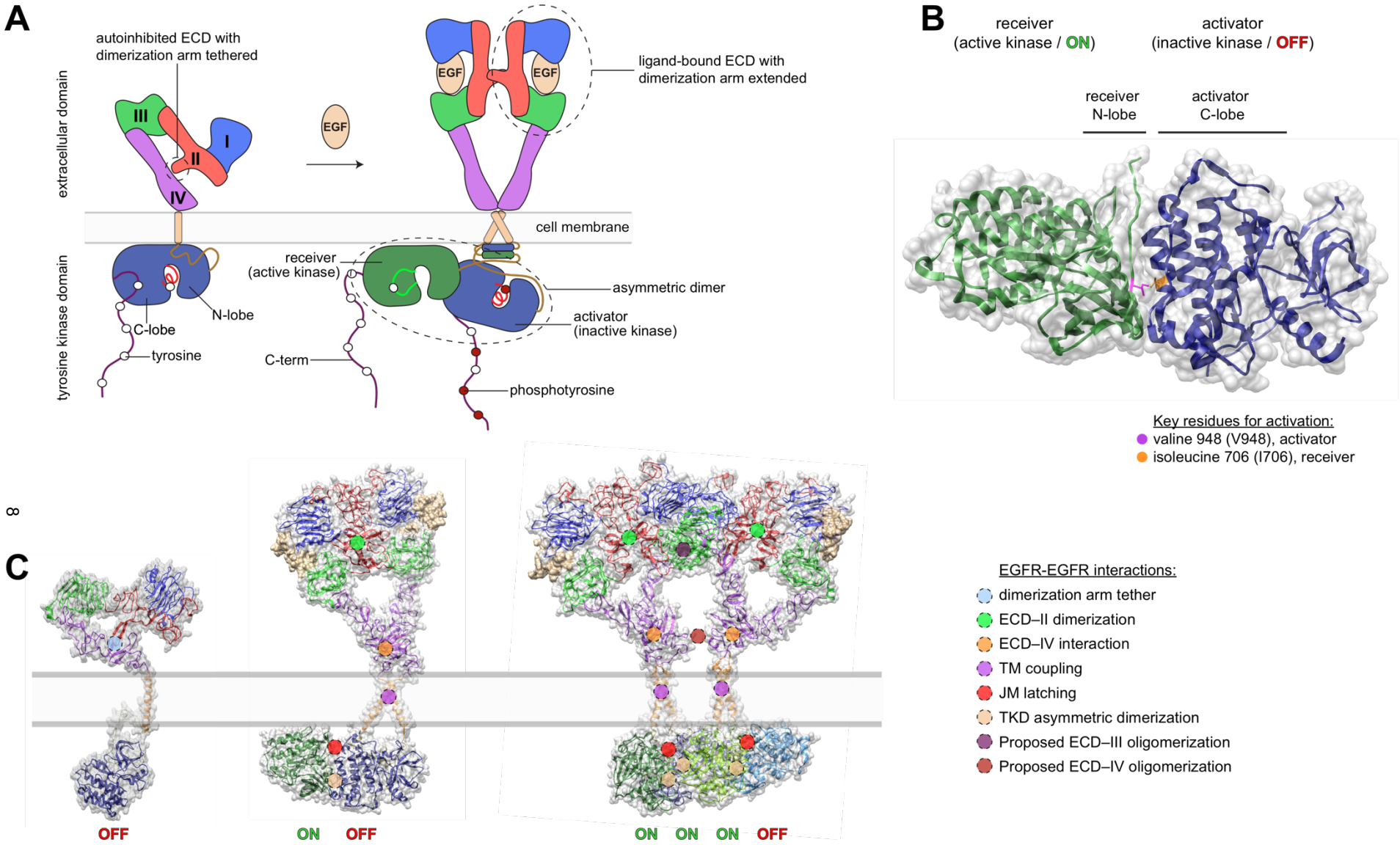


Figure 1.4: Activation and multimerization of EGFR

A, model of ligand-induced dimerization-mediated EGFR activation; adapted from (116). Colors correspond to EGFR protein domains from Figure 1.3. **B,** ribbon diagram and space filling model of the asymmetric dimer of the EGFR TKD (PDB 2GS6 (125)) generated using Chimera (124). The activator kinase is colored navy, and the receiver (enzymatically active) kinase is colored dark green. Key contact residues at the asymmetric interface are highlighted in orange (valine 948) and magenta (isoleucine 706). **C,** ribbon diagram and space filling models of the complete EGFR monomer (164), dimer (156), and tetramer (160) generated using Chimera(124). Ribbon colors correspond to EGFR protein domains from Figure 1.3, except for additional TKDs in dark green, chartreuse, and sky blue. EGF is presented as a space-filling beige model. EGFR-EGFR interactions are highlighted, including the proposed ECD sub-domain III (160) and IV (109) oligomerization interfaces.

binding) domains onto EGFR's C-term activates three main signaling pathways: the MAPK, PI3K, and PLC-γ signaling pathways (163). The proteins that directly interact with activated (auto-pY) EGFR can generally be categorized as: positive regulators of mitogenic signaling, such as the kinase SRC, the phospholipase PLC-γ, and the transcription factor STAT3; adaptor/scaffolding proteins with no catalytic activity, but with the ability to assemble and regulate protein complexes, such as GRB2 and SHC; and negative regulators of signaling such as the E3 ubiquitin ligase CBL and phosphatase SHP1 (163). At least 65 different proteins with SH2 or PTB domains have been validated as binding to EGFR's C-term pYs (165). These interacting proteins are often multivalent—having multiple SH2, PTB, SH3, PH, or other domains —and mediate interactions with additional proteins or with the cell membrane (166). When the multivalency of adaptor proteins (≥ 65 effectors), the possible variety of autophosphorylated sites (~12 tyrosines) on EGFR's C-term, and the receptor's regulation by different growth factors and oligo/hetero/homo-dimerization (611 possible combinations) are all considered, the picture of EGFR signaling becomes complex and nuanced.

Much work has gone into untangling and simplifying EGFR signaling (167,168). When the entirety of the known signaling pathway is mapped out, a notable feature becomes apparent: a bowtie (or hourglass) structure (169). The vast ligand–EGFR–SH2/PTB network of receptor signaling complexes converges onto just a handful of molecules, which themselves activate a variety of signaling cascades **(Figure 1.5B)**. The critical nodes in EGFR signaling are a series of previously mentioned non-receptor tyrosine kinases (non-RTKs), small GTPases, and phosphatidylinositol phosphates (PIPs). Specifically, the nodes in EGFR signaling are the non-RTKs PYK2 and SRC, the small GTPases RAS and RAC, and the PIPs PI4,5-P_2, and PI3,4,5-P_3 (169)[x]. As the complexity of EGFR signaling via PYK2 and SRC is being untangled (170,171), the field has forged ahead with the simplified view that EGFR signals primarily through the MAPK (RAS node) and or PI3K/AKT (PIP node) signaling pathways (172)[y].

<u>Learning from EGFR Mutations to Design Better Therapies</u>

Since EGFR was recognized as an oncogene due to its homology to v-erb-B, a retroviral protein that enables avian erythroblastosis virus to transform chicken blood cells (173), the study of cancer and EGFR have inexorably been linked (174,175). EGFR overexpression in brain, head & neck, lung, pancreatic, and colorectal cancers led to myriad clinical trials aimed at targeting the protein for therapeutic effects (113). While EGFR-TKIs are FDA-approved against cancers in a couple of different body sites (113,176), the inhibitors have their most significant effects against EGFR-mutant lung adenocarcinoma. At the same time, cetuximab, the anti-EGFR antibody, has been found to be most clinically effective for head & neck (177) and colorectal cancers (178). These clinical findings lead to basic questions: why are EGFR-TKIs most effective in lung cancer and why are anti-EGFR antibodies most effective in head & neck and colorectal cancers? The answers to these questions may have to do with EGFR's mechanism of activation. Anti-EGFR antibodies are thought to be effective in colorectal and head & neck cancers because these cancers express locally high levels of EGFR-activating ligands and because anti-EGFR antibodies block ligand binding (175). At the same time, EGFR-TKIs are effective against certain lung adenocarcinomas because these cancers harbor recurrent mutations in the kinase domain—where TKIs bind (63-65). While EGFR is implicated in colorectal, head & neck, and pancreatic cancers, EGFR mutations in these malignancies are not consistently identified[z]. Recurrent genetic alterations are useful in cancer biology because they hint at how an oncogene's product is activated. Along these lines, study of lung adenocarcinoma associated EGFR mutations has been critical in pushing forward understanding of EGFR biology.

Studies of the common lung adenocarcinoma associated EGFR mutations (L858R and ex19del) have provided further insights into the biology of the receptor. Studies of EGFR-L858R have demonstrated that it is ≥ 50 times more active than wild-type (WT) EGFR (66) as a result of preferentially adapting the active conformation (180). More importantly, studies on EGFR-L858R have added a plethora of evidence to support EGFR's activation via receptor-mediated asymmetric dimerization. While monomeric EGFR-L858R is active in the absence of EGF (181,182), the protein is even more active as part of an asymmetric dimer (183), especially when it adopts the receiver position (184). In fact, the oncogenic properties of EGFR-L858R may be entirely driven by the greater propensity of these mutants to form active dimers (185). The ECD also appears to play a key role in the activation of mutant EGFR, even in the absence of ligand (186). Study of EGFR ex19del has likewise been generalized to help with the understanding of the mechanism of activation of similar deletions in other oncogenes (187). In due turn, increased knowledge about the structure and function of EGFR mutations has enabled the development of more potent irreversible second-generation (188) and third-generation (189,190) EGFR-TKIs for the treatment of lung adenocarcinoma.

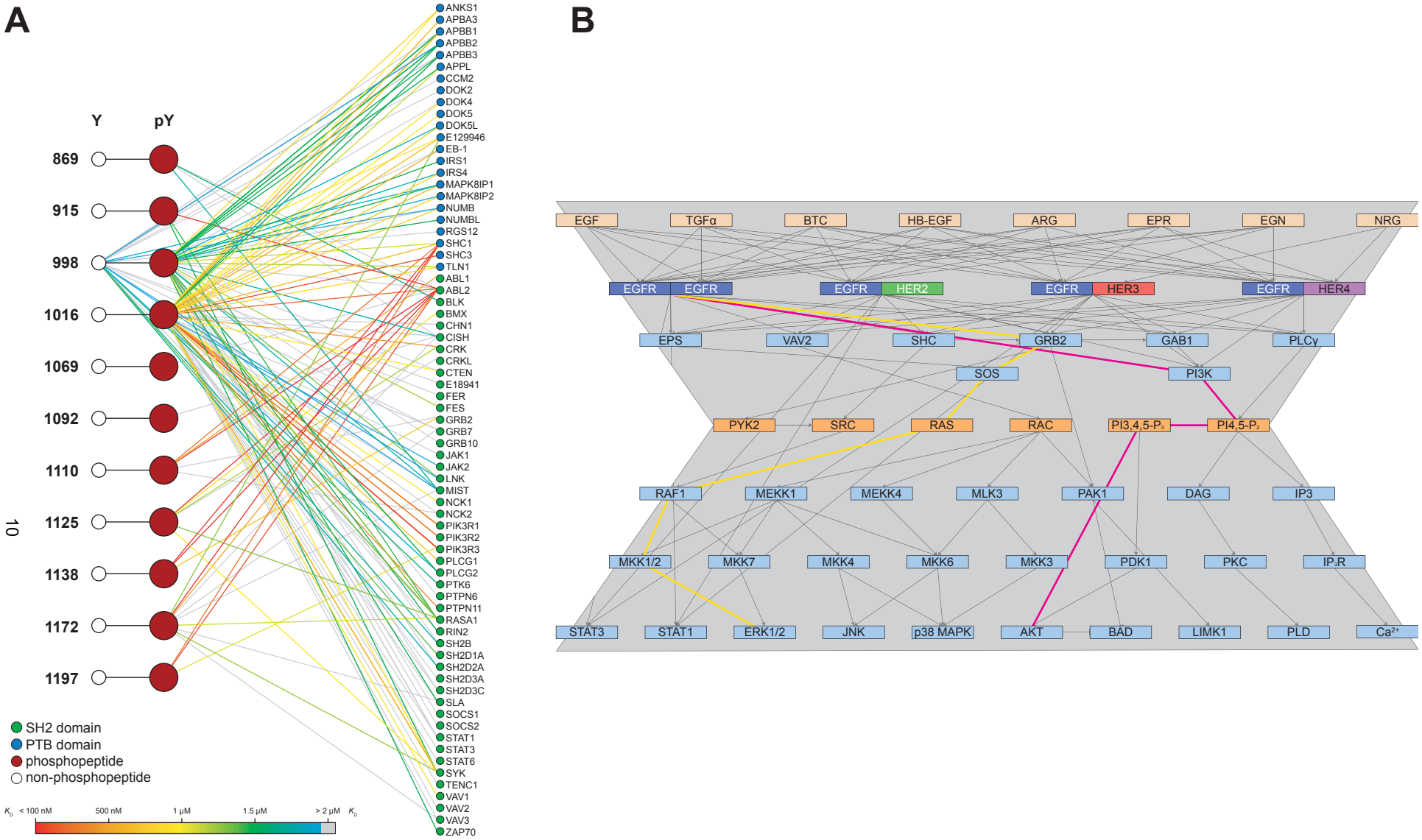


Figure 1.5: Overview of EGFR signaling

A, graphical representation of phosphotyrosine-mediated recruitment of adaptors to EGFR; lines connect phosphotyrosines to adaptors and provide a quantitative protein interaction network for EGFR, adapted from (165). Dark red circles represent phosphopeptides at each indicated EGFR tyrosine residue; clear circles represent the non-phosphorylated version of each phosphopeptide; green circles represent SH2 domains; and blue circles represent PTB domains. Lines connecting peptides to domains indicate observed interactions, colored according to the affinity of the interaction (see legend). **B,** simplified bow-tie architecture of the EGFR signaling pathway, adapted from (169). Beige = ligands; light blue = signaling intermediaries; orange = critical nodes in EGFR signaling [non-RTKs (PYK2 and SRC), small GTPases (RAS and RAC), and PIPs (PI3,4,5-P_3 and PI4,5-P_2)]; yellow = MAPK signaling pathway (RAS node); and magenta = PI3K/AKT signaling pathway (PIP node).

There now exist over 20 well-characterized EGFR-TKIs that have undergone some clinical development in NSCLC **(Table 1.1)**. First generations EGFR-TKIs were a fallout of attempting to study WT-EGFR with small molecules (191). Although there is still some controversy (192), the majority of first-generation EGFR-TKIs are thought to selectively bind the active mode of the TKD (which explains the effectiveness in EGFR-mutant lung adenocarcinoma) (180). This has been attributed to two factors: 1) the recognition of the active state of the kinase by certain EGFR-TKIs (193,194) and 2) the higher ATP affinity of mutant enzyme relative to the WT-EGFR (195). Unlike the ATP-competitive first generation EGFR-TKIs, second generation EGFR-TKIs, were designed to bind the TKD irreversibly in attempts to overcome acquired resistance. While these agents proved efficacious against models of T790M-mediated EGFR-TKI resistance (196,197), they failed to meet endpoints against T790M-positive EGFR-mutant lung adenocarcinoma in clinical trials—due to the limited therapeutic window offered by these agents (198,199). As such, third generation EGFR-TKIs were developed: irreversible EGFR-TKIs with a puckered ring structure that can bind the TKD active site in the presence of T790M (189,190,200). These agents have proven to be active in patients (95,201), and one of them, osimertinib, is FDA approved. However, resistance to osimertinib is emerging, suggesting that alternative strategies and/or EGFR-TKIs are needed (202).

Recent work on EGFR mutants in glioblastoma has presented another possible way to inhibit EGFR: by using inhibitors selective for the mutant-specific conformation of the receptor (203). TKIs can be categorized by generation and mode of binding, but they can also be categorized as whether they bind the active form (Type I inhibitors) or inactive form (Type II inhibitors) of EGFR (or their target, in general (204)). EGFR mutations found in glioblastoma are more responsive to Type II EGFR-TKIS (the best characterized being lapatinib) than type I inhibitors (such as erlotinib) (203). Though head to head data is sparse, it appears that lung adenocarcinoma associated EGFR mutations are, dose for dose, more potently inhibited by Type I as opposed to Type II EGFR-TKIs; however, lapatinib still is able to inhibit certain EGFR-mutant lung cancer cell line models at clinically achievable doses (203,205,206). Although clinical development of lapatinib was shifted away from NSCLC (as the agent did not meet predetermined ORR in NSCLC), initial trials were performed in heavily pre-treated patients and failed to test for EGFR mutation status (205,207,208). Because there is no structural reason why EGFR-TKIs could not bind the inactive EGFR TKD effectively (182), more study is needed into Type II inhibitors in the setting of EGFR-mutant lung adenocarcinoma. This is especially true as EGFR mutants are not all alike and there exists a spectrum of activation between the active and inactive confirmations of the TKD (180). Since imatinib (a BCR-ABL TKI) was first recognized as an inactive site TKD binder (209), it has been generally recognized that targeting inactive states of TKDs offers clear advantages in terms of target selectivity, patient safety, and time to resistance (210-212). In light of the mounting success of Type II TKIs[aa], the question is whether the inactive state of the EGFR TKD is a better target for drug discovery and therapy.

Purpose of these Studies

We now know that EGFR mutations (≥ 85% being L858R or ex19del (179)) afflict ~20% of patients with lung adenocarcinoma (214). However, because EGFR mutations historically were interrogated by 'hot-spot' PCR-based methods, most known mutations are biased to fall between exons 18 and 21 (the regions that encodes for L858R and ex19del and which encodes the majority of the TKD). As such, there is an outstanding amount of basic science and clinical data regarding mutations in these exons, and there are less data regarding mutations elsewhere in the protein—such as in the ECD or C-term—even though hundreds of such mutations have been catalogued in NSCLC patients (179,215,216). At the same time, due to the limitations in companion diagnostics and sequencing technologies (82,217), it is likely that other EGFR variants exist in lung adenocarcinoma. This idea is backed by clinical data: a small proportion (up to 20%, depending on the trial) of patients with no detected EGFR-activating mutations show a radiographic response when treated with EGFR-TKIs (62,218-220). While knowledge about common mutations has allowed for rational treatment of specific cohorts (L858R or ex19del) of EGFR-mutant NSCLC, little is known about the remaining 10–15% of EGFR mutations in NSCLC. The rest of these mutations—deemed rare or atypical mutations—are primarily scattered throughout the EGFR TKD and can mediate either sensitivity or primary resistance to EGFR-TKI therapy (221). The goal of these studies is to fill in this knowledge gap: to identify novel EGFR mutations in lung adenocarcinoma, to determine the sensitivity of these alterations to therapy, and, along the way, to uncover more about the biology of EGFR.

generation	compound name	generic name	1º target(s)	type	class	binding	furthest clinical development in NSCLC
1	OSI744 (59)	erlotinib	EGFR	I	EGFR	reversible	FDA-approved (70,71)
	ZD1839 (60)	gefitinib	EGFR	I	EGFR	reversible	FDA-approved (62,68,69)
	GW572016 (194)	lapatinib	EGFR/HER2	II	dual-HER	reversible	Phase I (208)
	BPI2009H (222)	icotinib	EGFR	I	EGFR	reversible	Phase III (223)
	AZD8931 (224)	sapitinib	EGFR/HER2/HER4	I	pan-HER	reversible	Phase I (225)
	XL647 (226)	tesevatinib	EGFR/HER2/VEGFR	I	multitargeted	reversible	Phase II (227)
	ZD6474 (228)	vandetanib	VEGFR/EGFR/RET	I	multitargeted	reversible	Phase III (229)
	AEE788 (180,230)	—	EGFR/HER2	I	dual-HER	reversible	Phase I (231)
	AZD3759 (232)	—	EGFR	I	CNS-penetrant	reversible	Phase I (233)
	BMS599626 (234,235)	—	EGFR/HER2	I	dual-HER	reversible	Phase I (236)
	BMS690514 (237)	—	EGFR/HER2/VEGFR	I	multitargeted	reversible	Phase II (238)
	CUDC101 (239)	—	EGFR/HER2/HDAC	I	multitargeted	reversible	Phase I (240)
2	BIBW2992 (188)	afatinib	EGFR/HER2	I	dual-HER	irreversible	FDA-approved (72,73,75)
	PF299804 (197)	dacomitinib	EGFR/HER2/HER4	I	pan-HER	irreversible	Phase III (199)
	CI1033 (241)	canertinib	EGFR/HER2/HER4	I	pan-HER	irreversible	Phase II (242)
	HKI272 (243)	neratinib	EGFR/HER2/HER4	II	pan-HER	irreversible	Phase II (244)
	EKB569 (245)	pelitinib	EGFR/HER2	II	dual-HER	irreversible	Phase I (246)
	AST1306 (247)	allitinib	EGFR/HER2	II	dual-HER	irreversible	Phase I (248)
	HM781 (249)	poziotinib	EGFR/HER2	I	dual-HER	irreversible	Phase I (250)
	TAK285 (251)	—	EGFR/HER2	II	dual-HER	irreversible	Phase I (252)
3	AZD9291 (190,253)	osimertinib	EGFR T790M	I	mutant-selective	irreversible	FDA-approved (95,254)
	CO1686 (189)	rociletinib	EGFR T790M	I	mutant-selective	irreversible	Phase II (201)
	AC0010 (255)	avitinib	EGFR T790M	I	mutant-selective	irreversible	Phase I (256)
	EGF816 (257)	—	EGFR T790M	I	mutant-selective	irreversible	Phase I (258)

Table 1.1: EGFR-TKIs for the treatment of NSCLC
Inhibitors are categorized on generation based on their being 1st generation: reversible; 2nd generation: irreversible; and 3rd generation: mutant-EGFR selective. In cases where compounds have several names, only the most recent compound name is listed. 1º targets are the targets which the compound was designed to inhibit (listed in order of cell-free IC_{50}) with slashes indicating near-equal potency. The compound type (I=binds active confirmation of EGFR; II=binds inactive confirmation (204)) was determined based on available crystal structures and/or whether the compound was designed against the structure of active EGFR (PDB 2GS6) or inactive EGFR (2GS7). The inhibitor class refers to the common language used to describe the inhibitor. Note that clinical development is specific to NSCLC; some inhibitors, such as vandetanib, are FDA-approved for other indications. Not detailed in this table: natural EGFR inhibitors (e.g. erbstatin (259)), early tyrphostins (e.g. AG1478 (260)), molecules that never underwent clinical testing (e.g. EAI045 (261), EKI-785 (262), PD153035 (263), WZ4002 (200)), molecules never developed for their ability to inhibit EGFR in patients with NSCLC (e.g. BGB283 (264), brigatinib (265)), named secondary metabolites (AZD5104 (190), OSI420 (266)), and/or molecules with only abstract-level data (ASLAN001, ASP8273, CNX2006, epitinib, olmutinib, PD168393, PF06747775, and theliatinib).

References

1. Wain, H. M. *et al.* Guidelines for human gene nomenclature. *Genomics* **79,** 464–470 (2002).

2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66,** 7–30 (2016).

3. Travis, W. D., World Health Organization, International Agency for Research on Cancer, Cancer, I. A. F. T. S. O. L. & Pathology, I. A. O. *Pathology and genetics of tumours of the lung, pleura, thymus and heart*. (Lyon : IARC Press, Oxford : Oxford University Press, 2015).

4. Travis, W. D. *et al.* International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol* **6,** 244–285 (2011).

5. SEER Program (National Cancer Institute). *SEER cancer statistics review.* (Bethesda, MD: U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute, 2016).

6. Moore, K. L. & Dalley, A. F. *Clinically oriented anatomy*. (Philadelphia : Lippincott Williams & Wilkins, 2009).

7. Netter, F. H. *Atlas of human anatomy*. (Philadelphia, PA : Saunders/Elsevier, 2011).

8. Carbone, P. P., Forst, J. K., Feinstein, A. R., Higgins, G. A. & Selawry, O. S. Lung Cancer: Perspectives and Prospects. *Ann Intern Med* **73,** 1003–1024 (1970).

9. Hyde, L. & Hyde, C. I. Clinical manifestations of lung cancer. *Chest* **65,** 299–306 (1974).

10. Chute, C. G. *et al.* Presenting conditions of 1539 population-based lung cancer patients by cell type and stage in New Hampshire and Vermont. *Cancer* **56,** 2107–2111 (1985).

11. Buccheri, G. & Ferrigno, D. Lung cancer: clinical presentation and specialist referral time. *Eur. Respir. J.* **24,** 898–904 (2004).

12. Kocher, F. *et al.* Longitudinal analysis of 2293 NSCLC patients: a comprehensive study from the TYROL registry. *Lung Cancer* **87,** 193–200 (2015).

13. Howington, J. A., Blum, M. G., Chang, A. C., Balekian, A. A. & Murthy, S. C. Treatment of Stage I and II Non-small Cell Lung Cancer. *Chest* **143,** e278S–e313S (2013).

14. Ramnath, N. *et al.* Treatment of Stage III Non-small Cell Lung Cancer. *Chest* **143,** e314S–e340S (2013).

15. Socinski, M. *et al.* Treatment of Stage IV Non-small Cell Lung Cancer. *Chest* **143,** e341S–e368S (2013).

16. Zornosa, C. *et al.* First-line systemic therapy practice patterns and concordance with NCCN guidelines for patients diagnosed with metastatic NSCLC treated at NCCN institutions. *J Natl Compr Canc Netw* **10,** 847–856 (2012).

17. Sandler, A. *et al.* Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med* **355,** 2542–2550 (2006).

18. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45,** 228–247 (2009).

19. Johnson, J. R., Williams, G. & Pazdur, R. End points and United States Food and Drug Administration approval of oncology drugs. *J. Clin. Oncol.* **21,** 1404–1411 (2003).

20. Food and Drug Administration Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER). Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics. 1–19 (2015).

21. Schiller, J. H. *et al.* Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med* **346,** 92–98 (2002).

22. Smit, E. F. *et al.* Three-arm randomized study of two cisplatin-based regimens and paclitaxel plus gemcitabine in advanced non-small-cell lung cancer: a phase III trial of the European Organization for Research and Treatment of Cancer Lung Cancer Group--EORTC 08975. *J. Clin. Oncol.* **21,** 3909–3917 (2003).

23. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* **375,** 1823–1833 (2016).

24. Pass, H. I., Pass, H. I. L. C. & Cancer, I. A. F. T. S. O. L. *Principles & practice of lung cancer: the official reference text of the IASLC*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2010).

25. Garraway, L. A. & Sellers, W. R. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat. Rev. Cancer* **6,** 593–602 (2006).

26. Imielinski, M. *et al.* Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell* **150,** 1107–1120 (2012).

27. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339,** 1546–1558 (2013).

28. Thiagalingam, S. *et al.* Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nat Genet* **13,** 343–346 (1996).

29. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

30. Hu, H.-D. *et al.* Histological subtypes of solitary pulmonary nodules of adenocarcinoma and their clinical relevance. *J Thorac Dis* **5,** 841–846 (2013).

31. Meador, C. B. *et al.* Beyond histology: translating tumor genotypes into clinically effective targeted therapies. *Clin. Cancer Res.* **20,** 2264–2275 (2014).

32. Lundberg, A. S. *et al.* Immortalization and transformation of primary human airway epithelial cells by gene transfer. *Oncogene* **21,** 4577–4586 (2002).

33. Ramirez, R. D. *et al.* Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer Research* **64,** 9027–9034 (2004).

34. Kendall, S. D., Linardic, C. M., Adam, S. J. & Counter, C. M. A network of genetic events sufficient to convert normal human cells to a tumorigenic state. *Cancer Research* **65,** 9824–9828 (2005).

35. Sato, M. *et al.* Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Research* **66,** 2116–2128 (2006).

36. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347,** 78–81 (2015).

37. Wang, C., Lisanti, M. P. & Liao, D. J. Reviewing once more the c-myc and Ras collaboration: converging at the cyclin D1-CDK4 complex and challenging basic concepts of cancer biology. *Cell Cycle* **10,** 57–67 (2011).

38. Weinstein, I. B. Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science* **297,** 63–64 (2002).

39. Weinstein, I. B., Joe, A. & Felsher, D. Oncogene Addiction. *Cancer Research* **68,** 3077–3080 (2008).

40. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455,** 1069–1075 (2008).

41. Seo, J.-S. *et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Research* **22,** 2109–2119 (2012).

42. Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* **10,** 1243–1260 (2015).

43. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).

44. Tsao, A. S. *et al.* Scientific Advances in Lung Cancer 2015. *Journal of Thoracic Oncology* **11,** 613–638 (2016).

45. Carr, T. H. *et al.* Defining actionable mutations for oncology therapeutic development. *Nat. Rev. Cancer* **16,** 319–329 (2016).

46. Unni, A. M., Lockwood, W. W., Zejnullahu, K., Lee-Lin, S.-Q. & Varmus, H. Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. *Elife* **4,** e06907 (2015).

47. Lammers, P. E., Lovly, C. M. & Horn, L. A patient with metastatic lung adenocarcinoma harboring concurrent EGFR L858R, EGFR germline T790M, and PIK3CA mutations: the challenge of interpreting results of comprehensive mutational testing in lung cancer. *J Natl Compr Canc Netw* **12,** 6–11– quiz 11 (2014).

48. Klempner, S. J. *et al.* Emergence of RET rearrangement co-existing with activated EGFR mutation in EGFR-mutated NSCLC patients who had progressed on first- or second-generation EGFR TKI. *Lung Cancer* **89,** 357–359 (2015).

49. Kawano, O. *et al.* PIK3CA mutation status in Japanese lung cancer patients. *Lung Cancer* **54,** 209–215 (2006).

50. Levy, M. A., Lovly, C. M. & Pao, W. Translating genomic information into clinical medicine: Lung cancer as a paradigm. *Genome Research* **22,** 2101–2108 (2012).

51. Sequist, L. V. *et al.* Response to treatment and survival of patients with non-small cell lung cancer undergoing somatic EGFR mutation testing. *Oncologist* **12,** 90–98 (2007).

52. Hendler, F. J. & Ozanne, B. W. Human squamous cell lung cancers express increased epidermal growth factor receptors. *J. Clin. Invest.* **74,** 647–651 (1984).

53. Gill, G. N. *et al.* Monoclonal anti-EGF receptor antibodies which are inhibitors of epidermal growth factor binding and antagonists of epidermal growth factor binding and antagonists of epidermal growth factor-stimulated tyrosine protein kinase activity. *J. Biol. Chem.* **259,** 7755–7760 (1984).

54. Pirker, R. *et al.* Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. *Lancet* **373,** 1525–1531 (2009).

55. Lynch, T. J. *et al.* Cetuximab and first-line taxane/carboplatin chemotherapy in advanced non-small-cell lung cancer: results of the randomized multicenter phase III trial BMS099. *Journal of Clinical Oncology* **28,** 911–917 (2010).

56. Janjigian, Y. Y. *et al.* Dual inhibition of EGFR with afatinib and cetuximab in kinase inhibitor-resistant EGFR-mutant lung cancer with and without T790M mutations. *Cancer Discovery* **4,** 1036–1045 (2014).

57. Shepherd, F. A. *et al.* Erlotinib in previously treated non-small-cell lung cancer. *N Engl J Med* **353,** 123–132 (2005).

58. Thatcher, N. *et al.* Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* **366,** 1527–1537 (2005).

59. Moyer, J. D. *et al.* Induction of apoptosis and cell cycle arrest by CP-358,774, an inhibitor of epidermal growth factor receptor tyrosine kinase. *Cancer Research* **57,** 4838–4848 (1997).

60. Barker, A. J. *et al.* Studies leading to the identification of ZD1839 (IRESSA): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorg. Med. Chem. Lett.* **11,** 1911–1914 (2001).

61. Miller, V. A. *et al.* Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. *J. Clin. Oncol.* **22,** 1103–1109 (2004).

62. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361,** 947–957 (2009).

63. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304,** 1497–1500 (2004).

64. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350,** 2129–2139 (2004).

65. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 13306–13311 (2004).

66. Carey, K. D. *et al.* Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. *Cancer Research* **66,** 8163–8171 (2006).

67. Loong, H. H. & Mok, T. S. K. Dropping bars or rising hoops—phase III outcomes of NSCLC. *Nature Publishing Group* **11,** 381–382 (2014).

68. Maemondo, M. *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* **362,** 2380–2388 (2010).

69. Mitsudomi, T. *et al.* Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol.* **11,** 121–128 (2010).

70. Zhou, C. *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR

mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol.* **12,** 735–742 (2011).

71. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13,** 239–246 (2012).

72. Sequist, L. V. *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *Journal of Clinical Oncology* **31,** 3327–3334 (2013).

73. Wu, Y.-L. *et al.* Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol.* **15,** 213–222 (2014).

74. Wu, Y.-L. *et al.* First-line erlotinib versus gemcitabine/cisplatin in patients with advanced EGFR mutation-positive non-small-cell lung cancer: analyses from the phase III, randomized, open-label, ENSURE study. *Annals of Oncology* **26,** 1883–1889 (2015).

75. Park, K. *et al.* Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. *Lancet Oncology* 1–13 (2016).

76. Hansen, H. H. *et al.* Combination chemotherapy of advanced lung cancer: a randomized trial. *Cancer* **38,** 2201–2207 (1976).

77. Lin, J. J. *et al.* Five-Year Survival in EGFR-Mutant Metastatic Lung Adenocarcinoma Treated with EGFR-TKIs. *Journal of Thoracic Oncology* **11,** 556–565 (2016).

78. Lee, C. K. *et al.* Impact of EGFR inhibitor in non-small cell lung cancer on progression-free and overall survival: a meta-analysis. *J. Natl. Cancer Inst.* **105,** 595–605 (2013).

79. Haaland, B., Tan, P. S., de Castro, G. & Lopes, G. Meta-analysis of first-line therapies in advanced non-small-cell lung cancer harboring EGFR-activating mutations. *J Thorac Oncol* **9,** 805–811 (2014).

80. Kuan, F.-C. *et al.* Overall survival benefits of first-line EGFR tyrosine kinase inhibitors in EGFR-mutated non-small-cell lung cancers: a systematic review and meta-analysis. *Br. J. Cancer* **113,** 1519–1528 (2015).

81. Lindeman, N. I. *et al.* Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J Thorac Oncol* **8,** 823–859 (2013).

82. Tan, D. S. W. *et al.* The International Association for the Study of Lung Cancer Consensus Statement on Optimizing Management of EGFR Mutation-Positive Non-Small Cell Lung Cancer: Status in 2016. *Journal of Thoracic Oncology* **11,** 946–963 (2016).

83. Fukuoka, M. *et al.* Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). *Journal of Clinical Oncology* **29,** 2866–2874 (2011).

84. Inoue, A. *et al.* Updated overall survival results from a randomized phase III trial comparing gefitinib with carboplatin-paclitaxel for chemo-naïve non-small cell lung cancer with sensitive EGFR gene mutations (NEJ002). *Annals of Oncology* **24,** 54–59 (2013).

85. Yang, J. C.-H. *et al.* Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials. *Lancet Oncol.* **16,** 141–151 (2015).

86. Sebastian, M., Schmittel, A. & Reck, M. First-line treatment of EGFR-mutated nonsmall cell lung cancer: critical review on study methodology. *Eur Respir Rev* **23,** 92–105 (2014).

87. Gettinger, S. & Politi, K. PD-1 Axis Inhibitors in EGFR- and ALK-Driven Lung Cancer: Lost Cause? *Clin. Cancer Res.* **22,** 4539–4541 (2016).

88. Oxnard, G. R. *et al.* Acquired resistance to EGFR tyrosine kinase inhibitors in EGFR-mutant lung cancer: distinct natural history of patients with tumors harboring the T790M mutation. *Clin. Cancer Res.* **17,** 1616–1622 (2011).

89. Yu, H. A. *et al.* Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin. Cancer Res.* **19,** 2240–2247 (2013).

90. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2,** e73 (2005).

91. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* **352,** 786–792 (2005).

92. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Science Translational Medicine* **3,** 75ra26–75ra26 (2011).

93. Yu, H. A., Riely, G. J. & Lovly, C. M. Therapeutic strategies utilized in the setting of acquired resistance to EGFR tyrosine kinase inhibitors. *Clin. Cancer Res.* **20,** 5898–5907 (2014).

94. Lovly, C. M. & Shaw, A. T. Molecular pathways: resistance to kinase inhibitors and implications for therapeutic strategies. *Clin. Cancer Res.* **20,** 2249–2256 (2014).

95. Jänne, P. A. *et al.* AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med* **372,** 1689–1699 (2015).

96. Ullrich, A. *et al.* Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* **309,** 418–425 (1984).

97. Cohen, S., Ushiro, H., Stoscheck, C. & Chinkers, M. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. *J. Biol. Chem.* **257,** 1523–1531 (1982).

98. Lemmon, M. A. & Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141,** 1117–1134 (2010).

99. Lemmon, M. A., Schlessinger, J. & Ferguson, K. M. The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harbor Perspectives in Biology* **6,** a020768–a020768 (2014).

100. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298,** 1912–1934 (2002).

101. Jura, N. *et al.* Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. *Molecular Cell* **42,** 9–22 (2011).

102. Gotoh, N., Tojo, A., Hino, M., Yazaki, Y. & Shibuya, M. A highly conserved tyrosine residue at codon 845 within the kinase domain is not required for the transforming activity of human epidermal growth

factor receptor. *Biochemical and Biophysical Research Communications* **186,** 768–774 (1992).

103. Tice, D. A., Biscardi, J. S., Nickles, A. L. & Parsons, S. J. Mechanism of biological synergy between cellular Src and epidermal growth factor receptor. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 1415–1420 (1999).

104. Cohen, S. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. *J. Biol. Chem.* **237,** 1555–1562 (1962).

105. Sibilia, M. *et al.* The epidermal growth factor receptor: from development to tumorigenesis. *Differentiation* **75,** 770–787 (2007).

106. Schechter, A. L. *et al.* The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. *Nature* **312,** 513–516 (1984).

107. Kraus, M. H., Issing, W., Miki, T., Popescu, N. C. & Aaronson, S. A. Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 9193–9197 (1989).

108. Plowman, G. D. *et al.* Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal growth factor receptor family. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 1746–1750 (1993).

109. Huang, Y. *et al.* Molecular basis for multimerization in the activation of the epidermal growth factor receptor. *Elife* **5,** 18756 (2016).

110. Lemmon, M. A. *et al.* Two EGF molecules contribute additively to stabilization of the EGFR dimer. *EMBO J.* **16,** 281–294 (1997).

111. Schlessinger, J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110,** 669–672 (2002).

112. Harris, R. C., Chung, E. & Coffey, R. J. EGF receptor ligands. *Exp. Cell Res.* **284,** 2–13 (2003).

113. Roskoski, R., Jr. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacological Research* **79,** 34–74 (2014).

114. Beerli, R. R. & Hynes, N. E. Epidermal growth factor-related peptides activate distinct subsets of ErbB receptors and differ in their biological activities. *J. Biol. Chem.* **271,** 6071–6076 (1996).

115. Leahy, D. J. Structure and function of the epidermal growth factor (EGF/ErbB) family of receptors. *Adv. Protein Chem.* **68,** 1–27 (2004).

116. Kovacs, E., Zorn, J. A., Huang, Y., Barros, T. & Kuriyan, J. A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor. *Annu. Rev. Biochem.* **84,** 739–764 (2015).

117. Jorissen, R. N. *et al.* Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp. Cell Res.* **284,** 31–53 (2003).

118. Honegger, A. M., Schmidt, A., Ullrich, A. & Schlessinger, J. Evidence for epidermal growth factor (EGF)-induced intermolecular autophosphorylation of the EGF receptors in living cells. *Mol. Cell. Biol.* **10,** 4035–4044 (1990).

119. Jura, N. *et al.* Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell* **137,** 1293–1307 (2009).

120. Red Brewer, M. *et al.* The juxtamembrane region of the EGF receptor functions as an activation

domain. *Molecular Cell* **34,** 641–651 (2009).

121. Hubbard, S. R. & Till, J. H. Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* **69,** 373–398 (2000).

122. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109,** 275–282 (2002).

123. Kornev, A. P. & Taylor, S. S. Defining the conserved internal architecture of a protein kinase. *Biochim. Biophys. Acta* **1804,** 440–444 (2010).

124. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

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

126. Kornev, A. P., Haste, N. M., Taylor, S. S. & Eyck, L. F. T. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 17783–17788 (2006).

127. Kornev, A. P., Taylor, S. S. & Eyck, Ten, L. F. A helix scaffold for the assembly of active protein kinases. *Proceedings of the National Academy of Sciences* **105,** 14377–14382 (2008).

128. Endicott, J. A., Noble, M. E. M. & Johnson, L. N. The structural basis for control of eukaryotic protein kinases. *Annu. Rev. Biochem.* **81,** 587–613 (2012).

129. Burgess, A. W. *et al.* An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Molecular Cell* **12,** 541–552 (2003).

130. Garrett, T. P. J. *et al.* Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* **110,** 763–773 (2002).

131. Ogiso, H. *et al.* Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110,** 775–787 (2002).

132. Dawson, J. P. *et al.* Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. *Mol. Cell. Biol.* **25,** 7734–7742 (2005).

133. Cho, H.-S. & Leahy, D. J. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* **297,** 1330–1333 (2002).

134. Ferguson, K. M. *et al.* EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization. *Molecular Cell* **11,** 507–517 (2003).

135. Mattoon, D., Klein, P., Lemmon, M. A., Lax, I. & Schlessinger, J. The tethered configuration of the EGF receptor extracellular domain exerts only a limited control of receptor function. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 923–928 (2004).

136. Walker, F. *et al.* CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor. *J. Biol. Chem.* **279,** 22387–22398 (2004).

137. Elleman, T. C. *et al.* Identification of a determinant of epidermal growth factor receptor ligand-binding specificity using a truncated, high-affinity form of the ectodomain. *Biochemistry* **40,** 8930–8939 (2001).

138. Schlessinger, J. Autoinhibition Control. *Science* **300,** 750–752 (2003).

139. Kohda, D. *et al.* A 40-kDa epidermal growth factor/transforming growth factor alpha-binding domain produced by limited proteolysis of the extracellular domain of the epidermal growth factor receptor. *J. Biol. Chem.* **268,** 1976–1981 (1993).

140. Li, S. *et al.* Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7,** 301–311 (2005).

141. Sunada, H., Magun, B. E., Mendelsohn, J. & MacLeod, C. L. Monoclonal antibody against epidermal growth factor receptor is internalized without stimulating receptor phosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* **83,** 3825–3829 (1986).

142. Kurai, J. *et al.* Antibody-dependent cellular cytotoxicity mediated by cetuximab against lung cancer cell lines. *Clin. Cancer Res.* **13,** 1552–1561 (2007).

143. Burke, C. L., Lemmon, M. A., Coren, B. A., Engelman, D. M. & Stern, D. F. Dimerization of the p185neu transmembrane domain is necessary but not sufficient for transformation. *Oncogene* **14,** 687–696 (1997).

144. Leo C Groenen, Francesca Walker, Antony W Burgess, A. & Treutlein, H. R. A Model for the Activation of the Epidermal Growth Factor Receptor Kinase: Involvement of an Asymmetric Dimer? †. *Biochemistry* **36,** 3826–3836 (1997).

145. Clayton, A. H. A. *et al.* Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis. *J. Biol. Chem.* **280,** 30392–30399 (2005).

146. Clayton, A. H. A., Orchard, S. G., Nice, E. C., Posner, R. G. & Burgess, A. W. Predominance of activated EGFR higher-order oligomers on the cell surface. *Growth Factors* **26,** 316–324 (2008).

147. Kozer, N. *et al.* Exploring higher-order EGFR oligomerisation and phosphorylation--a combined experimental and theoretical approach. *Mol. BioSyst.* **9,** 1849–1863 (2013).

148. Chung, I. *et al.* Spatial control of EGF receptor activation by reversible dimerization on living cells. *Nature* **464,** 783–787 (2010).

149. Singh, B. & Coffey, R. J. Trafficking of epidermal growth factor receptor ligands in polarized epithelial cells. *Annu. Rev. Physiol.* **76,** 275–300 (2014).

150. Li, X. & Carthew, R. W. A microRNA Mediates EGF Receptor Signaling and Promotes Photoreceptor Differentiation in the Drosophila Eye. *Cell* **123,** 1267–1277 (2005).

151. Breindel, J. L. *et al.* EGF receptor activates MET through MAPK to enhance non-small cell lung carcinoma invasion and brain metastasis. *Cancer Research* **73,** 5053–5065 (2013).

152. Paulsen, C. E. *et al.* Peroxide-dependent sulfenylation of the EGFR catalytic site enhances kinase activity. *Nature Chemical Biology* **8,** 57–64 (2011).

153. Runkle, K. B. *et al.* Inhibition of DHHC20-Mediated EGFR Palmitoylation Creates a Dependence on EGFR Signaling. *Molecular Cell* **62,** 385–396 (2016).

154. Mendrola, J. M., Berger, M. B., King, M. C. & Lemmon, M. A. The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J. Biol. Chem.* **277,** 4704–4712 (2002).

155. Mineev, K. S. *et al.* Spatial structure of the transmembrane domain heterodimer of ErbB1 and ErbB2 receptor tyrosine kinases. *Journal of Molecular Biology* **400,** 231–243 (2010).

156. Arkhipov, A. *et al.* Architecture and membrane interactions of the EGF receptor. *Cell* **152,** 557–569 (2013).

157. Endres, N. F. *et al.* Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. *Cell* **152,** 543–556 (2013).

158. Bessman, N. J., Bagchi, A., Ferguson, K. M. & Lemmon, M. A. Complex relationship between ligand binding and dimerization in the epidermal growth factor receptor. *CellReports* **9,** 1306–1317 (2014).

159. Thiel, K. W. & Carpenter, G. Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. *Proceedings of the National Academy of Sciences* **104,** 19238–19243 (2007).

160. Needham, S. R. *et al.* EGFR oligomerization organizes kinase-active dimers into competent signalling platforms. *Nature Communications* **7,** 13307 (2016).

161. Kovacs, E. *et al.* Analysis of the Role of the C-Terminal Tail in the Regulation of the Epidermal Growth Factor Receptor. *Mol. Cell. Biol.* **35,** 3083–3102 (2015).

162. Kozer, N. *et al.* Recruitment of the adaptor protein Grb2 to EGFR tetramers. *Biochemistry* **53,** 2594–2604 (2014).

163. Wagner, M. J., Stacey, M. M., Liu, B. A. & Pawson, T. Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling. *Cold Spring Harbor Perspectives in Biology* **5,** a008987–a008987 (2013).

164. Kaszuba, K. *et al.* N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 4334–4339 (2015).

165. Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439,** 168–174 (2006).

166. Schlessinger, J. & Lemmon, M. A. SH2 and PTB domains in tyrosine kinase signaling. *Sci. STKE* **2003,** RE12–re12 (2003).

167. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2,** 127–137 (2001).

168. Yarden, Y. & Pines, G. The ERBB network: at last, cancer therapy meets systems biology. *Nat. Rev. Cancer* **12,** 553–563 (2012).

169. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* **1,** 2005.0010–E17 (2005).

170. Michael, N. & Jura, N. Src defines a new pool of EGFR substrates. *Nature Publishing Group* **22,** 945–947 (2015).

171. Verma, N. *et al.* Targeting of PYK2 Synergizes with EGFR Antagonists in Basal-like TNBC and Circumvents HER3-Associated Resistance via the NEDD4-NDRG1 Axis. *Cancer Research* (2016).

172. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7,** 169–181 (2007).

173. Downward, J. *et al.* Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. *Nature* **307,** 521–527 (1984).

174. Thompson, D. M. & Gill, G. N. The EGF receptor: structure, regulation and potential role in malignancy. *Cancer Surv.* **4,** 767–788 (1985).

175. Arteaga, C. L. & Engelman, J. A. ERBB Receptors: From Oncogene Discovery to Basic Science to Mechanism-Based Cancer Therapeutics. *Cancer Cell* **25,** 282–303 (2014).

176. Moore, M. J. *et al.* Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *Journal of Clinical Oncology* **25,** 1960–1966 (2007).

177. Vermorken, J. B. *et al.* Platinum-based chemotherapy plus cetuximab in head and neck cancer. *N Engl J Med* **359,** 1116–1127 (2008).

178. Van Cutsem, E. *et al.* Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med* **360,** 1408–1417 (2009).

179. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research* **43,** D805–D811 (2015).

180. Yun, C.-H. *et al.* Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **11,** 217–227 (2007).

181. Wang, Z. *et al.* Mechanistic insights into the activation of oncogenic forms of EGF receptor. *Nature Publishing Group* **18,** 1388–1393 (2011).

182. Gajiwala, K. S. *et al.* Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. *Structure* **21,** 209–219 (2013).

183. Shan, Y. *et al.* Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell* **149,** 860–870 (2012).

184. Red Brewer, M. *et al.* Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proceedings of the National Academy of Sciences* **110,** E3595–604 (2013).

185. Kim, Y. *et al.* Temporal resolution of autophosphorylation for normal and oncogenic forms of EGFR and differential effects of gefitinib. *Biochemistry* **51,** 5212–5222 (2012).

186. Valley, C. C. *et al.* Enhanced dimerization drives ligand-independent activity of mutant epidermal growth factor receptor in lung cancer. *Molecular Biology of the Cell* **26,** 4087–4099 (2015).

187. Foster, S. A. *et al.* Activation Mechanism of Oncogenic Deletion Mutations in BRAF, EGFR, and HER2. *Cancer Cell* 1–49 (2016).

188. Li, D. *et al.* BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene* **27,** 4702–4711 (2008).

189. Walter, A. O. *et al.* Discovery of a mutant-selective covalent inhibitor of EGFR that overcomes T790M-mediated resistance in NSCLC. *Cancer Discovery* **3,** 1404–1415 (2013).

190. Cross, D. A. E. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discovery* **4,** 1046–1061 (2014).

191. Ward, W. H. *et al.* Epidermal growth factor receptor tyrosine kinase. Investigation of catalytic mechanism, structure-based searching and discovery of a potent inhibitor. *Biochem. Pharmacol.* **48,** 659–666 (1994).

192. Qiu, C. *et al.* In vitro enzymatic characterization of near full length EGFR in activated and inhibited states. *Biochemistry* **48,** 6624–6632 (2009).

193. Stamos, J., Sliwkowski, M. X. & Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **277,** 46265–46272 (2002).

194. Wood, E. R. *et al.* A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Research* **64,** 6652–6659 (2004).

195. Yun, C.-H. *et al.* The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences* **105,** 2070–2075 (2008).

196. Kwak, E. L. *et al.* Irreversible inhibitors of the EGF receptor may circumvent acquired resistance to gefitinib. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 7665–7670 (2005).

197. Engelman, J. A. *et al.* PF00299804, an irreversible pan-ERBB inhibitor, is effective in lung cancer models with EGFR and ERBB2 mutations that are resistant to gefitinib. *Cancer Research* **67,** 11924–11932 (2007).

198. Miller, V. A. *et al.* Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): a phase 2b/3 randomised trial. *Lancet Oncol.* **13,** 528–538 (2012).

199. Ramalingam, S. S. *et al.* Dacomitinib versus erlotinib in patients with advanced-stage, previously treated non-small-cell lung cancer (ARCHER 1009): a randomised, double-blind, phase 3 trial. *Lancet Oncol.* **15,** 1369–1378 (2014).

200. Zhou, W. *et al.* Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* **462,** 1070–1074 (2009).

201. Sequist, L. V. *et al.* Rociletinib in EGFR-mutated non-small-cell lung cancer. *N Engl J Med* **372,** 1700–1709 (2015).

202. Eberlein, C. A. *et al.* Acquired Resistance to the Mutant-Selective EGFR Inhibitor AZD9291 Is Associated with Increased Dependence on RAS Signaling in Preclinical Models. *Cancer Research* **75,** 2489–2500 (2015).

203. Vivanco, I. *et al.* Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discovery* **2,** 458–471 (2012).

204. Dar, A. C. & Shokat, K. M. The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling. *Annu. Rev. Biochem.* **80,** 769–795 (2011).

205. Burris, H. A. Phase I Safety, Pharmacokinetics, and Clinical Activity Study of Lapatinib (GW572016), a Reversible Dual Inhibitor of Epidermal Growth Factor Receptor Tyrosine Kinases, in Heavily Pretreated Patients With Metastatic Carcinomas. *Journal of Clinical Oncology* **23,** 5305–5313 (2005).

206. Zejnullahu, K. *et al.* Disparate effects of gefitinib and lapatinib on egfr mutant lung cancer. *Cancer Research* **68,** 3636–3636 (2008).

207. Ross, H. J. *et al.* Randomized phase II multicenter trial of two schedules of lapatinib as first- or second-line monotherapy in patients with advanced or metastatic non-small cell lung cancer. *Clin. Cancer Res.* **16,** 1938–1949 (2010).

208. Ramlau, R. *et al.* Phase I Study of Lapatinib and Pemetrexed in the Second-Line Treatment of Advanced or Metastatic Non-Small-Cell Lung Cancer With Assessment of Circulating Cell Free Thymidylate Synthase RNA as a Potential Biomarker. *Clin Lung Cancer* **16,** 348–357 (2015).

209. Schindler, T. *et al.* Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289,** 1938–1942 (2000).

210. Liu, Y. & Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nature Chemical Biology* **2,** 358–364 (2006).

211. Kufareva, I. & Abagyan, R. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. *J. Med. Chem.* **51,** 7921–7932 (2008).

212. Novotny, C. J. *et al.* Overcoming resistance to HER2 inhibitors through state-specific kinase binding. *Nature Chemical Biology* 1–10 (2016).

213. Roskoski, R., Jr. Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacological Research* **103,** 26–48 (2016).

214. Morgensztern, D., Politi, K. & Herbst, R. S. EGFR Mutations in Non-Small-Cell Lung Cancer: Find, Divide, and Conquer. *JAMA Oncol* **1,** 146–148 (2015).

215. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* **2,** 401–404 (2012).

216. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* **6,** pl1–pl1 (2013).

217. Pao, W. & Ladanyi, M. Epidermal growth factor receptor mutation testing in lung cancer: searching for the ideal method. *Clin. Cancer Res.* **13,** 4954–4955 (2007).

218. Han, S.-W. *et al.* Predictive and prognostic impact of epidermal growth factor receptor mutation in non-small-cell lung cancer patients treated with gefitinib. *J. Clin. Oncol.* **23,** 2493–2501 (2005).

219. Yang, C.-H. *et al.* Specific EGFR mutations predict treatment outcome of stage IIIB/IV patients with chemotherapy-naive non-small-cell lung cancer receiving first-line gefitinib monotherapy. *Journal of Clinical Oncology* **26,** 2745–2753 (2008).

220. Pao, W. & Chmielecki, J. Rational, biologically based treatment of EGFR-mutant non-small-celllung cancer. 1–15 (2010).

221. Massarelli, E., Johnson, F. M., Erickson, H. S., Wistuba, I. I. & Papadimitrakopoulou, V. Uncommon epidermal growth factor receptor mutations in non-small cell lung cancer and their mechanisms of EGFR tyrosine kinase inhibitors sensitivity and resistance. *Lung Cancer* **80,** 235–241 (2013).

222. Tan, F. *et al.* Icotinib (BPI-2009H), a novel EGFR tyrosine kinase inhibitor, displays potent efficacy in preclinical studies. *Lung Cancer* **76,** 177–182 (2012).

223. Shi, Y. *et al.* Icotinib versus gefitinib in previously treated advanced non-small-cell lung cancer (ICOGEN): a randomised, double-blind phase 3 non-inferiority trial. *Lancet Oncology* **14,** 953–961 (2013).

224. Barlaam, B. *et al.* Discovery of AZD8931, an Equipotent, Reversible Inhibitor of Signaling by EGFR, HER2, and HER3 Receptors. *ACS Med Chem Lett* **4,** 742–746 (2013).

225. Tjulandin, S. *et al.* Phase I, dose-finding study of AZD8931, an inhibitor of EGFR (erbB1), HER2

(erbB2) and HER3 (erbB3) signaling, in patients with advanced solid tumors. *Invest New Drugs* **32,** 145–153 (2014).

226. Gendreau, S. B. *et al.* Inhibition of the T790M Gatekeeper Mutant of the Epidermal Growth Factor Receptor by EXEL-7647. *Clin. Cancer Res.* **13,** 3713–3723 (2007).

227. Pietanza, M. C. *et al.* XL647--a multitargeted tyrosine kinase inhibitor: results of a phase II study in subjects with non-small cell lung cancer who have progressed after responding to treatment with either gefitinib or erlotinib. *J Thorac Oncol* **7,** 219–226 (2012).

228. Wedge, S. R. *et al.* ZD6474 inhibits vascular endothelial growth factor signaling, angiogenesis, and tumor growth following oral administration. *Cancer Research* **62,** 4645–4655 (2002).

229. Herbst, R. S. *et al.* Vandetanib plus docetaxel versus docetaxel as second-line treatment for patients with advanced non-small-cell lung cancer (ZODIAC): a double-blind, randomised, phase 3 trial. *Lancet Oncology* **11,** 619–626 (2010).

230. Traxler, P. *et al.* AEE788: A dual family epidermal growth factor receptor/ErbB2 and vascular endothelial growth factor receptor tyrosine kinase inhibitor with antitumor and antiangiogenic activity. *Cancer Research* **64,** 4931–4941 (2004).

231. Baselga, J. *et al.* Using pharmacokinetic and pharmacodynamic data in early decision making regarding drug development: a phase I clinical trial evaluating tyrosine kinase inhibitor, AEE788. *Clin. Cancer Res.* **18,** 6364–6372 (2012).

232. Zeng, Q. *et al.* Discovery and Evaluation of Clinical Candidate AZD3759, a Potent, Oral Active, Central Nervous System-Penetrant, Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor. *J. Med. Chem.* **58,** 8200–8215 (2015).

233. Yang, Z. *et al.* AZD3759, a BBB-penetrating EGFR inhibitor for the treatment of EGFR mutant NSCLC with CNS metastases. *Science Translational Medicine* **8,** 368ra172 (2016).

234. Fink, B. E. *et al.* New dual inhibitors of EGFR and HER2 protein tyrosine kinases. *Bioorg. Med. Chem. Lett.* **15,** 4774–4779 (2005).

235. Gavai, A. V. *et al.* Discovery and preclinical evaluation of [4-[[1-(3-fluorophenyl)methyl]-1H-indazol-5-ylamino]-5-methylpyrrolo[2,1-f][1,2,4]triazin-6-yl]carbamic acid, (3S)-3-morpholinylmethyl ester (BMS-599626), a selective and orally efficacious inhibitor of human epidermal growth factor receptor 1 and 2 kinases. *J. Med. Chem.* **52,** 6527–6530 (2009).

236. Soria, J.-C. *et al.* Phase I safety, pharmacokinetic and pharmacodynamic trial of BMS-599626 (AC480), an oral pan-HER receptor tyrosine kinase inhibitor, in patients with advanced solid tumors. *Ann. Oncol.* **23,** 463–471 (2012).

237. Wong, T. W. *et al.* Antitumor and antiangiogenic activities of BMS-690514, an inhibitor of human EGF and VEGF receptor kinase families. *Clin. Cancer Res.* **17,** 4031–4041 (2011).

238. Soria, J.-C. *et al.* Phase I-IIa study of BMS-690514, an EGFR, HER-2 and -4 and VEGFR-1 to -3 oral tyrosine kinase inhibitor, in patients with advanced or metastatic solid tumours. *Eur. J. Cancer* **49,** 1815–1824 (2013).

239. Lai, C.-J. *et al.* CUDC-101, a multitargeted inhibitor of histone deacetylase, epidermal growth factor receptor, and human epidermal growth factor receptor 2, exerts potent anticancer activity. *Cancer Research* **70,** 3647–3656 (2010).

240. Shimizu, T. *et al.* Phase I first-in-human study of CUDC-101, a multitargeted inhibitor of HDACs,

EGFR, and HER2 in patients with advanced solid tumors. *Clin. Cancer Res.* **20,** 5032–5040 (2014).

241. Fry, D. W. *et al.* Specific, irreversible inactivation of the epidermal growth factor receptor and erbB2, by a new class of tyrosine kinase inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 12022–12027 (1998).

242. Jänne, P. A. *et al.* Multicenter, randomized, phase II trial of CI-1033, an irreversible pan-ERBB inhibitor, for previously treated advanced non small-cell lung cancer. *Journal of Clinical Oncology* **25,** 3936–3944 (2007).

243. Rabindran, S. K. *et al.* Antitumor activity of HKI-272, an orally active, irreversible inhibitor of the HER-2 tyrosine kinase. *Cancer Research* **64,** 3958–3965 (2004).

244. Sequist, L. V. *et al.* Neratinib, an Irreversible Pan-ErbB Receptor Tyrosine Kinase Inhibitor: Results of a Phase II Trial in Patients With Advanced Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **28,** 3076–3083 (2010).

245. Wissner, A. *et al.* Synthesis and Structure−Activity Relationships of 6,7-Disubstituted 4-Anilinoquinoline-3-carbonitriles. The Design of an Orally Active, Irreversible Inhibitor of the Tyrosine Kinase Activity of the Epidermal Growth Factor Receptor (EGFR) and the Human Epidermal Growth Factor Receptor-2 (HER-2). *J. Med. Chem.* **46,** 49–63 (2003).

246. Erlichman, C. *et al.* Phase I Study of EKB-569, an Irreversible Inhibitor of the Epidermal Growth Factor Receptor, in Patients With Advanced Solid Tumors. *J. Clin. Oncol.* **24,** 2252–2260 (2006).

247. Xie, H. *et al.* AST1306, a novel irreversible inhibitor of the epidermal growth factor receptor 1 and 2, exhibits antitumor activity both in vitro and in vivo. *PLoS ONE* **6,** e21487 (2011).

248. Zhang, J. *et al.* A phase I study of AST1306, a novel irreversible EGFR and HER2 kinase inhibitor, in patients with advanced solid tumors. *J Hematol Oncol* **7,** 22 (2014).

249. Nam, H.-J. *et al.* Antitumor activity of HM781-36B, an irreversible Pan-HER inhibitor, alone or in combination with cytotoxic chemotherapeutic agents in gastric cancer. *Cancer Lett.* **302,** 155–165 (2011).

250. Noh, Y.-H., Lim, H.-S., Jung, J.-A., Song, T. H. & Bae, K.-S. Population pharmacokinetics of HM781-36 (poziotinib), pan-human EGF receptor (HER) inhibitor, and its two metabolites in patients with advanced solid malignancies. *Cancer Chemother. Pharmacol.* **75,** 97–109 (2015).

251. Ishikawa, T. *et al.* Design and synthesis of novel human epidermal growth factor receptor 2 (HER2)/epidermal growth factor receptor (EGFR) dual inhibitors bearing a pyrrolo[3,2-d]pyrimidine scaffold. *J. Med. Chem.* **54,** 8030–8050 (2011).

252. LoRusso, P. *et al.* Phase 1 dose-escalation, pharmacokinetic, and cerebrospinal fluid distribution study of TAK-285, an investigational inhibitor of EGFR and HER2. *Invest New Drugs* **32,** 160–170 (2014).

253. Ward, R. A. *et al.* Structure- and Reactivity-Based Development of Covalent Inhibitors of the Activating and Gatekeeper Mutant Forms of the Epidermal Growth Factor Receptor (EGFR). *J. Med. Chem.* **56,** 7025–7048 (2013).

254. Mok, T. S. *et al.* Osimertinib or Platinum–Pemetrexed in EGFRT790M–Positive Lung Cancer. *N Engl J Med* NEJMoa1612674–12 (2016).

255. Xu, X. *et al.* AC0010, an Irreversible EGFR Inhibitor Selectively Targeting Mutated EGFR and Overcoming T790M-Induced Resistance in Animal Models and Lung Cancer Patients. *Molecular Cancer Therapeutics* **15,** 2586–2597 (2016).

256. Zhang, L., Zhao, H., Hu, B., Jiang, J. & Zheng, X. First-in-human study of AC0010, a novel irreversible, mutant-selective EGFR inhibitor in patients with 1st generation EGFR TKI-resistant non-small cell lung cancer. *Annals of Oncology* **27,** vi114–vi135 (2016).

257. Jia, Y. *et al.* EGF816 Exerts Anticancer Effects in Non-Small Cell Lung Cancer by Irreversibly and Selectively Targeting Primary and Acquired Activating Mutations in the EGF Receptor. *Cancer Research* **76,** 1591–1602 (2016).

258. Shao-Weng Tan, D. *et al.* P3.02b-117 Phase Ib Results from a Study of Capmatinib (INC280) + EGF816 in Patients with EGFR-Mutant Non-Small Cell Lung Cancer (NSCLC) Topic: EGFR RES. *Journal of Thoracic Oncology* **12,** S1264–S1265 (2017).

259. Umezawa, H. *et al.* Studies on a new epidermal growth factor-receptor kinase inhibitor, erbstatin, produced by MH435-hF3. *J. Antibiot.* **39,** 170–173 (1986).

260. Levitzki, A. & Gazit, A. Tyrosine Kinase Inhibition - an Approach to Drug Development. *Science* **267,** 1782–1788 (1995).

261. Jia, Y. *et al.* Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature* **534,** 129–132 (2016).

262. Discafani, C. M. *et al.* Irreversible inhibition of epidermal growth factor receptor tyrosine kinase with in vivo activity by N-[4-[(3-bromophenyl)amino]-6-quinazolinyl]-2-butynamide (CL-387,785). *Biochem. Pharmacol.* **57,** 917–925 (1999).

263. Fry, D. W. *et al.* A specific inhibitor of the epidermal growth factor receptor tyrosine kinase. *Science* **265,** 1093–1095 (1994).

264. Tang, Z. *et al.* BGB-283, a Novel RAF Kinase and EGFR Inhibitor, Displays Potent Antitumor Activity in BRAF-Mutated Colorectal Cancers. *Molecular Cancer Therapeutics* **14,** 2187–2197 (2015).

265. Ceccon, M. *et al.* Treatment Efficacy and Resistance Mechanisms Using the Second-Generation ALK Inhibitor AP26113 in Human NPM-ALK-Positive Anaplastic Large Cell Lymphoma. *Mol. Cancer Res.* **13,** 775–783 (2015).

266. Czejka, M., Sahmanovic, A., Buchner, P., Steininger, T. & Dittrich, C. Disposition of Erlotinib and Its Metabolite OSI420 in a Patient with High Bilirubin Levels. *Case Rep Oncol* **6,** 602–608 (2013).

Notes

[a] While human gene symbols generally are italicized with all letters in uppercase (*EGFR*), I will refer to genes in uppercase only (EGFR). As opposed to nomenclature guidelines (1), proteins also will be designated in non-italicized uppercase. This simplified all-caps notation is meant to circumvent the ambiguity related to interchanging references of EGFR alterations, constructs, and mutants.

[b] Patients usually are asymptomatic in the early stages of the disease because of the sparse pain fiber innervation in the lungs and because adenocarcinomas most often originate in the periphery of the lungs—leaving patients with sufficient respiratory reserve (6,7). This presentation contrasts with that of SCLC and squamous NSCLC; cough, dyspnea, and hemoptysis present most frequently in patients with squamous cell and SCL carcinomas because of these tumors' tendency to involve central airways (8-12).

[c] The most used platinum doublets against lung adenocarcinoma are cisplatin/carboplatin+pemetrexed and carboplatin+paclitaxel (16). The latter is sometimes combined with bevacizumab, an anti-angiogenic therapeutic antibody, which helps boost response rates and survival (17).

[d] Defined as the percent of patients experiencing objectively decreased tumor size per repeat imaging of the same tumor. ORR is typically the combined percentage of patients experiencing a complete response (CR; loss of detectable tumor per repeat imaging of the same tumor) and partial response (PR; loss of ≥ 30% of tumor per repeat imaging of the same tumor) (18).

[e] Defined as the average time between diagnosis (or treatment initiation) and death (19,20).

[f] Due to their signaling redundancy and synthetic lethality (46), these mutations are rarely co-occurring. Still, there are case reports of co-occurring driver alterations in NSCLC (47,48), especially with PIK3CA (49).

[g] Anti-EGFR antibodies have demonstrated efficacy against lung cancer cells in vitro for decades (53). However, for reasons that are not completely understood, anti-EGFR therapeutic antibodies have proven less efficacious against NSCLC in the clinic (54,55). Still, there may be a place for anti-EGFR antibodies treating defined molecular cohorts of NSCLC (56).

[h] Also known as time to progression (TTP); defined as the average time between diagnosis (or treatment initiation) and recurrence (radiological or clinical) (19,20). PFS is increasingly used as an endpoint in molecular-based clinical trials, where demonstration of efficacy is the main goal (67).

[i] A measure of how often a particular event happens in one group compared to how often it happens in the control group. In this case, the risk of an event—progression or death—while being treated with gefitinib versus (control) platinum doublets at 12 months is 0.48—just under 50%.

[j] This is to say: patients with EGFR-mutant lung adenocarcinoma treated with a first generation EGFR-TKI (erlotinib or gefitinib) have the same overall survival as matched patients (EGFR-mutant lung adenocarcinoma) treated with chemotherapy. It is unclear whether this finding is due to patient selection, poor trial design, and/or patient crossover (86). It should also be noted that patients with EGFR-mutant NSCLC have a longer OS than patients with EGFR-mutant-negative NSCLC, regardless of treatment modality (83-85).

[k] Immunotherapy (such as PD/L-1 inhibitors) has shown promising results (23) with respect to increasing OS in patients with NSCLC (partly because the trials are designed/powered to test OS); however, the role of these agents in EGFR-mutant lung adenocarcinoma remains contentious (87).

[l] In fact, in various unsupervised clustering analyses, EGFR is more closely related to ACK or SRC than to any RTK (100). EGFR's inactive confirmation bears striking resemblance to inactive CDK and/or SRC (101).

[m] Still, in most mouse models, defects are seen in bone, brain, heart, and various epithelia—notably in the skin, hair, lungs, and eyes. Indeed, EGF was originally identified as a protein that induces early eyelid opening

and teeth eruption in mice (104).

[n] This is the classical view accepted by the field. However, emerging evidence suggests that EGFR may be forming higher order multimers (109).

[o] Extracellular domains I and III alternatively are referred to as LR1 and LR2—the LR being short for leucine-rich (115).

[p] In other words, the EGFR TKD primarily catalyzes trans-autophosphorylation (118). As such, throughout this work, the measurement of EGFR's phosphotyrosines is used as a readout of EGFR activity.

[q] The DFG residues relative orientation (compared to the a-loop) does change; they simply remain in the catalytic site in both the in/active confirmations.

[r] For this reason, it was thought that EGFR was always in the active conformation and that dimerization of the receptor simply enabled trans-phosphorylation of the C-terms of two receptors (101).

[s] Although exposure of the dimerization arm is an important part of the HER family activation mechanism, it is not sufficient. Mutations that disrupt the domain II–IV inhibitory tether do not activate EGFR (135,136). Moreover, deleting domain IV (and thus constitutively exposing the dimerization arm) does not cause ligand-independent dimerization of EGFR (137).

[t] Most of the binding energy between EGF and EGFR is mediated by domain III (139), which is also the binding cite for cetuximab, the anti-EGFR therapeutic antibody (140). While cetuximab mainly inhibits EGFR signaling by preventing ligand binding and receptor-mediated dimerization, there is also a role for the antibody in attenuating signal by causing internalization/downregulation of the receptor and immune-mediated effects (141,142).

[u] These interactions were postulated to exist long before the crystal structure of EGFR was solved (144).

[v] This is not to mention the myriad cellular processes involved in modifying EGFR signaling in the cell including receptor internalization (149), miRNA modulation of enzyme levels (150), and/or cross talk with other receptor types (151). Beyond phosphorylation, glycosylation, and ubiquitylation, unique post-translational modifications also effect EGFR's activity, with sulfenylation (152) and palmitoylation (153) being but two examples. This is to say: EGFR activity is regulated in a multitude of ways beyond autophosphorylation.

[w] By extension, this increases receptor activity by removing inhibitory proximal C-term interactions (161) and providing additional sites for adaptor binding.

[x] Alternatively, phosphatidylinositol-4,5-bisphosphate (PI4,5-P_2), phosphatidylinositol-3,4, 5-triphosphate (PI3,4,5-P_3).

[y] Robust readouts of MAPK and/or PI3K/AKT activation are used as a readout of EGFR downstream signaling throughout this work.

[z] EGFR is mutated in ≤ 2 % of these cancers, and few of these mutations are recurrent (179).

[aa] Besides erlotinib and gefitinib, most other FDA-approved reversible TKIs bind to the inactive confirmation of their targets (182,213).

CHAPTER II: EGFR-RAD51

Adapted from: Konduri K*, Gallant JN*, Chae YK, Giles FJ, Gitlitz BJ, Gowen K, Ichihara E, Owonikoko TK, Peddareddigari V, Ramalingam SS, Reddy SK, Eaby-Sandy B, Vavala T, Whiteley A, Chen H, Yan Y, Sheehan JH, Meiler J, Morosini D, Ross JS, Stephens PJ, Miller VA, Ali SM, and Lovly CM. EGFR Fusions as Novel Therapeutic Targets in Lung Cancer. Cancer Discovery. 2016 Jun 2;6(6):601–11.
*co-first authors

Abstract

Here, we report that novel epidermal growth factor receptor (EGFR) gene fusions comprising the N-terminal of EGFR linked to various fusion partners, most commonly RAD51, are recurrent in lung cancer. We describe five patients with metastatic lung cancer whose tumors harbored EGFR fusions, four of whom were treated with EGFR tyrosine kinase inhibitors (TKIs) with documented anti-tumor responses. In vitro, EGFR-RAD51 fusions are oncogenic and can be therapeutically targeted with available EGFR-TKIs and therapeutic antibodies. These results support the dependence of EGFR-rearranged tumors on EGFR-mediated signaling and suggest several therapeutic strategies for patients whose tumors harbor this novel alteration.

Statement of Significance

We report for the first time the identification and therapeutic targeting of EGFR C-terminal fusions in patients with lung cancer and document responses to the EGFR inhibitor, erlotinib, in 4 patients whose tumors harbored EGFR fusions. Findings from these studies will be immediately translatable to the clinic as there are already several approved EGFR inhibitors.

Introduction

Oncogenic mutations in the epidermal growth factor receptor (EGFR) are found in a subset of patients with non-small cell lung cancer (NSCLC) and serve as important predictive biomarkers in this disease (1-3). These mutations, which most commonly occur as either small in frame deletions in exon 19 or point mutations in exon 21 (L858R) within the EGFR tyrosine kinase domain, confer constitutive activity and sensitivity to EGFR tyrosine kinase inhibitors (TKIs). Several large phase III clinical trials have shown that patients with EGFR-mutant lung cancer derive superior clinical responses when treated with EGFR-TKIs as compared with standard chemotherapy (4-6), and several EGFR inhibitors are already FDA-approved. These trials used PCR-based 'hotspot' testing, which typically interrogate for EGFR point mutations and small indels in exons 18-21. More recently, next-generation sequencing (NGS) of tumor samples has allowed for the identification of additional mechanisms whereby the EGF receptor may become aberrantly activated (7,8), further documenting the importance of EGFR signaling in the pathogenesis of lung cancer. Here, we report, for the first time in lung cancer, the presence of oncogenic EGFR fusions, most commonly EGFR-RAD51, which contain the entire EGFR tyrosine kinase domain fused to RAD51, a protein involved in DNA damage responses. We demonstrate that these fusions are oncogenic in pre-clinical studies and show that patients whose tumors harbor EGFR fusions derive significant clinical benefit from treatment with EGFR-TKI therapy.

Frequency of EGFR Alterations in Lung Cancer

To determine the frequency of EGFR fusions in lung cancer, we analyzed data from ~10,000 clinical cases **(Table S2.1)**. Fusion events, defined by a genomic breakpoint in EGFR exons 23 through intron 25, were detected in 5 patients, each of which is described below.

Case Reports

Patient 1, a 35-year-old female, was diagnosed with metastatic lung adenocarcinoma after presenting with generalized weakness and worsening vision. Imaging studies revealed widespread disease in the bone, liver, lymph nodes, adrenal glands, and hard palate **(Table 2.1)**. MRI showed innumerable metastases in the brain, dura, and left globe, resulting in retinal detachment. She was initially treated with radiotherapy to the brain and spine. Due to significant debility in the setting of tumor-induced disseminated intravascular coagulation (DIC), she was a poor candidate for cytotoxic chemotherapy. A lymph node biopsy was sent for genomic profiling using an extensively validated hybrid capture- based NGS diagnostic platform (FoundationOne) (9) and found to harbor a novel EGFR rearrangement at exon 25, resulting in the formation of an EGFR-RAD51 fusion gene **(Figures 2.1A–B and Table S2.2)**. The patient was treated with the EGFR-TKI,

Patient No.	Age	Gender	Ethnicity	Diagnosis	Smoking Status	Sites of disease	Prior Treatment	EGFR fusion	EGFR TKI	Best response	Duration of EGFR TKI therapy
1	35	Female	South Asian	Stage IV lung adenocarcinoma	Never	• Lung • Lymph nodes • Bone • Brain • Adrenal gland • Breast • Peritoneum • Eye	• RT to the whole brain and T8 vertebra	EGFR-RAD51	erlotinib	PR	8 months
2	21	Female	Caucasian	Stage IV lung adenocarcinoma	3 pack years	• Lung • Lymph nodes • Bone • Brain	• RT to the thoracic spine • Radiosurgery to brain mets • RT to the right hip	EGFR-RAD51	erlotinib	PR	5 months
3	43	Female	Caucasian	Stage IV lung adenocarcinoma	10 pack years, quit > 10 years ago	• Lung • Bone • Brain	• WBI • Carboplatin, pemetrexed, bevacizumab, • 6 cycles	EGFR-PURB	erlotinib	PR	20 months, on-going
4	38	Male	Caucasian	Stage IV lung adenocarcinoma	Former light smoker (3 pack years)	• Lung • Lymph nodes • Pleura • Bone	• Cisplatin/ Pemetrexed, • 6 cycles • Maintenance pemetrexed, • 11 cycles	EGFR-RAD51	erlotinib	PR	6 months, on-going
5	60	Female	Caucasian	Stage IV lung adenocarcinoma	Never	• Lung • Lymph nodes • Brain	• Craniotomy • RT to surgical bed • Carboplatin/ pemetrexed, • 4 cycles • Maintenance pemetrexed, • 6 cycles	EGFR-RAD51	N/A	N/A	N/A

Table 2.1: Clinical characteristics of patients with NSCLC harboring EGFR kinase fusions
Abbreviations: RT, radiation therapy; WBI, whole-brain irradiation; PR, partial response, N/A, not applicable; mets, metastases. This table corresponds to Table 1 from Konduri, Gallant et al., Cancer Discovery, 2016 (10).

erlotinib. Within two weeks of erlotinib initiation, DIC had resolved **(Figure S2.1A)** and the patient experienced clinical improvement with a noticeable decrease in supraclavicular lymphadenopathy and a hard palate metastatic lesion. After six months of treatment, the primary left lung mass and largest two liver lesions had decreased by 69% per RECIST (11) **(Figure 2.1C and Figure S2.1B)**, and the patient experienced an improvement in her functional status. She remained on erlotinib for 8 months, after which she experienced disease progression.

Patient 2, a 21-year-old female, was diagnosed with metastatic lung adenocarcinoma after presenting with right shoulder pain and unintentional weight loss. MRI revealed extensive metastatic disease in the spine and a right paraspinal mass extending into neuroforamina. Additional imaging studies showed metastatic disease in the brain, innumerable lung nodules, lymph nodes, and right acetabulum. Biopsy of an axillary lymph node showed metastatic adenocarcinoma consistent with lung primary. NGS testing revealed an EGFR-RAD51 fusion. The patient received palliative radiotherapy to the spine and brain metastases. Subsequently, the patient reported hemoptysis and dyspnea with exertion. Complete blood count showed a marked drop in platelet number and elevated lactate dehydrogenase, consistent with DIC. She was not a candidate for systemic chemotherapy. She was started on erlotinib approximately 6 weeks after initial presentation. Thrombocytopenia resolved within 10 days **(Figure S2.2A)**, and the patient experienced symptomatic improvement. CT scans obtained 3 months after the initiation of erlotinib showed a significant regression of bilateral miliary nodules as well as a 43% decrease in the index lesions of the left lower lobe (LLL), subcarinal lymph node, and right apical soft tissue mass compared to baseline **(Figure 2.1C and Figure S2.2B)**. The patient remained on erlotinib for 5 months with response, but she is no longer taking this medication due to nonmedical issues.

Patient 3, a 42-year-old female, was diagnosed with metastatic lung adenocarcinoma after presenting with right hip pain. Imaging studies revealed widespread disease including the primary left lower lobe (LLL) lesion, lytic lesions in the right pelvis and acetabulum, and brain metastases. Biopsy of a lung mass was positive for adenocarcinoma. She was initially treated with whole brain radiotherapy and platinum based chemotherapy with a partial response. While receiving chemotherapy, her tumor biopsy sample was sent for NGS testing and found to harbor an EGFR rearrangement at exon 25, resulting in the formation of a fusion gene between EGFR and PURB **(Table S2.2 and Figures S2.3A–B)**. At the time of disease progression on chemotherapy, the patient was treated with erlotinib, resulting in a 48% decrease in the LLL index lesion on-going for 20 months **(Figure 2.1C and Figure S2.3C)**.

Patient 4, a 38-year-old male, was diagnosed with metastatic lung adenocarcinoma after presenting with dyspnea and progressive weakness. Imaging studies showed metastatic disease to the lungs, lymph nodes, pleura, and bone. A pleural biopsy was performed, and NGS testing identified an EGFR-RAD51 fusion. He was initially treated with cisplatin/pemetrexed followed by maintenance pemetrexed. At the time of disease progression, the patient was started on erlotinib, with partial response after 2 cycles of therapy **(Figure 2.1C and Figure S2.4)**. The patient has now received erlotinib for 6 months with continued response.

Patient 5, a 60-year-old female, initially presented with headache, slurred speech, and left foot drag. MRI revealed three enhancing cerebral masses with midline shift. Further imaging studies showed a 4 cm mass in the lingula and lymphadenopathy. Biopsy of the lung mass revealed adenocarcinoma. The patient underwent resection of a right frontal tumor followed by radiotherapy. She was treated with four cycles of carboplatin/pemetrexed with partial response followed by pemetrexed maintenance therapy. During this treatment, NGS testing was completed and revealed an EGFR-RAD51 fusion. The patient continues to receive benefit from pemetrexed therapy; she has not yet been treated with an EGFR-TKI.

EGFR-RAD51 is Oncogenic

We stably expressed EGFR variants in Ba/F3 cells and detected expression of EGFR-RAD51 at the expected molecular weight as compared to EGFR wild-type (WT) and the known oncogenic EGFR- L858R mutation **(Figure 2.2A)**. We observed that, analogous to EGFR-L858R, EGFR-RAD51 was able to activate downstream signaling through the MAPK and PI3K/AKT pathways. EGFR-RAD51 was also able to sustain IL-3-independent proliferation of Ba/F3 cells, an activity phenotype associated with the transforming function of other oncogenic tyrosine kinases **(Figure 2.2B)** (12). In parallel, we expressed the same EGFR variants in NR6 cells (which lack endogenous EGFR (13)) **(Figure S2.5A–B)**. EGFR-RAD51 significantly increased colony formation of NR6 cells in soft agar—a hallmark of tumor cells—as compared to control cells and those expressing EGFR-WT **(Figure S2.5C–D)**.

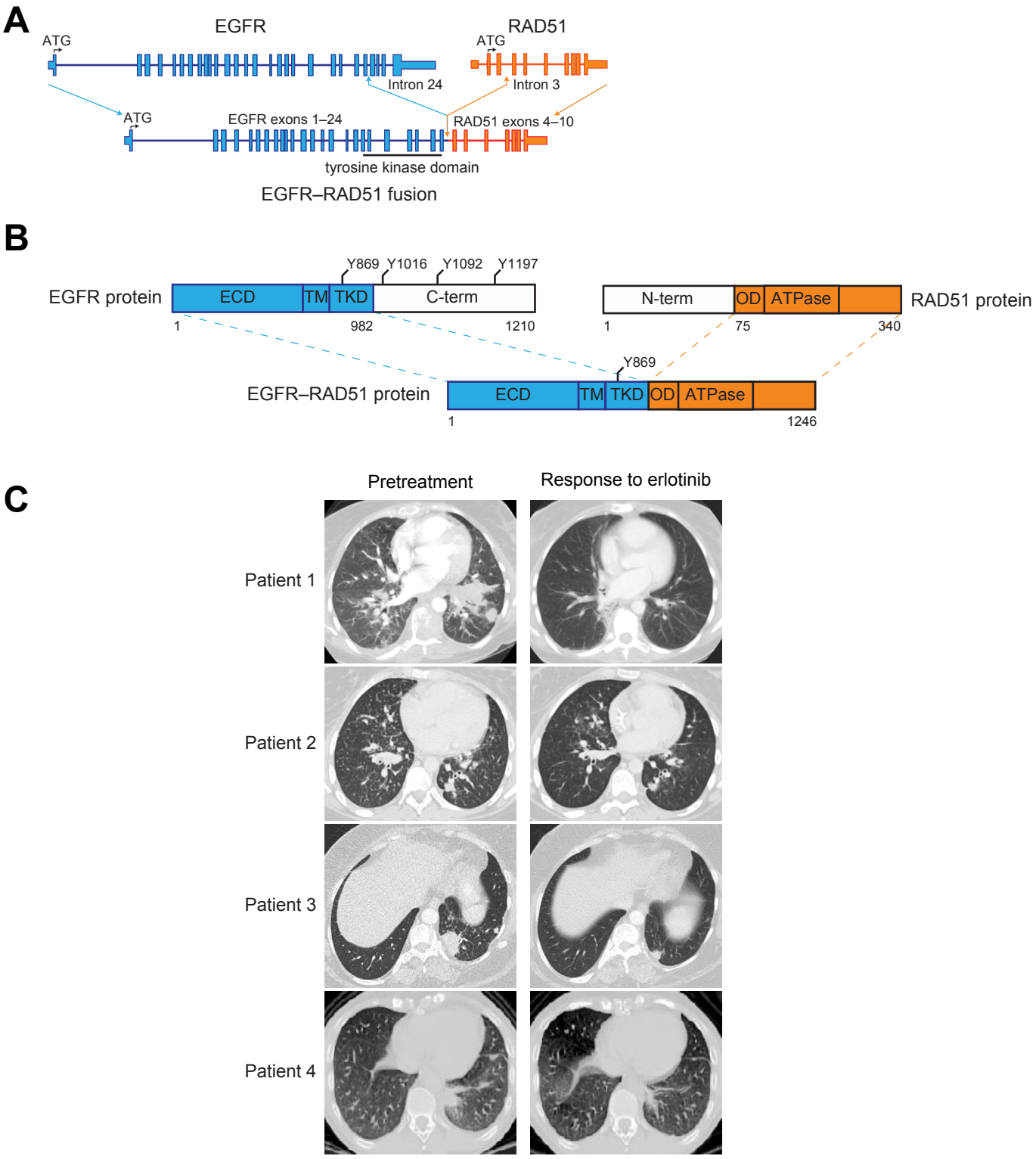


Figure 2.1: EGFR fusions are clinically actionable.
A, scaled representation of EGFR-RAD51 depicting the genomic structure of the fusion. ATG, translational start site; blue, EGFR; orange, RAD51. **B,** schematic representation of the EGFR-RAD51 fusion protein domain structure. Numbers correspond to amino acid residues. Y, tyrosine residue; ECD, extracellular domain; TM, transmembrane helix; TKD, tyrosine kinase domain; C-term, carboxyl terminus region; N-term, amino terminus; OD, oligomerization domain/section; ATPase, adenylpyrophosphatase; blue, EGFR; orange, RAD51. **C,** serial CT scans from index patients with lung adenocarcinoma harboring EGFR fusions, documenting response to the EGFR-TKI erlotinib. Left images, scans obtained prior to initiation of erlotinib. Right images, scans obtained during erlotinib therapy. This figure corresponds to Figure 1 from Konduri, Gallant et al., Cancer Discovery, 2016 (10).

EGFR contains several autophosphorylation sites in the C-terminal tail of the receptor (including tyrosines 1016, 1092, and 1197) that positively regulate the transforming activity of EGFR by mediating downstream proliferative signaling (14). These autophosphorylation sites, which serve as docking sites for signaling adaptor proteins, are lacking in the EGFR-RAD51 fusion **(Figure 2.1B and Figure 2.2C)**. Notably, however, EGFR-RAD51 contains tyrosine 869, a phosphorylation site within the kinase domain which is critical for complete EGFR function and transformation in NSCLC (15). The presence of tyrosine 869 may explain why EGFR-RAD51 is still able to activate downstream oncogenic signaling via the PI3K/AKT and MAPK pathways **(Figure 2.2A and C)**. Notably, these EGFR fusions also lack tyrosine 1069, the CBL binding site which targets EGFR for degradation. Indeed, EGFR-RAD51 is more stable (has a slower turnover rate) compared with the WT EGF receptor **(Figure S2.6A–B)**, suggesting that EGFR-RAD51 receptor stability might also play a role in its oncogenic properties. Taken together, these data support that EGFR-RAD51 is able to activate tumorigenic signaling and confer an oncogenic phenotype.

Computational Modeling of EGFR-RAD51

The EGFR tyrosine kinase is activated through ligand-mediated formation of asymmetric (N-lobe to C-lobe) dimers (16). Kinase fusions, on the other hand, commonly share a mechanism of activation whereby the fusion partner drives dimerization of the kinase and leads to ligand-independent activation (17). Given the presence of RAD51, a self-assembling filamentous protein (18), we hypothesized that the EGFR-RAD51 fusion protein can form such partner-driven dimers. To validate this hypothesis, we modeled EGFR-RAD51 based on available experimental structures of RAD51 and the active asymmetric EGFR dimer. Conformational loop sampling with Rosetta demonstrates that it is geometrically feasible for EGFR kinase subunits to adopt the asymmetric (active) dimeric conformation when fused to RAD51 **(Figure 2.2D)**. Further, the concatenation of RAD51 subunits could bring tethered EGFR kinase domains close together, increasing their local concentration, and leading to further EGFR activation **(Figure S2.7)**. Although this structural modeling demonstrates that the EGFR-RAD51 is geometrically capable of forming active dimers, further experimental data are needed to confirm this mechanism.

EGFR-RAD51 can be Therapeutically Targeted with Existing EGFR Inhibitors

The finding of recurrent EGFR fusions in lung cancer is of particular interest because EGFR-TKIs have proven an effective therapeutic strategy for tumors harboring certain EGFR mutations. Therefore, we sought to determine the effectiveness of EGFR-TKIs against EGFR-RAD51. We treated Ba/F3 cells expressing EGFR-RAD51 with erlotinib (1st-generation reversible EGFR-TKI), afatinib (2nd- generation irreversible EGFR/HER2-TKI), and osimertinib (3rd-generation irreversible EGFR-TKI) to assess the effects of these inhibitors on cellular proliferation. EGFR-L858R served as a positive control in this experiment. Each TKI effectively inhibited the growth of Ba/F3 cells expressing EGFR-RAD51 to varying degrees **(Figure 2.3A and Table S2.3)**. Downstream MAPK and PI3K/AKT signaling was also inhibited with TKI treatment **(Figure 2.3B)**. The on-target effect of EGFR-TKIs could be observed when blotting for phospho-tyrosine from immunoprecipitated EGFR-RAD51 protein **(Figure 2.3C)** and when observing the phosphorylation status of tyrosine 869, which is included in the fusion protein **(Figure S2.8)**. Finally, we tested the effects of the FDA-approved EGFR antibody, cetuximab, in our cell culture models. Cetuximab binds to the EGFR extracellular domain and blocks the binding of growth factors, such as EGF (19). In contrast to EGFR-L858R, proliferation of Ba/F3 cells expressing EGFR-RAD51 was potently inhibited by cetuximab **(Figure 2.3D and Figure S2.9)**. Together, these results show that the EGFR-RAD51 can be potently inhibited by a variety of EGFR-targeted agents, suggesting several intriguing clinical avenues.

Discussion

Just over ten years ago, ‘canonical’ EGFR point mutations and short indels in the kinase domain were detected retrospectively by PCR-based ‘hotspot’ testing in patients with lung cancer who responded to EGFR-TKI therapy (1-3). Assessing for these ‘canonical’ EGFR mutations is now the accepted standard of care worldwide for patients with lung cancer. Here, by utilizing a comprehensive NGS assay that interrogates the entire coding region of EGFR (as well as introns 7, 15, 24, 25, and 26), we identified novel EGFR fusions in lung cancer patients—EGFR-RAD51 and EGFR-PURB—that would otherwise have gone undetected by the standard of care.

While distinct EGFR fusions have previously been observed in glioma (20), this is the first documentation of patients with EGFR fusion positive tumors that derived significant and sustained anti-tumor responses from treatment with the EGFR-TKI, erlotinib. Our in vitro work also hints at afatinib being potent

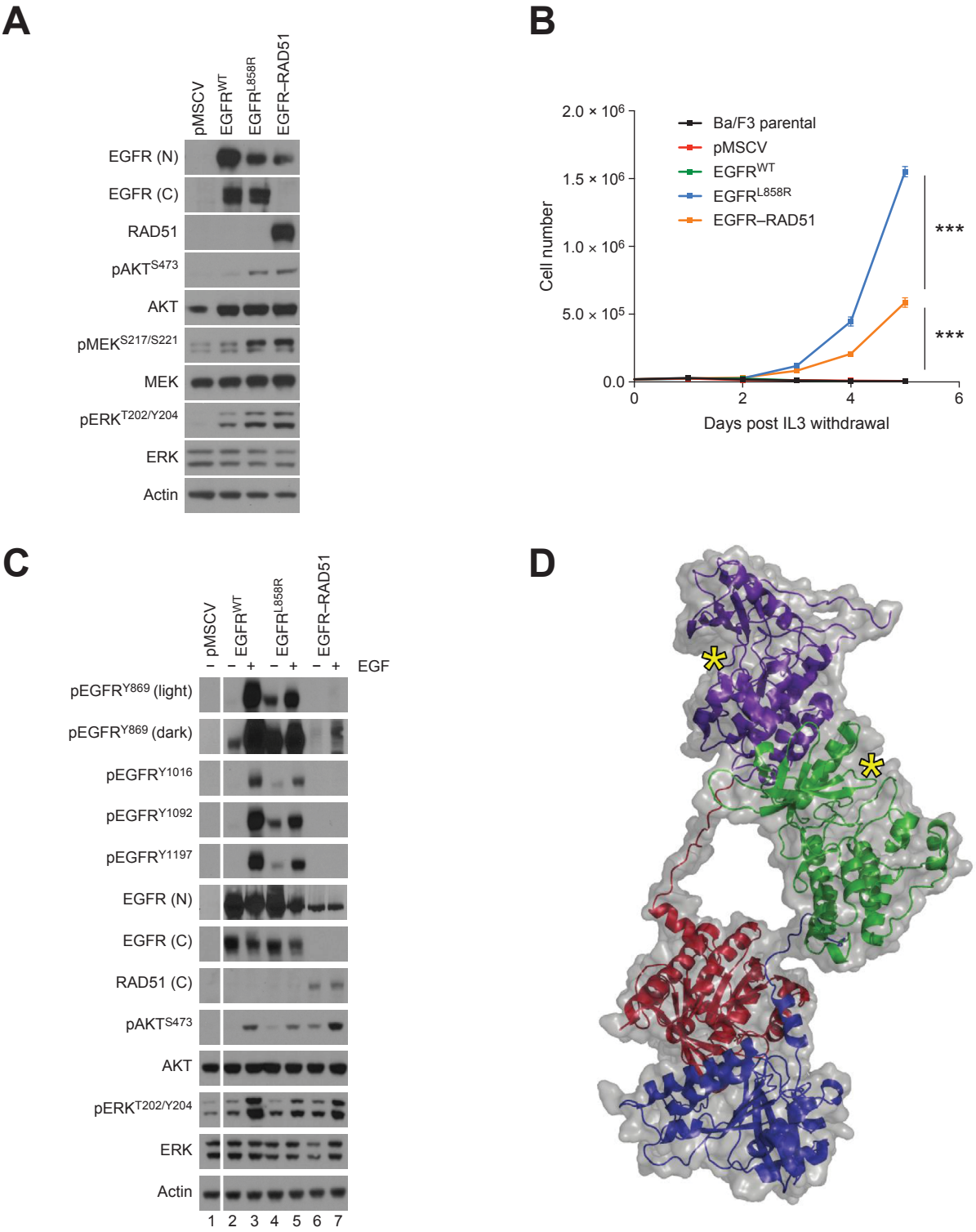


Figure 2.2: EGFR-RAD51 is an oncogenic EGFR alteration.
A, Ba/F3 lines stably expressing pMSCV (vector only), EGFR-WT, EGFR-L858R or EGFR-RAD51 were subjected to Western blot analysis with indicated antibodies. The three distinct EGFR variants were detected at the anticipated molecular weight (MW) of ~150 kD. EGFR-RAD51 fusion is detected with both the N-terminal EGFR antibody [EGFR(N)] and with the RAD51 antibody. **B,** Ba/F3 cells transduced with indicated constructs (pMSCV, vector only) were grown in the absence of IL3 and counted every 24 hours. ***, P < 0.0001. **C,** Ba/F3-expressing EGFR variants were serum starved for 16 hours, treated with 50 ng/mL EGF for 5 minutes, and subjected to Western blot analysis with indicated antibodies. **D,** ribbon diagram and space-filling model of the EGFR-RAD51 kinase domains illustrating the proposed mechanism of activation. Purple, first EGFR kinase domain; green, second EGFR kinase domain; red, first RAD51 partner; blue, second RAD51 partner; yellow asterisks, active sites. This figure corresponds to Figure 2 from Konduri, Gallant et al., Cancer Discovery, 2016 (10).

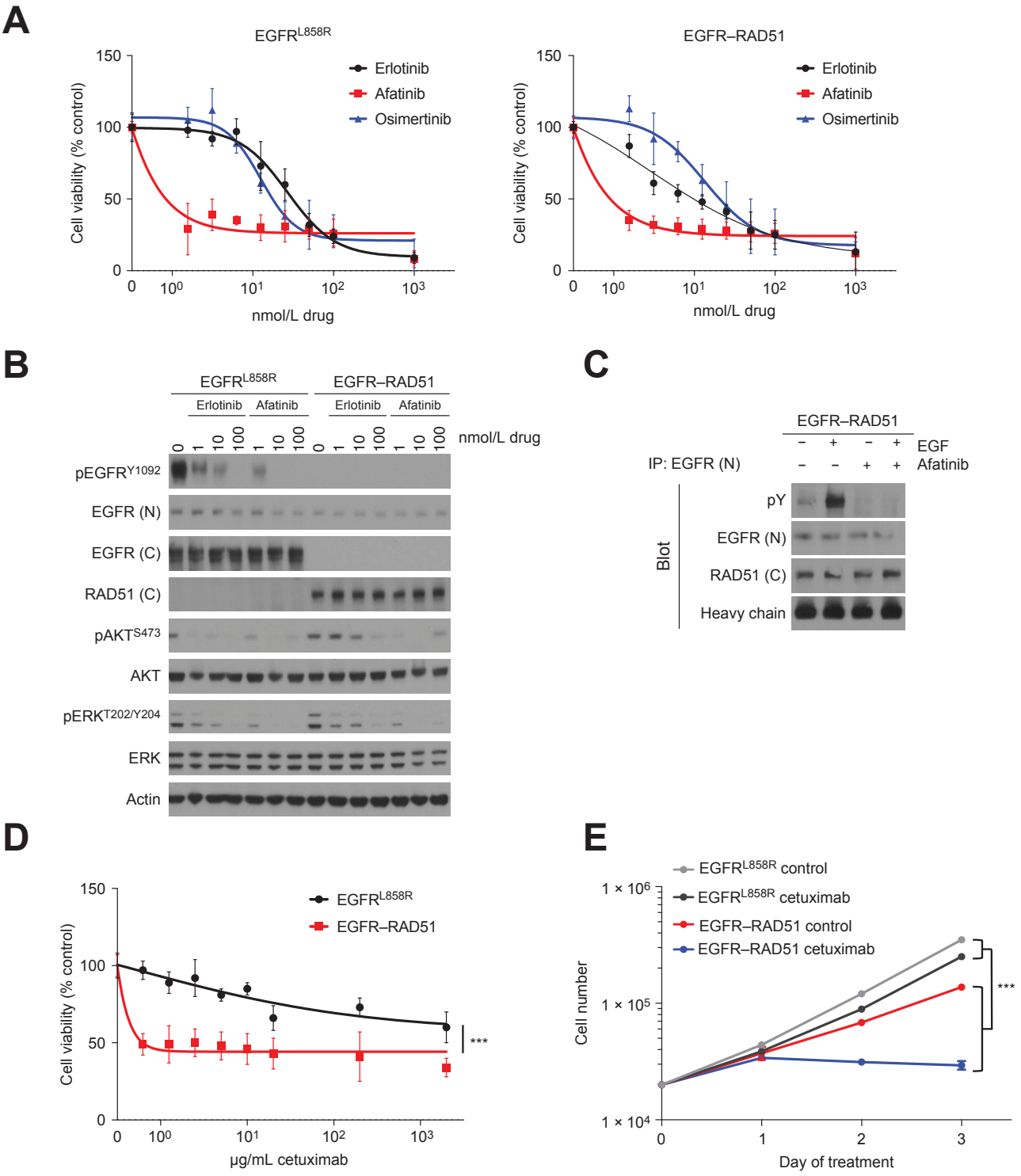


Figure 2.3: EGFR-RAD51 is therapeutically targetable with EGFR inhibitors.
A, Ba/F3 lines stably expressing EGFR-L858R or EGFR-RAD51 were treated with increasing doses of erlotinib, afatinib, or osimertinib for 72 hours. Each point represents six replicates. Data are presented as the mean percentage of viable cells compared with vehicle control ± SD. **B,** Ba/F3 lines stably expressing EGFR-L858R or EGFR-RAD51 were treated with increasing doses of erlotinib or afatinib for 2 hours and subjected to Western blot analysis with indicated antibodies. **C,** Ba/F3 cells stably expressing EGFR-RAD51 were serum starved for 16 hours and then treated with 100 nmol/L afatinib for 1 hour followed by 50 ng/mL EGF for 5 minutes. EGFR was immunoprecipitated from cellular lysates with an antibody targeting the EGFR N-terminus and then subjected to Western blot analysis with indicated antibodies. **D,** Ba/F3 cells stably expressing EGFR-L858R or EGFR-RAD51 were treated with increasing doses of cetuximab for 72 hours. Each point represents six replicates. Data are presented as the mean percentage of viable cells compared with vehicle control ± SD. ***, P < 0.0001. **E,** Ba/F3 cells stably expressing EGFR-L858R or EGFR-RAD51 were treated with 5 μg/mL cetuximab and counted every 24 hours. Each point represents the average of three replicates ± SD. ***, P < 0.0001. This figure corresponds to Figure 3 from Konduri, Gallant et al., Cancer Discovery, 2016 (10).

against EGFR-RAD51—highlighting that the structural effects of the EGFR mutation, the resultant conformation of the EGFR kinase domain, and the type of EGFR inhibitor are all important factors in determining the efficacy of a specific EGFR-TKI against a particular EGFR mutation (21).

Interestingly, EGFR-RAD51 fusions were markedly sensitive to both EGFR-TKIs and the EGFR antibody, cetuximab. While some unselected NSCLC patients respond to cetuximab, the monoclonal anti-EGFR antibody adds minimal benefit for most patients—even when combined with chemotherapy (22,23). Previous work has shown that cetuximab sensitivity is correlated with asymmetric dimerization (24). As our modeling suggests that the EGFR-RAD51 fusion is activated by virtue of constitutive dimerization, these findings suggest a unique molecular cohort that may benefit from cetuximab. Further, these findings provide a rationale for therapeutically targeting this subset of lung cancers with a wide array of anti-EGFR therapies, many of which are already FDA-approved. On-going work will elucidate whether the deregulated RAD51 component of the EGFR-RAD51 fusion protein may also be a therapeutic vulnerability for treatment with platinum-based and PARP-inhibitor therapies.

Our experimental work also demonstrates that EGFR-RAD51 fusions are oncogenic and able to mediate downstream signaling through the MAPK and PI3K/AKT pathways. While this may seem surprising, given that EGFR-RAD51 fusions lack the C-terminal tail known to be important for EGFR signal transduction, analyses of several cancer types have identified EGFR C-terminal deletions as recurrent and transforming events (25-27). The exact mechanism whereby these EGFR C-terminal deletions activate the EGFR kinase domain and confer transforming properties remains unclear. In the case of EGFR-RAD51 fusions, we believe there is a unique role for RAD51 given the specific fusion event's recurrence. A common characteristic of tyrosine kinase fusions is that the fusion partner (here, RAD51) contributes an oligomerization domain, which promotes activation of the kinase (28). As demonstrated by our structural modeling, EGFR-RAD51 fusion proteins could be activated by virtue of RAD51 oligomerization. Similarly, structural modeling has shown that PURB proteins can self-associate in the absence of nucleic acids . Additional experimental work will be required to determine whether activation of EGFR fusion proteins is driven by asymmetric dimerization of the EGFR tyrosine kinase domain, dimerization through the known partner oligomerization interface, or both.

The cases presented here highlight that adjusting our strategy and using newly available tools, such as comprehensive NGS tests, could prove useful in detecting alternative ways in which the EGFR pathway is altered (and can be targeted) in tumors. Based on the observation that the EGFR-RAD51 fusions detected in the first two patients occurred with breakpoints in EGFR intron 24, we defined further fusion events by the presence of a genomic breakpoint in EGFR exons 23 through intron 25. While we limited our investigation to these parameters, we cannot exclude that other EGFR rearrangements may exist in lung cancer outside of these parameters. Refinements in the assay may help discover more clinically relevant EGFR fusions or alterations in the future.

Methods

EGFR plasmid construction

A cDNA encoding the EGFR-RAD51 fusion was synthesized by Life Technologies based on the combined Refseq sequences of EGFR (NM_ 005228) and RAD51 (NM_002875) **(Figure S2.10)**. The pMSCV-puro vector backbone (Clontech) was used to construct all retroviruses. Assembly of pMSCV-puro-EGFR-WT and pMSCV-puro-EGFR-L858R was previously described (29). EGFR-RAD51 was subcloned from the pMA synthesis vector (Life Technologies) into the HpaI site of pMSCV-puro using blunt end ligation. All plasmids were sequence verified in the forward and reverse directions (10).

Cell Culture

Ba/F3 cells were purchased from DSMZ. Plat-GP cells were purchased from CellBioLabs. NR6 cells have been previously described (13). Ba/F3 cells were maintained in RPMI 1640 medium (Mediatech, Inc.). NR6 and Plat-GP cells were maintained in DMEM (Gibco). Media was supplemented with 10% heat inactivated fetal bovine serum (Atlanta Biologicals) and penicillin-streptomycin (Mediatech, Inc.) to final concentrations of 100 U/ml and 100 µg/ml, respectively. The Ba/F3 cell line was supplemented with 1 ng/mL murine IL-3 (Gibco). The Plat-GP cell line was cultured in the presence of 10 µg/mL blasticidin (Gibco). All cell lines were maintained in a humidified incubator with 5% CO2 at 37°C and routinely evaluated for mycoplasma contamination.

Ba/F3 and NR6 Cell Line Generation

The empty pMSCV-puro retroviral vector or pMSCV-puro vectors encoding EGFR (either EGFR-WT, EGFR-L858R, or EGFR-RAD51) were transfected, along with the envelope plasmid pCMV-VSV-G (Addgene), into cells Plat-GP packaging cells (CellBioLabs). Viral media was harvested 48 hours after transfection, spun down to remove debris, and supplemented with 2 μg/mL polybrene (Santa Cruz). 2.5×10^6 Ba/F3 cells (or 1×10^6 NR6 cells) were re-suspended in 10 mL viral media. Transduced cells were selected for 1 week in 2 μg/mL puromycin (Invivogen), and Ba/F3 cells were selected for an additional week in the absence of IL-3. Stable polyclonal populations were used for experiments and routinely tested for expression of EGFR constructs.

Compounds

Erlotinib, afatinib, and osimertinib (AZD9291) were purchased from Selleck Chemicals. Cetuximab was purchased from ImClone.

Antibodies, Immunoprecipitation, and Immunoblotting

The following antibodies were obtained from Cell Signaling Technology: phospho-EGFR tyrosine 845 (#2231, 1:1000 dilution), phospho-EGFR tyrosine 992 (#2235, 1:500 dilution), phospho-EGFR tyrosine 1068 (#2234, 1:1000 dilution), phospho-EGFR tyrosine 1173 (#4407, 1:1000 dilution), C-term EGFR clone D38 (#4267, 1:2000 dilution), phospho-AKT serine 473 (#9271, 1:500 dilution), AKT (#9272, 1:1000 dilution), phospho-ERK threonine 202/tyrosine 204 (#9101, 1:2000 dilution), ERK (#9102, 1:2000 dilution), HRP-conjugated anti-mouse (#7076, 1:5000 dilution), and HRP-conjugated anti-rabbit (#7074, 1:5000 dilution). β-actin antibody (#A2066, 1:5000 dilution) was purchased from Sigma-Aldrich. N-term EGFR clone H11 (MA-13070, 1:500 dilution) and HRP-conjugated anti-goat (PA-129617, 1:5000 dilution) were purchased from Thermo Fisher. N-term EGFR clone 528 (#120, 1:500 dilution) and C-term RAD51 clone C20 (#6862, 1:1000 dilution) were purchased from Santa-Cruz Biotechnologies. Anti-phosphotyrosine clone 4G10 (#05-321; 1:2000 dilution) was purchased from Millipore.

For immunoprecipitation, cells were harvested, washed in PBS, and lysed in NDLB buffer (1% Triton-X-100, 137 mM NaCl, 10% Glycerol, 20 mM Tris·HCl, pH 8.0) with freshly added 40 mM NaF, 1 mM Na-orthovanadate, and protease inhibitor mini tablets (Thermo Scientific). Protein was quantified using protein assay reagent and a SmartSpec plus spectrophotometer (Bio-Rad) per the manufacturer's protocol. 300–500 μg of lysates were subjected to overnight immunoprecipitation with 2 μg N-term EGFR clone 528 (Santa Cruz#120; 10 μL). Antibody was precipitated with Protein A Dynabeads (Invitrogen). Immunoblotting was then performed as described below.

For immunoblotting, cells were harvested, washed in PBS, and lysed in RIPA buffer (150 mM NaCl, 1% Triton-X-100, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris·HCl, pH 8.0) with freshly added 40 mM NaF, 1 mM Na-orthovanadate, and protease inhibitor mini tablets (Thermo Scientific). Protein was quantified (as detailed above) and 20μg of lysates were subjected to SDS-PAGE. Protein was transferred to PVDF membranes (Millipore) at 1000 mA·hr, blocked in 5% BSA, and incubated with antibodies as detailed above. Detection was performed using Western Lightning ECL reagent (Perkin Elmer) and autoradiography film paper (Denville). Samples analyzed with N-term EGFR clone 528 were prepared under non-reducing and non-boiled conditions.

Cell Viability, Counting, and Clonogenic Assays

For viability experiments, cells were seeded at 5,000 cells/well in 96-well plates and exposed to treatment the following day. At 72 hours post drug addition, Cell Titer Blue reagent (Promega) was added, and fluorescence at 570 nm was measured on a Synergy MX microplate reader (Biotek) according to the manufacturer's instructions.

For cell counting experiments, cells were seeded at 10,000 cells/well in 24-well plates in the absence of 1 ng/mL IL-3. Every 24 hours, cells were diluted 20 fold and counted using a Z1 Coulter Counter (Danaher).

For cell counting experiments with drug, cells were seeded at 20,000 cells/well in 12-well plates in the absence of 1 ng/mL IL-3. Drugs were added at the following concentrations: erlotinib 500 nM, afatinib 50 nM, osimertinib 500 nM, cetuximab 5 μg/mL. Every 24 hours, cells were diluted 20 fold and counted using a Z1 Coulter Counter (Danaher).

Soft Agar Assays

1.5 mL of 0.5% agar/DMEM was layered in each well of a 6-well dish. A total of 5,000 NR6 cells in 1.5 mL of 0.33% agar/DMEM were seeded on top of the initial agar and allowed to grow for three weeks. Each cell

line was plated in triplicate. Colonies were counted using GelCount (Oxford Optronix) with identical acquisition and analysis settings.

Structural Modeling of the EGFR-RAD51 Fusion

The sequence of the EGFR-RAD51 fusion protein was used to generate a structural model based on the crystal structure templates 1SZP.PDB (S. cerevisiae Rad51) (18) and 2GS6.PDB (human EGFR) (16). PyMOL version 1.5.0.3 (Schrödinger) was used to combine two monomers of yeast RAD51 and two kinase domains of EGFR into a single template structure for input to modeling. The N-termini of the RAD51 domains were positioned close to the C-termini of the EGFR domains to represent the fusion result. Modeller version 9.14 (30) was then used to generate the dimeric model of the fusion protein structure. The conformational space for the dimer was then sampled using Rosetta version 2015.05 (31). A total of 20,000 independent modeling runs were performed using kinematic loop closure. To illustrate the arrangement of the filament, the crystallographic symmetry records from 1SZP. PDB were then used to construct eight additional copies of the complex in PyMOL.

Clinical Data and Tumor Genotyping

Protected health information was reviewed according to the Health Insurance Portability and Accountability Act (HIPAA) guidelines. Genomic profiling of tumor samples was performed using a hybrid capture-based NGS diagnostic platform (FoundationOne) (9). The database of >56,000 Foundation Medicine clinical cases (Primary_150929_114735, November 2nd, 2015) was interrogated for rearrangement class events to identify those cases likely to harbor a EGFR fusion event using the FoundationOne Molecular Information Browser v0.8 (9). Cases involving an event with a genomic breakpoint in EGFR exons 23 through intron 25 were manually investigated to evaluate the potential of each individual rearrangement. Exon boundaries were chosen on the basis of the index cases (EGFR-RAD51 harboring a genomic breakpoint in EGFR intron 24).

Statistics and Data Presentation

All experiments were performed using at least three technical replicates and at least two independent times (biological replicates). For statistical analyses, all biological and technical replicates were pooled to perform an integrated assessment on the differences among groups. To determine the IC_{50}s, a logistic regression was applied after data was transformed using log transformation. To determine the differences in cell counts, IC_{50}s, time, and dose trends, and in order to account for the dependence of technical replications and repeated measurements, the linear mixed model was used to perform the analysis. The assumption of normality for mixed model was investigated. If necessary, data was transformed using log transformation for the linear mixed model. R3.2.2 (www.R-project.org) was used to perform all statistical analyses.

Each figure or panel shows a single representative experiment with the statistical significance derived from integrated experimental analyses—as described above. Unless indicated otherwise, data is presented as mean ± standard deviation. Western blot autoradiography films were scanned in full color at 600 dpi, desaturated in Adobe Photoshop CC, and cropped in Powerpoint. Genomic and proteomic diagrams were created in Adobe Illustrator CC. Patient images were cropped to highlight the region of interest. No other image alterations were made.

References

1. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304,** 1497–1500 (2004).

2. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350,** 2129–2139 (2004).

3. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 13306–13311 (2004).

4. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361,** 947–957 (2009).

5. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13,** 239–246 (2012).

6. Sequist, L. V. *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *Journal of Clinical Oncology* **31,** 3327–3334 (2013).

7. Gallant, J. N. *et al.* EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib. *Cancer Discovery* **5,** 1155–1163 (2015).

8. Baik, C. S., Wu, D., Smith, C., Martins, R. G. & Pritchard, C. C. Durable Response to Tyrosine Kinase Inhibitor Therapy in a Lung Cancer Patient Harboring Epidermal Growth Factor Receptor Tandem Kinase Domain Duplication. *Journal of Thoracic Oncology* **10,** e97–e99 (2015).

9. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology* **31,** 1023–1031 (2013).

10. Konduri, K. *et al.* EGFR Fusions as Novel Therapeutic Targets in Lung Cancer. *Cancer Discovery* **6,** 601–611 (2016).

11. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45,** 228–247 (2009).

12. Daley, G. Q. & Baltimore, D. Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 9312–9316 (1988).

13. Pruss, R. M. & Herschman, H. R. Variants of 3T3 cells lacking mitogenic response to epidermal growth factor. *Proc. Natl. Acad. Sci. U.S.A.* **74,** 3918–3921 (1977).

14. Jorissen, R. N. *et al.* Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp. Cell Res.* **284,** 31–53 (2003).

15. Chung, B. M. *et al.* The role of cooperativity with Src in oncogenic transformation mediated by non-small cell lung cancer-associated EGF receptor mutants. *Oncogene* **28,** 1821–1832 (2009).

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

17. Shaw, A. T., Hsu, P. P., Awad, M. M. & Engelman, J. A. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat. Rev. Cancer* **13,** 772–787 (2013).

18. Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nat. Struct. Mol. Biol.* **11,** 791–796 (2004).

19. Li, S. *et al.* Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7,** 301–311 (2005).

20. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nature Communications* **5,** 1–10 (2014).

21. Vivanco, I. *et al.* Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discovery* **2,** 458–471 (2012).

22. Pirker, R. *et al.* Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. *Lancet* **373,** 1525–1531 (2009).

23. Lynch, T. J. *et al.* Cetuximab and first-line taxane/carboplatin chemotherapy in advanced non-small-cell lung cancer: results of the randomized multicenter phase III trial BMS099. *Journal of Clinical Oncology* **28,** 911–917 (2010).

24. Cho, J. *et al.* Cetuximab response of lung cancer-derived EGF receptor mutants is associated with asymmetric dimerization. *Cancer Research* **73,** 6770–6779 (2013).

25. Cho, J. *et al.* Glioblastoma-Derived Epidermal Growth Factor Receptor Carboxyl-Terminal Deletion Mutants Are Transforming and Are Sensitive to EGFR-Directed Therapies. *Cancer Research* **71,** 7587–7596 (2011).

26. Imielinski, M. *et al.* Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell* **150,** 1107–1120 (2012).

27. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155,** 462–477 (2013).

28. Soda, M. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448,** 561–566 (2007).

29. Red Brewer, M. *et al.* Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proceedings of the National Academy of Sciences* **110,** E3595–604 (2013).

30. Eswar, N., Eramian, D., Webb, B., Shen, M.-Y. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426,** 145–159 (2008).

31. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology* **268,** 209–225 (1997).

CHAPTER III: EGFR-KDD

Adapted from: Gallant JN, Sheehan JH, Shaver TM, Bailey M, Lipson D, Chandramohan R, Brewer MR, York SJ, Kris MG, Pietenpol JA, Ladanyi M, Miller VA, Ali SM, Meiler J, and Lovly CM. EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib. Cancer Discovery. 2015 Nov 1;5(11):1155–63.

Abstract

Oncogenic EGFR mutations are found in 10-35% of lung adenocarcinomas. Such mutations, which present most commonly as small in-frame deletions in exon 19 or point mutations in exon 21 (L858R), confer sensitivity to EGFR tyrosine kinase inhibitors (TKIs). In analyzing the tumor from a 33-year-old male never smoker, we identified a novel EGFR alteration in lung cancer: EGFR exon 18-25 kinase domain duplication (EGFR-KDD). Through analysis of a larger cohort of tumor samples, we detected additional cases of EGFR-KDD in lung, brain, and other cancers. In vitro, EGFR-KDD is constitutively active, and computational modeling provides potential mechanistic support for its auto-activation. EGFR-KDD-transformed cells are sensitive to EGFR-TKIs and, consistent with these in vitro findings, the index patient had a partial response to the EGFR-TKI, afatinib. The patient eventually progressed, at which time, re-sequencing revealed an EGFR-dependent mechanism of acquired resistance to afatinib, thereby validating EGFR-KDD as a driver alteration and therapeutic target.

Statement of Significance

We identified oncogenic and drug sensitive EGFR exon 18-25 kinase domain duplications (EGFR-KDD) that are recurrent in lung, brain, and soft tissue cancers and documented that a patient with metastatic lung adenocarcinoma harboring the EGFR-KDD derived significant anti-tumor response from treatment with the EGFR inhibitor, afatinib. Findings from these studies will be immediately translatable as there are already several approved EGFR inhibitors in clinical use.

Introduction

The prospective identification and rational therapeutic targeting of tumor genomic alterations have revolutionized the care of patients with lung cancer and other malignancies. Oncogenic mutations in the epidermal growth factor receptor (EGFR) tyrosine kinase domain are found in an important subset of non-small cell lung cancer (NSCLC), and several large phase III clinical trials have shown that patients with EGFR-mutant lung cancer derive superior clinical responses when treated with EGFR tyrosine kinase inhibitors (TKIs) as compared with standard chemotherapy (1-3). Such mutations, which most commonly occur as either small in-frame deletions in exon 19 or point mutations in exon 21 (L858R), confer constitutive activity to the EGFR tyrosine kinase and sensitivity to EGFR-TKIs (4). Other oncogenic alterations, including ALK and ROS1 gene rearrangements, have similarly allowed for the rational treatment of molecular cohorts of NSCLC (5,6). Unfortunately, despite these significant advances in defining clinically relevant molecular cohorts of lung cancer, the currently identified genomic alterations account for only 50-60% of all tumors (7). Additional analyses are necessary to identify therapeutically actionable molecular alterations in these tumors.

Here, we describe the case of a 33-year-old male never smoker with metastatic lung adenocarcinoma whose tumor lacked all previously described actionable genomic alterations in this disease. Targeted next generation sequencing (NGS) based genomic profiling identified a novel in-frame tandem duplication of EGFR exons 18-25, the exons that encode the EGFR tyrosine kinase domain. This EGFR kinase domain duplication (EGFR-KDD) had not previously been reported in lung cancer, and there were no pre-clinical data or clinical evidence to support the use of EGFR inhibitors in patients whose tumors harbor the EGFR-KDD. However, the index patient was treated with the EGFR-TKI, afatinib, with rapid symptomatic improvement and significant decrease in tumor burden. Notably, upon disease progression, the patient's tumor harbored an increase in the copy number of the EGFR-KDD, solidifying the role of this EGFR alteration as a novel driver in this disease. Through analysis of a large set of annotated tumors, we demonstrate that the EGFR-KDD is recurrent in lung, brain, and soft tissue tumors. Overall, our data show, for the first time, that EGFR-KDD is an oncogenic and therapeutically actionable alteration.

Case Report

A 33-year-old male never smoker was diagnosed with stage IV lung adenocarcinoma after presenting

with cough and fatigue. His tumor biopsy was sent for genomic profiling using an extensively validated hybrid-capture-based NGS diagnostic assay (FoundationOne) (8). The patient's tumor was found to be negative for any previously described actionable alterations in lung cancer, including negative for the presence of previously described EGFR alterations such as L858R, G719A/C/S, and L861Q point mutations, exon 19 deletion/insertion, and exon 20 insertion (9). Interestingly, however, the patient's tumor was found to harbor an intragenic alteration in EGFR resulting in the tandem duplication of exons 18-25 **(Figure S3.1A)**. The presence of this alteration was confirmed by direct sequencing **(data not shown)** and by an independent clinical NGS assay (MSK-IMPACT™)(10)**(Figure S3.1B)**. Since exons 18-25 of EGFR encode the entire tyrosine kinase domain, this alteration results in an EGFR protein that has an in-frame kinase domain duplication **(Figure 3.1A and Figure S3.2)**. Notably, this EGFR alteration had not previously been reported in lung cancer; the EGFR-KDD (as duplication of exons 18-25 or 18-26) had only been reported in isolated cases of glioma to date (11-15). There were no data regarding the frequency of this alteration in tumor samples, nor were there data regarding the efficacy of EGFR-targeted agents against the EGFR-KDD.

Frequency of EGFR-KDD in Lung and Other Cancers

To determine the frequency of the EGFR-KDD in lung cancer and other tumors, we analyzed data from >38,000 clinical cases, each of which had results from FoundationOne targeted sequencing, analogous to the index patient. The EGFR-KDD was detected in 5 tumors from ~7,200 total lung cancers tested. In addition, EGFR-KDD was identified in 3 gliomas, 1 sarcoma, 1 peritoneal carcinoma, and 1 Wilms' Tumor **(Table 3.1)**. Among samples in The Cancer Genome Atlas (TCGA), we found previously unreported cases of the EGFR-KDD in lung adenocarcinoma (16) and glioblastoma multiforme (17,18) **(Table 3.1 and Figures S3.3A–D)**. Two additional cases were found by MSK-IMPACT™ sequencing **(Table 3.1)**. Together, these data show that the EGFR-KDD is a recurrent mutation in lung cancer, glioma, and other human malignancies. It is important to note, however, that since most conventional (exomic) sequencing platforms would not routinely detect this particular EGFR alteration (due to its intronic breakpoints), these numbers are likely an underestimate, and the true prevalence of the EGFR-KDD remains unknown.

The EGFR-KDD is Oncogenic

We expressed EGFR-KDD in NR6 and Ba/F3 cells **(Figures 3.1B–E)**. We observed expression of EGFR-KDD at the expected molecular weight as compared to EGFR wild-type (WT) and the well-characterized EGFR-L858R mutation **(Figures 3.1B and D)**. In contrast to EGFR-WT, both EGFR-L858R and EGFR-KDD displayed high levels of autophosphorylation. Consistent with these data, EGFR-KDD protein is constitutively autophosphorylated in the absence of serum in A1235 cells, a glioma cell line which harbors endogenous EGFR-KDD (11) **(Figure S3.4)**. To address whether EGFR-KDD is an oncogenic alteration, we tested its ability to confer anchorage-independent growth to NR6 cells. EGFR-KDD significantly increased colony formation in soft agar as compared to both EGFR-WT and the known oncogenic EGFR-L858R mutation **(Figure 3.1C and Figures S3.5A–D)**. Expression of a kinase dead version of the EGFR-KDD (called EGFR-KDD-dead) abrogated the growth of NR6 cells in soft agar, consistent with the requirement of kinase activity for anchorage-independent growth. In parallel, we expressed the same EGFR variants in Ba/F3 cells **(Figure 3.1D)**. EGFR-KDD, but not its kinase dead counterpart, induced IL-3-independent proliferation of Ba/F3 cells, an activity phenotype associated with the transforming function of other oncogenic tyrosine kinases (19) **(Figure 3.1E)**. As previously reported, EGFR-L858R, but not EGFR-WT, was able to support IL-3 independent growth of Ba/F3 cells (19).

Computational Modeling Demonstrates that EGFR-KDD can form Intra-Molecular Dimers

To provide insight into the mechanism of activation of EGFR-KDD, we examined the structure of the EGF receptor. The EGFR tyrosine kinase is known to be activated either due to increased local concentration of EGFR (e.g., as a result of ligand binding or overexpression), mutations in the activation loop (e.g., L858R) (20), or through formation of asymmetric (N-lobe to C-lobe) inter-molecular dimers between two EGFR proteins (21). Given the presence of two tandem in-frame kinase domains within the EGFR-KDD structure, we hypothesized that EGFR-KDD could form an intra-molecular dimer. To test this hypothesis, we modeled the EGFR-KDD based on the available experimental structure of the active asymmetric EGFR dimer (21). Conformational loop sampling with Rosetta demonstrates that the linker between the tandem tyrosine kinase domains allows for the proper positioning of the two domains necessary for asymmetric dimerization and intra-

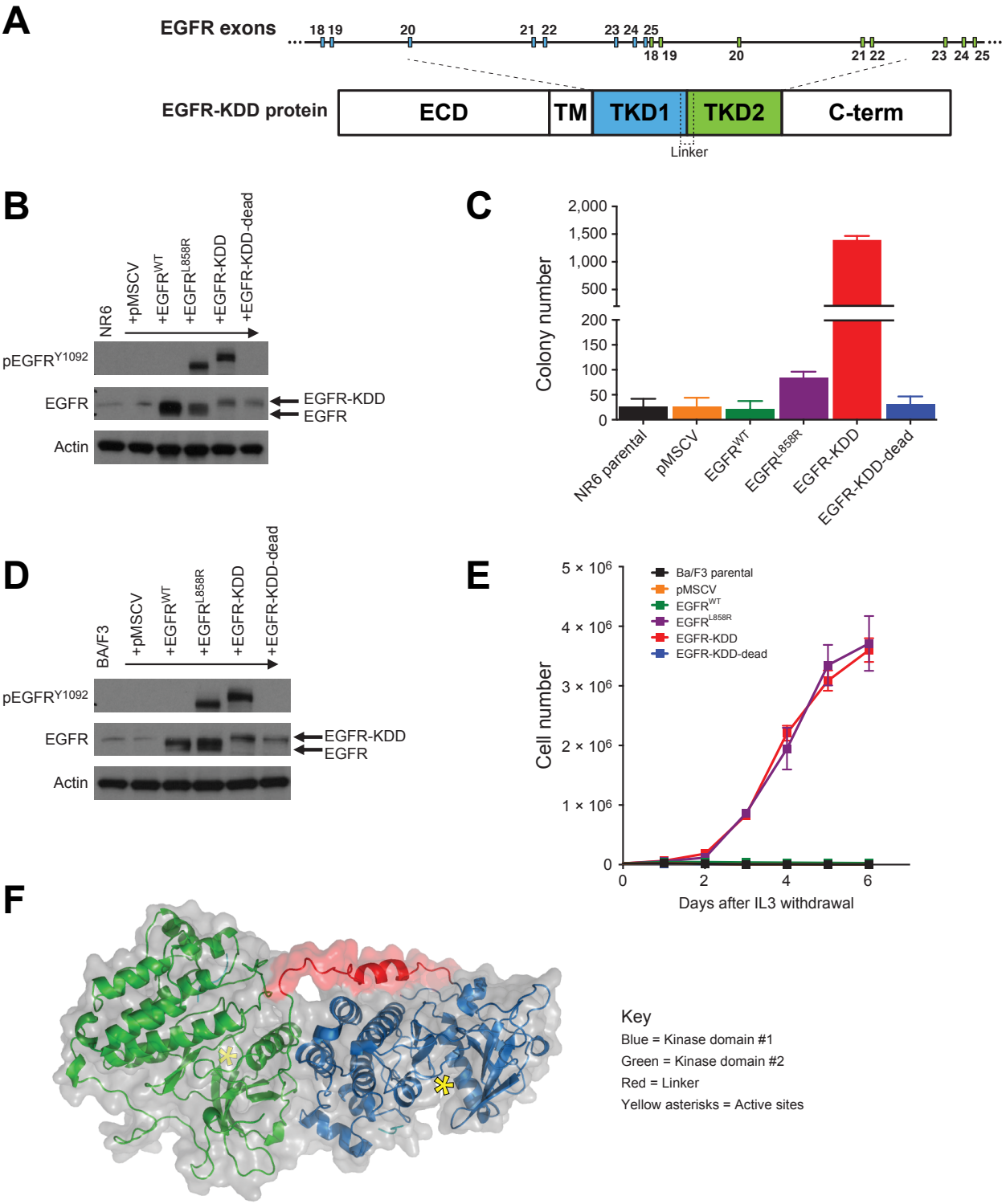


Figure 3.1: The EGFR-KDD is an oncogenic EGFR alteration.
A, schematic representation of EGFR-KDD depicting the genetic and protein domain structures. ECD, extracellular domain; TM, transmembrane helix; TKD1, first tyrosine kinase domain; TKD2, second tyrosine kinase domain; C-term, carboxyl terminus region. Blue, EGFR exons 18–25 #1; green, EGFR exons 18–25 #2. **B,** representative Western blot of NR6 cells stably expressing indicated EGFR constructs. EGFR-KDD-dead is a kinase-dead version of EGFR-KDD. **C,** NR6 cells stably expressing the indicated constructs (pMSCV = vector only) were plated in triplicate in soft agar, grown for 15 days, and quantified for colony formation. **D,** representative Western blot of BA/F3 cells expressing indicated EGFR constructs. **E,** BA/F3 cells transfected with indicated constructs (pMSCV = vector only) were grown in the absence of IL3 and counted every 24 hours. **F,** ribbon diagram and space-filling model of the EGFR-KDD kinase domains (GLY 696—PRO 1370) illustrating the proposed mechanism of autoactivation. This figure corresponds to Figure 1 from Gallant et al. Cancer Discovery, 2015 (22).

Dataset	Identification #	Age	Gender	Reported diagnosis
Foundation Medicine	FM-1	52	Female	Lung adenocarcinoma
	FM-2*	33	Male	Lung adenocarcinoma
	FM-3	53	Female	Lung adenocarcinoma
	FM-4	57	Female	Lung adenocarcinoma
	FM-5	29	Female	Lung NSCLC (NOS)
	FM-6	53	Female	Brain astrocytoma
	FM-7	49	Male	Brain glioblastoma
	FM-8	54	Male	Brain glioblastoma
	FM-9	2	Female	Kidney Wilms' tumor
	FM-10	63	Female	Peritoneal serous carcinoma
	FM-11	27	Female	Soft tissue sarcoma (NOS)
TCGA	TCGA-49-4512	69	Male	Lung adenocarcinoma
	TCGA-12-0821	62	Female	Brain glioblastoma
MSKCC	MSKCC-1*	33	Male	Lung adenocarcinoma
	MSKCC-2	67	Female	Lung adenocarcinoma
	MSKCC-3†	53	Male	Brain glioblastoma

Table 3.1: The EGFR-KDD is a recurrent alteration.
Characteristics of EGFR-KDD exons 18–25 patients from Foundation Medicine, TCGA, and Memorial Sloan Kettering Cancer Center datasets. NOS = not otherwise specified. * = index patient. † = this patient's tumor also contained high level amplification of wild-type EGFR and an EGFR G719C mutation. The EGFR-KDD and EGFR G719C alterations were below the level of wild-type EGFR amplification, and presumably reflect sub-clonal populations. This table corresponds to Table 1 from Gallant et al. Cancer Discovery, 2015 (22).

molecular EGFR activation **(Figure 3.1F)**. Therefore, our model suggests that EGFR-KDD is an oncogenic variant of the EGF receptor likely by virtue of its ability to form intramolecular asymmetric (activated) dimers. Although modeling demonstrates that the EGFR-KDD is geometrically capable of forming intra-molecular asymmetric dimers, further experimental data would be needed to confirm this mechanism

The EGFR-KDD can be Therapeutically Targeted with Existing EGFR-TKIs

We sought to determine whether EGFR-TKIs are an effective therapeutic strategy for tumors harboring the EGFR-KDD. We treated Ba/F3 cells expressing EGFR-WT, EGFR-L858R, and EGFR-KDD with erlotinib (1st generation reversible EGFR-TKI) (23), afatinib (2nd generation irreversible inhibitor of EGFR/HER2) (24), and AZD9291 (3rd generation mutant specific EGFR-TKI) (25) to assess the effects of these inhibitors on the autophosphorylation (and by extension, the kinase activity) and downstream signaling properties of the EGFR kinase. All three EGFR-TKIs were able to inhibit EGFR-KDD tyrosine phosphorylation in a dose-dependent manner, albeit to different levels **(Figure 3.2A)**. Afatinib was the most potent inhibitor of EGFR-KDD autophosphorylation, at doses similar to those required for EGFR-L858R inhibition. Activation of downstream MAPK signaling was also inhibited in Ba/F3 cells expressing EGFR-KDD, as shown by decreased ERK phosphorylation after drug treatment. Similar results were observed in 293T cells transfected with EGFR variants **(Figure S3.6A)** and in A1235 cells which harbor endogenous EGFR-KDD **(Figure S3.6B)**. To determine whether inhibition of EGFR autophosphorylation translated to inhibition of cellular proliferation, we treated Ba/F3 cells expressing various EGFR constructs with erlotinib, afatinib, and AZD9291. All three distinct EGFR-TKIs effectively inhibited the growth of EGFR-KDD Ba/F3 cells **(Figure 3.2B and Table S3.1)**. Consistent with our signaling data, the growth inhibition observed was most pronounced with afatinib. Analogous results were seen in A1235 cells **(Figure S3.6C)**. Together, these results show that the EGFR-KDD can be potently inhibited by afatinib, leading to decreased cell viability.

Treatment of the Index Patient with Afatinib

Since there were no data regarding the use of EGFR-TKIs or monoclonal antibodies in the setting of a tumor harboring the EGFR-KDD, the index patient was initially treated with standard first line chemotherapy for stage IV lung adenocarcinoma (cisplatin/pemetrexed/bevacizumab) (26). However, at the time of disease progression on this treatment regimen, the patient was treated with afatinib. Immediately after beginning afatinib, the patient reported feeling markedly better with improvements in his symptoms of cough and fatigue. After two cycles of afatinib, the patient showed a partial radiographic response (~50% tumor shrinkage) per RECIST criteria (27) **(Figure 3.3A)**. This clinical activity is consistent with our in vitro studies and provides rationale for further clinical investigation.

Acquired Resistance to Afatinib

The index patient developed acquired resistance to afatinib after 7 cycles of therapy **(Figure 3.3A)**. This duration of response is in line with the typical responses observed in other EGFR-mutant lung cancers treated with EGFR-TKIs (1-3). Molecular profiling was performed on the afatinib resistant tumor biopsy sample, and this testing uncovered significant amplification of the EGFR-KDD allele as the only genomic alteration that differed from his pre-treatment tumor sample **(Figure 3.3B)**. This sequencing result was confirmed via EGFR fluorescence in situ hybridization (FISH) **(Figure 3.3C)**. Amplification of the mutant EGFR allele has been reported as a mechanism of acquired resistance in the context of canonical EGFR mutations (e.g. exon 19 deletion, L858R) in lung cancer (28). Therefore, amplification of the EGFR-KDD in this post-treatment sample suggests an EGFR-dependent mechanism of resistance, thereby further validating this EGFR alteration as a driver and therapeutic target in patients.

Discussion

Although much progress has been made over the past several decades, lung cancer remains the leading cause of cancer deaths worldwide (29). The discovery of oncogenic EGFR mutations that sensitize lung cancers to EGFR-TKIs heralded the dawn of molecularly-targeted therapy in this disease (30-32). Indeed, numerous phase III studies have now documented that patients with EGFR-mutant tumors derive significant clinical and radiographic benefit from treatment with EGFR-TKIs, such as gefitinib, erlotinib, and afatinib (1-3). The majority of previously described activating mutations in EGFR are a series of small deletions in exon 19 or leucine to arginine substitutions at position 858 (L858R) in exon 21 (33). However, because mutations

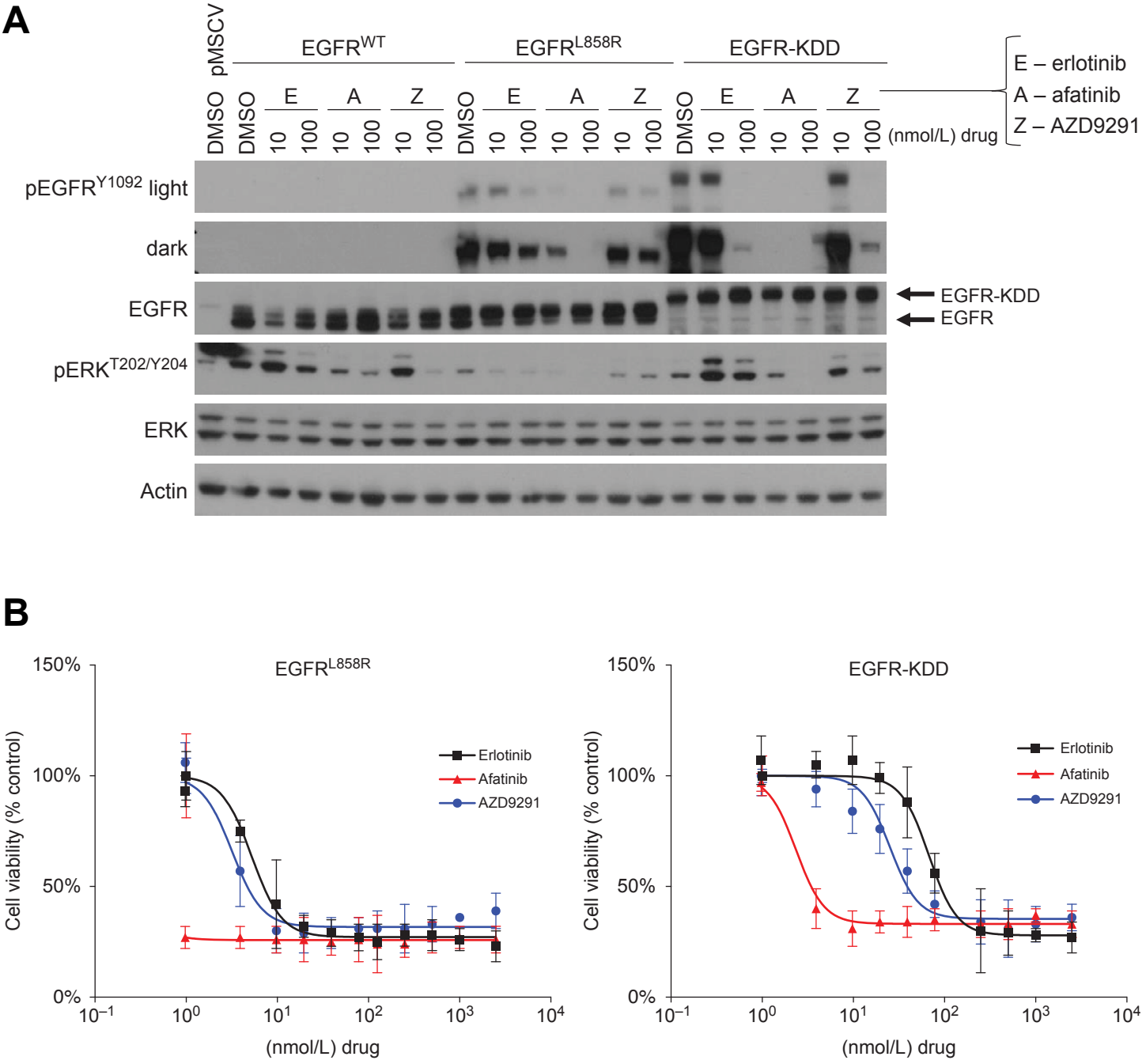


Figure 3.2: The EGFR-KDD can be therapeutically targeted with existing EGFR TKIs.
A, Ba/F3 lines stably expressing EGFRWT, EGFRL858R, or EGFR-KDD were treated with increasing doses of erlotinib, afatinib, or AZD9291 for 2 hours and lysed for Western blot analysis with the indicated antibodies. **B,** Ba/F3 lines stably expressing EGFRL858R or EGFR-KDD were treated with increasing doses of erlotinib, afatinib, or AZD9291 for 72 hours. Cell titer blue assays were performed to assess cell viability. Each point represents quadruplicate replicates. Data are presented as the mean percentage of viable cells com- pared with vehicle control ± SD. This figure corresponds to Figure 2 from Gallant et al. Cancer Discovery, 2015 (22).

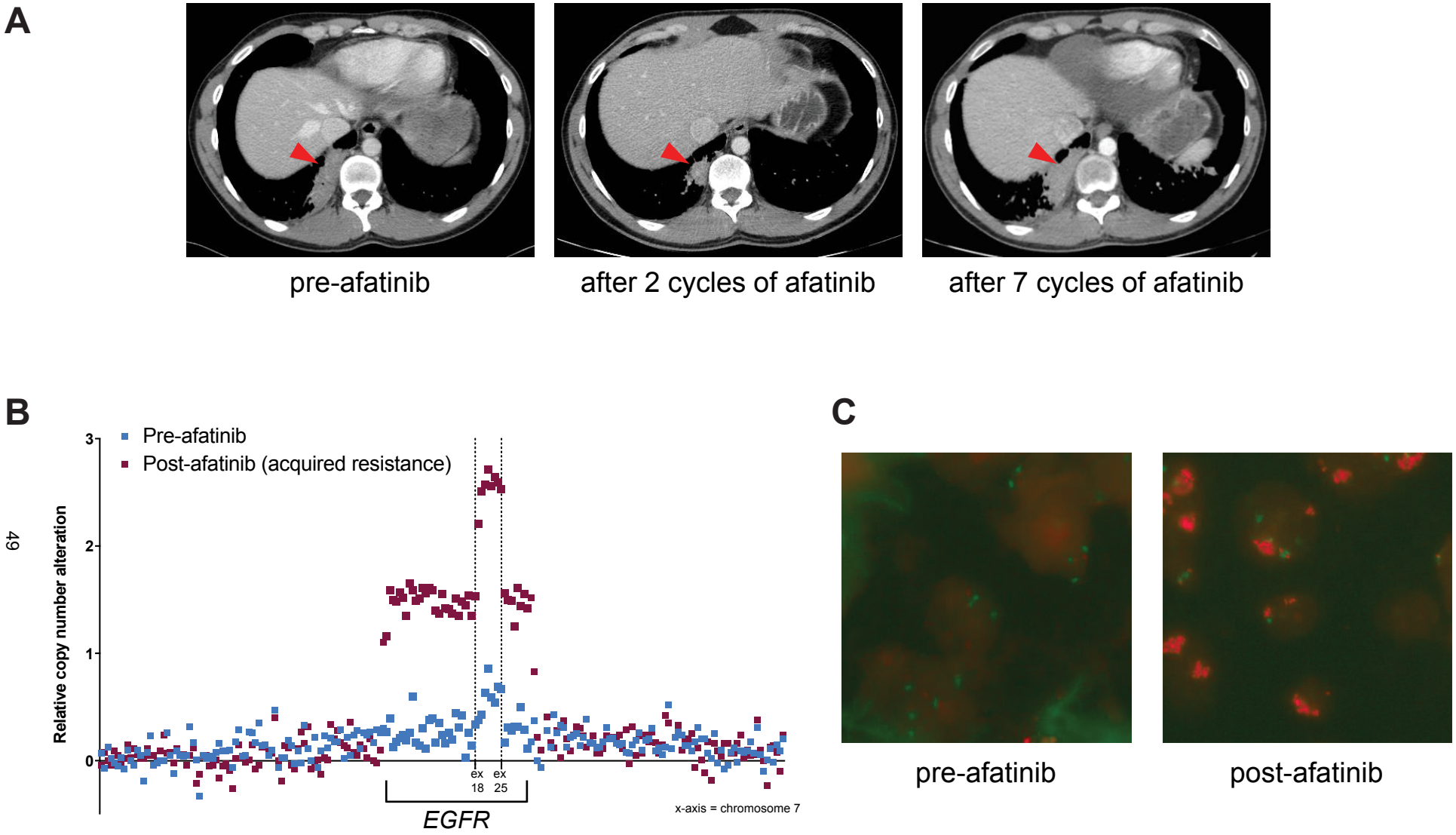


Figure 3.3: Serial chest CT scans of 33-year- old male with lung adenocarcinoma harboring EGFR-KDD documenting response to afatinib and subsequent acquired resistance.

A, left image, patient images after six cycles of cisplatin/pemetrexed/bevacizumab (largest mass diameter = 6.62 cm). Middle image, patient images after two cycles of afatinib (largest mass diameter = 2.72 cm). Right image, patient images after seven cycles of afatinib (largest mass diameter = 6.20 cm). The red arrowheads are pointing to the largest mass used for RECIST evaluation. **B,** copy-number data from FoundationOne NGS targets along chromosome 7 demonstrating amplification of the EGFR-KDD allele at the time of acquired resistance to afatinib (maroon squares) compared with the pre-afatinib tumor biopsy sample (blue squares). The x-axis represents chromosome 7. **C,** EGFR FISH of pre-afatinib (left image) and post-afatinib (right image) tumor biopsy samples used for the NGS analysis shown in B. Pre-afatinib, 1.6 copies of EGFR per chromosome 7 centromere (1.6 EGFR/CEP7); post-afatinib, 4.2 EGFR/CEP7. Green puncta, CEP7; red puncta, EGFR. This figure corresponds to Figure 3 from Gallant et al. Cancer Discovery, 2015 (22).

historically have been interrogated by 'hot-spot' PCR-based methods, most EGFR mutations are biased to fall between exons 18 and 21.

Here, we report the EGFR kinase domain duplication (EGFR-KDD) for the first time in lung cancer. This EGFR alteration contains an in-tandem and in-frame duplication of exons 18-25, which encode the entire EGFR kinase domain. We demonstrate that the EGFR-KDD is an oncogenic and constitutively activated form of the EGF receptor. We provide a structural model whereby the EGFR-KDD can be activated by virtue of asymmetric intra-molecular dimerization, as opposed to the typical asymmetric inter-molecular dimerization between adjacent EGFR molecules. Furthermore, we demonstrate that the EGFR-KDD can be therapeutically targeted with EGFR-TKIs, many of which are already FDA-approved. Additionally, we establish that the EGFR-KDD alteration is not only recurrent in lung cancer but also in gliomas and other tumor types.

Most importantly, we provide the first documentation of a clinical response to EGFR inhibitor therapy in a lung cancer patient whose tumor harbored the EGFR-KDD alteration. In contrast to lung cancer patients with more common EGFR mutations (e.g., exon 19 deletion and L858R), prior to our study, there was no precedent to support the use of EGFR inhibitors in patients whose lung tumors harbor the EGFR-KDD alteration. Therefore, our patient was not eligible for first-line EGFR-TKI therapy and was instead treated with platinum based chemotherapy, the standard of care for metastatic lung adenocarcinoma (34). The index patient was treated with afatinib for second-line therapy because this agent is FDA approved for the treatment of EGFR-mutant NSCLC and because, interestingly, afatinib was consistently the most potent EGFR-TKI against the EGFR-KDD across several different assays. This was not unexpected as it has been shown that various EGFR mutations or truncations have differential sensitivity to EGFR-TKIs due to nuanced structural differences (35). The marked tumor regression and improved functional status seen with afatinib therapy provides important clinical validation for the EGFR-KDD as an actionable alteration in lung cancer. Overall, the index patient derived a partial response to afatinib for 7 cycles, after which there was progression of disease. His tumor was re-biopsied and found to contain amplification of the EGFR-KDD in this post-treatment sample—suggesting an EGFR-dependent mechanism of resistance and validating this EGFR alteration as a driver and therapeutic target in patients.

This case also reinforces the need to functionally validate and discern the therapeutic 'actionability' of genomic alterations as increasingly sophisticated methods of NGS-based assays are being brought to the forefront of clinical diagnostics. Notably, the EGFR-KDD would not have been recognized by the 'hot-spot' PCR-based methods for EGFR mutational analysis described above. Therefore, it is not surprising that this EGFR alteration had not previously been detected. In fact, the EGFR-KDD in the index patient's tumor was identified because of a fortuitous intronic breakpoint that lay close to the exonic probes of the NGS diagnostic assay. Therefore, we hypothesize that the EGFR-KDD may have gone and may continue to go undetected in other tumors because standard (exomic) sequencing platforms do not target this particular alteration due to its large intragenic repeat and intronic breakpoint. Thus, while our data show that the EGFR-KDD is recurrent in multiple tumor types, this alteration would not be detected with currently approved PCR-based methods and is difficult to detect using standard exomic sequencing, consequently making our reported frequency a likely underestimate. Future design of tumor sequencing platforms should incorporate intronic probes for EGFR in order to more reliably detect the EGFR-KDD.

In summary, we have identified a recurrent, oncogenic, and drug sensitive EGFR-KDD in a subset of patients with lung cancer, glioma, and other cancer types. Our findings provide a rationale for therapeutically targeting this unique subset of EGFR-KDD driven tumors with EGFR tyrosine kinase inhibitors, many of which are already FDA-approved. Therefore, findings from our studies are expected to be rapidly translated into the clinic as they provide a new avenue for precision medicine in these difficult-to-treat malignancies.

Methods

Cell culture

The human lung adenocarcinoma cell line, II-18, has been previously described and was verified to harbor the EGFR-L858R mutation direct by cDNA sequencing (36). A1235 cells were a kind gift from Drs. Fenstermaker and Ciesielski (Roswell Park Cancer Institute, Buffalo, NY) (11). 293T cells were purchased from ATCC. BA/F3 cells were purchased from DSMZ. Plat-GP cells were purchased from CellBioLabs. NR6 cells were a kind gift from Dr. William Pao (37). II-18 and BA/F3 cells were maintained in RPMI 1640 medium (Mediatech, Inc.). A1235, 293T, and NR6 cells were maintained in DMEM (Gibco). Media was supplemented with 10% heat inactivated fetal bovine serum (Atlanta Biologicals) and penicillin-streptomycin (Mediatech, Inc.) to final concentrations of 100 U/ml and 100 µg/ml, respectively. The BA/F3 cell line was supplemented with 1

ng/mL murine IL-3 (Gibco). The Plat-GP cell line was cultured in the presence of 1 μg/mL blasticidin (Gibco). All cell lines were maintained in a humidified incubator with 5% CO2 at 37°C and routinely evaluated for mycoplasma contamination. Besides verifying the status of EGFR mutations in cell lines, no additional cell line identification was performed.

Compounds

Erlotinib, afatinib, and osimertinib were purchased from Selleck Chemicals.

EGFR Plasmid Construction

A cDNA encoding the EGFR-KDD (exon 18-25 tandem duplication) was synthesized by Life Technologies based on the consensus coding sequence. The pMSCV-puro vector backbone (Clontech) was used to construct all retroviruses. Assembly of pMSCV-puro-EGFR-WT and pMSCV-puro-EGFR-L858R was previously described (38). The EGFR-KDD was subcloned from the pMA synthesis vector (Life Technologies) into the HpaI site of pMSCV-puro using blunt end ligation. The pcDNA3.1 vector was used for transient expression experiments in 293T cells. Assembly of pcDNA-EGFR-WT and pcDNA-EGFR-L858R was previously described (32). EGFR-KDD was subcloned from the pMA synthesis vector (Life Technologies) into the pcDNA3.1/V5 vector (Life Technologies) using Gateway cloning (Life Technologies). All plasmids were sequence verified in the forward and reverse directions. EGFR-KDD-dead was constructed using multi-site-directed mutagenesis (Agilent) of the catalytic lysines (K745 and K1096) to methionines using the following primer:
KM-F: 5'–AAAGTTAAAATTCCCGTCGCTATCATGGAATTAAGAGAAGCAAC–3'.
The plasmids were fully re-sequenced in each case to ensure that no additional mutations were introduced.

BA/F3 and NR6 Cell Line Generation

The empty pMSCV-puro retroviral vector or pMSCV-puro vectors encoding EGFR (either EGFR-WT, EGFR-L858R, EGFR-KDD, or EGFR-KDD-dead) were transfected, along with the envelope plasmid pCMV-VSV-G (CellBioLabs), into cells Plat-GP packaging cells (CellBioLabs). Viral media was harvested 48 hours after transfection, spun down to remove debris, and supplemented with 2 μg/mL polybrene (Santa Cruz). 2.5×10^6 BA/F3 cells (or 1×10^6 NR6 cells) were re-suspended in 10 mL viral media. Transduced cells were selected for 1 week in 2 μg/mL puromycin (Invitrogen) and BA/F3 cells were selected for an additional week in the absence of IL-3. Stable polyclonal populations were used for experiments and routinely tested for expression of EGFR constructs.

Antibodies and Immunoblotting

The following antibodies were obtained from Cell Signaling Technology: phospho-EGFR tyrosine 1068 (#2234, 1:1000 dilution), EGFR (#4267, 1:1250 dilution), phospho-ERK threonine 202/tyrosine 204 (#9101, 1:2000 dilution), ERK (#9102, 1:2000 dilution), HRP-conjugated anti-mouse (#7076, 1:5000 dilution), and HRP-conjugated anti-rabbit (#7074, 1:5000 dilution). The actin antibody (#A2066, 1:5000 dilution) was purchased from Sigma-Aldrich. The EGFR antibody (#610017, 1:2000 dilution) was purchased from BD Pharmingen. For immunoblotting, cells were harvested, washed in PBS, and lysed in RIPA buffer (150 mM NaCl, 1% Triton-X-100, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris·HCl, pH 8.0) with freshly added 40 mM NaF, 1 mM Na-orthovanadate, and protease inhibitor mini tablets (Thermo Scientific). Protein was quantified using protein assay reagent and a SmartSpec plus spectrophotometer (Bio-Rad) per the manufacturer's protocol. Lysates were subjected to SDS-PAGE followed by blotting with the indicated antibodies and detection by Western Lightning ECL reagent (Perkin Elmer).

Cell Viability, Counting, and Clonogenic assays

For viability experiments, cells were seeded at 5,000 cells/well in 96-well plates and exposed to treatment the following day. At 72 hours post drug addition, Cell Titer Blue reagent (Promega) was added, and fluorescence at 570 nm was measured on a Synergy MX microplate reader (Biotek) according to the manufacturer's instructions. For cell counting experiments, cells were seeded at 20,000 cells/well in 24-well plates in the presence or absence of 1 ng/mL IL-3. Every 24 hours, cells were diluted 20 fold and counted using a Z1 Coulter Counter (Danaher). For clonogenic assays, cells were seeded at 5,000 cells/well in 24-well plates and exposed to treatment the following day. Media and inhibitors were refreshed every 72 hours, and cells were grown for 1 week or until confluence in control wells. Cells were fixed with 4% v/v formalin and stained with 0.025% crystal violet. Dye intensity was quantified using an infrared imaging system (LI-COR).

Viability assays were set up in quadruplicate, clonogenic assays were set up in triplicate, and cell-counting assays in duplicate. All experiments were performed at least three independent times. Data are presented as the percentage of viable cells compared to control (vehicle only treated) cells. Regressions were generated as sigmoidal dose-response curves using Prism 6 (GraphPad) by normalizing data and constraining the top to 100.

Soft Agar Assays

1.5 mL of 0.5% agar/DMEM was layered in each well of a 6-well dish. A total of 10,000 NR6 cells in 1.5 mL of 0.33% soft agar/DMEM were seeded on top of the initial agar and allowed to grow for 15 days. Each cell line was plated in triplicate. Colonies were counted using GelCount (Oxford Optronix) with identical acquisition and analysis settings.

Transient Transfections

293T cells were transfected using Lipofectamine 2000 (Life Technologies) per the manufacturer's recommendations. 24 hours post-transfection, cells were treated for two hours, gathered for western blot analysis, and prepared as described above.

Structural Modeling of the EGFR-KDD

The linker residues in the EGFR-KDD protein sequence, FFSSPSTSRTPLLSSLLVEPLTPS, were defined as those between the two kinase domains and not present in the X-ray crystal structure of EGFR in its allosterically activated dimeric form, 2GS6.pdb (21).This linker was manually built and placed into the EGFR crystal structure using PyMOL 1.5.0.3. The conformational space for the linker was then sampled using the loop modeling functionality of Rosetta version 2015.05 (39). 20,000 independent loop modeling runs were performed using kinematic closure. The best model from these runs was still lacking residues 748-750, 992-1004, 1099-1101, and 1343-1355 because these surface-exposed loops were not resolved in the experimental structure. These four loops were reconstructed using Modeller 9.14, and the model with the lowest DOPE score was selected (40). Those loops were then sampled for an additional 20,000 runs using Rosetta to generate the complete energy-minimized model of EGFR-KDD residues GLY 696 - PRO 1370.

Tumor Biopsy Samples

All patient tumor biopsy samples were obtained under Institutional Review Board (IRB) approved protocols (Vanderbilt University IRB# 050644). Written informed consent was obtained from the index patient. All samples were de-identified, protected health information reviewed according to the Health Insurance Portability and Accountability Act (HIPAA) guidelines, and studies conducted in accordance with the Declaration of Helsinki.

Validation of EGFR-KDD

Nucleic acids were isolated from formalin-fixed, paraffin-embedded tissue sections using the AllPrep DNA/RNA FFPE Kit (Qiagen). Nucleic acids were isolated from A1235 cell pellets using RNAeasy and DNAeasy kits (Qiagen). RT-/PCR were used to amplify the 'breakpoint' region following exon 25 and preceding exon 18 of EGFR-KDD. RT-PCR was performed with the SuperScript III one-step RT-PCR system (Life Technologies) using 100 ng RNA and the following primers:
Ex24-25F: 5'–GTTTTCCCAGTCACGACCTTGTCATTCAGGGGGATG–3'
Ex18-19R: 5'–CAGGAAACAGCTATGACGGGATCCAGAGTCCCTTATACA–3'
PCR was performed with HotStarTaq master mix (Qiagen) using 100 ng DNA and the following primers:
Ex25F: 5'–GTTTTCCCAGTCACGACAGAACCAAGGGGGATTTCAT–3'
Ex18R: 5'–CAGGAAACAGCTATGACTTGGTCTCACAGGACCACTG–3'
RT-/PCR products, which included the EGFR-KDD 'breakpoint', were sequenced in both directions using M13 primers, BigDye Terminator chemistry, and 3730 DNA Analyzers (Applied Biosystems). Aside from the EGFRKDD,
no mutations or alterations were found in either the sequenced regions of the A1235 cell line or the patient sample (as compared to the EGFR CCDS).

Identifying EGFR-KDD in The Cancer Genome Atlas

Copy number data from the Broad Institute TCGA Genome Data Analysis Center 2015-04-02 run were visually inspected to identify samples with focal amplification of the EGFR-KDD region (exons 18-25). RNA-seq (e714b8a4-dd57-4b01-83fd-d3a9fb2d4ad1120 and c552b1e3-9158-4c4d-b02b-b16ff7903552) and whole-genome sequencing (27c2031a-39f1-473c-88a6-9e7a03cedf04) files were inspected to confirm the presence of tandem duplication reads. Raw data available at: doi:10.7908/C1K64H04 and doi:10.7908/C1MP525H.

EGFR Fluorescence in situ Hybridization (FISH)

EGFR FISH was performed by Integrated Oncology, a LabCorp specialty testing group, using the EGFR-CEP7 Dual Color DNA Probe (Vysis). A trained pathologist quantified the copies of CEP7 and EGFR in 60 nuclei per sample.

Statistics and Data Presentation

All experiments were performed using at least two technical replicates and at least three independent times (biological replicates). Each figure or panel shows a single representative experiment. Unless indicated otherwise, data is presented as mean ± standard deviation. Western blot autoradiography films were scanned in full color at 600 dpi, desaturated in Adobe Photoshop CC, and cropped in Powerpoint. EGFR-FISH images were normalized using ‘Match Color’ in Adobe Photoshop CC. No other image alterations were made.

References

1. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361,** 947–957 (2009).

2. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13,** 239–246 (2012).

3. Sequist, L. V. *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *Journal of Clinical Oncology* **31,** 3327–3334 (2013).

4. Carey, K. D. *et al.* Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. *Cancer Research* **66,** 8163–8171 (2006).

5. Shaw, A. T. *et al.* Crizotinib versus Chemotherapy in Advanced ALK-Positive Lung Cancer. *N Engl J Med* **368,** 2385–2394 (2013).

6. Shaw, A. T. *et al.* Crizotinib in ROS1-rearranged non-small-cell lung cancer. *N Engl J Med* **371,** 1963–1971 (2014).

7. Meador, C. B. *et al.* Beyond histology: translating tumor genotypes into clinically effective targeted therapies. *Clin. Cancer Res.* **20,** 2264–2275 (2014).

8. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology* **31,** 1023–1031 (2013).

9. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7,** 169–181 (2007).

10. Cheng, D. T. *et al.* Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J Mol Diagn* **17,** 251–264 (2015).

11. Fenstermaker, R. A., Ciesielski, M. J. & Castiglia, G. J. Tandem duplication of the epidermal growth factor receptor tyrosine kinase and calcium internalization domains in A-172 glioma cells. *Oncogene* **16,** 3435–3443 (1998).

12. Ciesielski, M. J. & Fenstermaker, R. A. Oncogenic epidermal growth factor receptor mutants with tandem duplication: gene structure and effects on receptor function. *Oncogene* **19,** 810–820 (2000).

13. Frederick, L., Wang, X. Y., Eley, G. & James, C. D. Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Research* **60,** 1383–1387 (2000).

14. Ozer, B. H., Wiepz, G. J. & Bertics, P. J. Activity and cellular localization of an oncogenic glioblastoma multiforme- associated EGF receptor mutant possessing a duplicated kinase domain. *Oncogene* **29,** 855–864 (2009).

15. Furgason, J. M. *et al.* Whole genome sequencing of glioblastoma multiforme identifies multiple structural variations involved in EGFR activation. *Mutagenesis* **29,** 341–350 (2014).

16. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).

17. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human

glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).

18. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155,** 462–477 (2013).

19. Greulich, H. *et al.* Oncogenic Transformation by Inhibitor-Sensitive and -Resistant EGFR Mutants. *PLoS Med.* **2,** e313 (2005).

20. Yun, C.-H. *et al.* Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **11,** 217–227 (2007).

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

22. Gallant, J. N. *et al.* EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib. *Cancer Discovery* **5,** 1155–1163 (2015).

23. Moyer, J. D. *et al.* Induction of apoptosis and cell cycle arrest by CP-358,774, an inhibitor of epidermal growth factor receptor tyrosine kinase. *Cancer Research* **57,** 4838–4848 (1997).

24. Li, D. *et al.* BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene* **27,** 4702–4711 (2008).

25. Cross, D. A. E. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discovery* **4,** 1046–1061 (2014).

26. Zornosa, C. *et al.* First-line systemic therapy practice patterns and concordance with NCCN guidelines for patients diagnosed with metastatic NSCLC treated at NCCN institutions. *J Natl Compr Canc Netw* **10,** 847–856 (2012).

27. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45,** 228–247 (2009).

28. Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Science Translational Medicine* **3,** 75ra26–75ra26 (2011).

29. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J Clin* **65,** 5–29 (2015).

30. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304,** 1497–1500 (2004).

31. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350,** 2129–2139 (2004).

32. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 13306–13311 (2004).

33. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research* **43,** D805–D811 (2015).

34. Ettinger, D. S. *et al.* Non-small cell lung cancer, version 6.2015. *J Natl Compr Canc Netw* **13,** 515–524 (2015).

35. Vivanco, I. *et al.* Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discovery* **2,** 458–471 (2012).

36. Ohashi, K. *et al.* Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor BRAF gene mutations but lack mutations in KRAS, NRAS, or MEK1. *Proceedings of the National Academy of Sciences* **109,** E2127–33 (2012).

37. Regales, L. *et al.* Dual targeting of EGFR can overcome a major drug resistance mutation in mouse models of EGFR mutant lung cancer. *J. Clin. Invest.* **119,** 3000–3010 (2009).

38. Red Brewer, M. *et al.* Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proceedings of the National Academy of Sciences* **110,** E3595–604 (2013).

39. Stein, A. & Kortemme, T. Improvements to robotics-inspired conformational sampling in rosetta. *PLoS ONE* **8,** e63090 (2013).

40. Eswar, N., Eramian, D., Webb, B., Shen, M.-Y. & Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426,** 145–159 (2008).

CHAPTER IV: FUTURE DIRECTIONS

Overview

This chapter is divided into two sections with the first focusing on EGFR-RAD51 and the second focusing on EGFR-KDD. Both sections describe ongoing work and future directions with an eye towards understanding the oncogenic proteins' mechanism of action. Specifically, each section details efforts aimed at creating cell line models of the EGFR re-arrangement, systems being developed to test for their mechanisms of activation, and hypotheses about how the molecules may be propagating downstream mitogenic signaling.

EGFR-RAD51: Ongoing Work and Future Directions

Creating a Cell Line Model of EGFR-RAD51

The main limitation in our current work on EGFR-RAD51 has to do with the cell lines used to model the oncogenic fusion. Firstly, both Ba/F3 and NR6 cell lines are murine cell lines that may not be representative of the human milieu in which EGFR-RAD51 leads to tumor formation. While the Ba/F3 model is a well validated predictor of oncogenic properties of a gene (1-3), and though growth in soft agar is correlated with the transformed phenotype (4), murine cells are more amenable to transformation than human cells (5,6). Our current models cannot identify potential differences that EGFR-RAD51 expression may have in human cells. Secondly, our pMSCV retroviral system (7) uses MMLV-based overexpression of EGFR-RAD51. Although the relative activity of promoters depends on the cell type and on the particular cDNA sequence, the long terminal repeats (LTRs) that drive MMLV gene expression are generally strong promoters (8). Therefore, it is conceivable that a relatively high amount of expressed protein is the factor responsible for cellular transformation. This presumed high expression is problematic because oncogenic fusions have been found to express at relatively low levels as compared to endogenous proteins (9). Lastly, it is possible that breakage of the RAD51 gene (which would occur in cells with EGFR-RAD51 fusions) is the key oncogenic property of EGFR-RAD51. Because loss or mutation of a single RAD51 allele is genotoxic, it could be that the gene fusion is oncogenic by virtue of its ability to cause RAD51-mediated loss of homologous recombination (HR) and increased amounts of chromosomal instability (10). Together, these factors point towards the need for a more nuanced approach toward the cellular study of EGFR-RAD51.

To circumvent cell line limitations, we are taking the following approaches (in increasing likelihood of success):

1. Searching for human lung cancer cell lines that harbor endogenous EGFR-RAD51;
2. Using the CRISPR/Cas9 system to create an in situ chromosomal rearrangement in AALE cells (normal lung epithelial cells);
3. Attempting to create cell lines from EGFR-RAD51 index patients; and
4. Reconstructing our stable EGFR-RAD51 cells using human bronchial epithelial cells (HBEC).

To find an available human cell line harboring an EGFR fusion, we will use a set of converging electronic database searches. First, we will search the Cancer Cell Line Encyclopedia (CCLE) (11) for cells that are sensitive to EGFR-targeted therapies and/or harbor predicted truncation mutations in EGFR. Cells with intronic mutations will have their sequences examined manually because fusions often occur in non-coding regions (12). The Genomics of Drug Sensitivity in Cancer (GDSC) (13,14) will also be queried for an alternative source of cell line sensitivity to EGFR-targeted therapies. Because the CCLE (174 lung cancer cell lines) and GDSC (217 cell lines) databases only have an overlapping fraction of the ≥ 300 available lung cancer cell lines (15), a manual search will need to be performed on the remaining cell lines. Because the N-term of EGFR is a recurrent feature in EGFR fusions and because functional protein expression is critical for driver oncogenes, we will perform a selective search for EGFR protein expression. First pass searches will be performed with the image searching capabilities of the most common search engines. Databases will be searched with the cell line names, and the terms EGFR, western blot, mutation, and/or lung cancer. Experiments and papers will be retrieved, experimental details checked, and cell lines expressing EGFR, as determined by known C-term antibodies, will be eliminated from the search. As the majority of antibodies used for western blotting bind the C-term of EGFR, and because EGFR is critical to lung epithelial cells and is usually expressed in all NSCLCs (16), any cell line that does not express C-term EGFR in a western blot image will be re-queried using MEDLINE and EMBASE. Finally, in parallel, GEO, and other databases, will be searched for any available genetic data on lung cancer cell lines not analyzed by the CCLE and GDSC. Sequencing data will be manually analyzed using validated fusion detection tools (17). Any potential hits will be acquired and tested for the presence of the expected EGFR fusion. While this search method has a low chance of success (based on our

published frequency of EGFR fusions in NSCLC (18)), finding an endogenous model of an EGFR fusion would be a boon to our research efforts.

As there currently exists no known lung cancer cell line harboring the endogenous EGFR-RAD51 fusion, we are utilizing CRISPR/Cas9 technology to create the fusion in normal human lung epithelial cells. This method has successfully been used to create endogenous lung cancer related fusions including the EML4-ALK and KI5B-RET inversions and the CD74-ROS1 translocation (19). Using standard methods (20), we created, synthesized, and tested 4 different Cas9 gRNAs: 2 targeting intron 24 of EGFR and 2 targeting intron 3 of RAD51 **(Figure 4.1A and Table S4.1)**. These gRNAs were subcloned into pX458 vectors and their efficacy tested in 293FT cells. After transient transfection into 293FT cells, all 4 gRNAs were able to induce double stranded DNA (dsDNA) breaks in their target locations as determined by SURVEYOR assay **(Figure 4.1B)**. As the transfection of a single plasmid containing multiple gRNAs is more efficient than transfecting two plasmids (21), multiplexed constructs, each containing one EGFR and one RAD51 gRNA, were created. These multiplexed constructs retained their ability to induce dsDNA breaks in 293FT cells, albeit, less efficiently **(Figure 4.1C)**. Using both multiplexed EGFR-RAD51 sgRNA constructs, we attempted to generate the translocation between chromosomes 15 (RAD51) and 7 (EGFR) in normal human lung cells (using the AALE cell line; hTERT and SV40 immortalized normal human bronchial epithelial cells (22)). To date, 140 putative EGFR-RAD51 AALE clones have been isolated and frozen[a]. While some clones tested positive for the fusion by breakpoint RT-PCR, none of the positive RT-PCR hits tested positive for expression of the protein by western blotting **(Figures 4.1D–E)**. In total, 48 putative clones have tested negative for the EGFR-RAD51 fusion; the other half remains untested and frozen as pellets or cells in DMSO. However, as the frequency of in vitro CRISPR/Cas9 induced translocations is rather low[b], current efforts should be maintained.

Because we already know of 5 patients with EGFR fusions, the most straightforward approach to creating a cell line would be to obtain tumor tissue from these patients. However, there are several logistical hurdles including patient privacy laws, IRB approval, and long distance shipping of tumor tissues. The ideal way to generate a cell line is from a large amount of fresh tumor tissue. Such tissue can be mechanically dissociated and grown into cell lines over time. Our lab has some success with this approach (23), and it has proven to be a generally reliable method to model human tumors in vitro (24). However, acquiring such tissue from our index patients is unlikely given the logistical hurdles mentioned above. Alternatively, a small piece of fresh tissue (such as a repeat biopsy) could be used with a feeder layer of cells and specialized media to grow indefinitely as conditionally reprogrammed cells (25-27). Unfortunately, this latter method introduces several artifacts into the cell culture system, such as the feeder cells. A potential solution could be to acquire freshly snap frozen and/or slowly frozen (in DMSO supplemented media) tumor biopsies; recent research has shown that tumor cells can be revived from such frozen tissue (28). As reviving frozen biopsies has been accomplished, and because we have documented the existence of several EGFR fusion patients, this approach may prove to be the most feasible and least artificial option.

The least sophisticated option (but still more nuanced than our original Ba/F3 and NR6 cells) would be to create a human model of EGFR-RAD51 using our retroviruses and established normal human lung cell lines. Using reagents in lab, we could re-create pMSCV based EGFR-RAD51 retrovirus and infect HBEC (29) or AALE cell lines at various MOI. Using varying doses of puromycin, we could then select human lung cells harboring EGFR-RAD51. Alternatively, we could subclone our construct into our lab's doxyxycline-inducible pRetroX-TRE3G system to create differently-expressing clones. Regardless of approach, selecting a variety of stable clones could allow us to see what happens at varying levels of EGFR-RAD51 expression; but, we would still be left with an overexpression system.

Testing Homologous Recombination in Cells Harboring EGFR-RAD51

Because RAD51 functions as a critical component of HR (30), we hypothesize that cancers driven by EGFR-RAD51 will be sensitive to therapies aimed at the DNA damage response. Being able to target an oncogenic fusion with targeted therapy (EGFR-TKIs) and synthetic lethality (DNA damaging agents) would be an exciting finding and would have implications in cases of acquired resistance. We hypothesize that EGFR-RAD51 antagonizes normal HR processes through RAD51 haploinsufficiency and/or through RAD51 mislocalization. To test whether EGFR-RAD51 could cause RAD51 haploinsufficiency, we would need a cell line of endogenous EGFR-RAD51, as outlined in the section above. However, testing sensitivity to therapies aimed at the DNA damage response and whether EGFR-RAD51 causes RAD51 mislocalization can be accomplished with our available reagents.

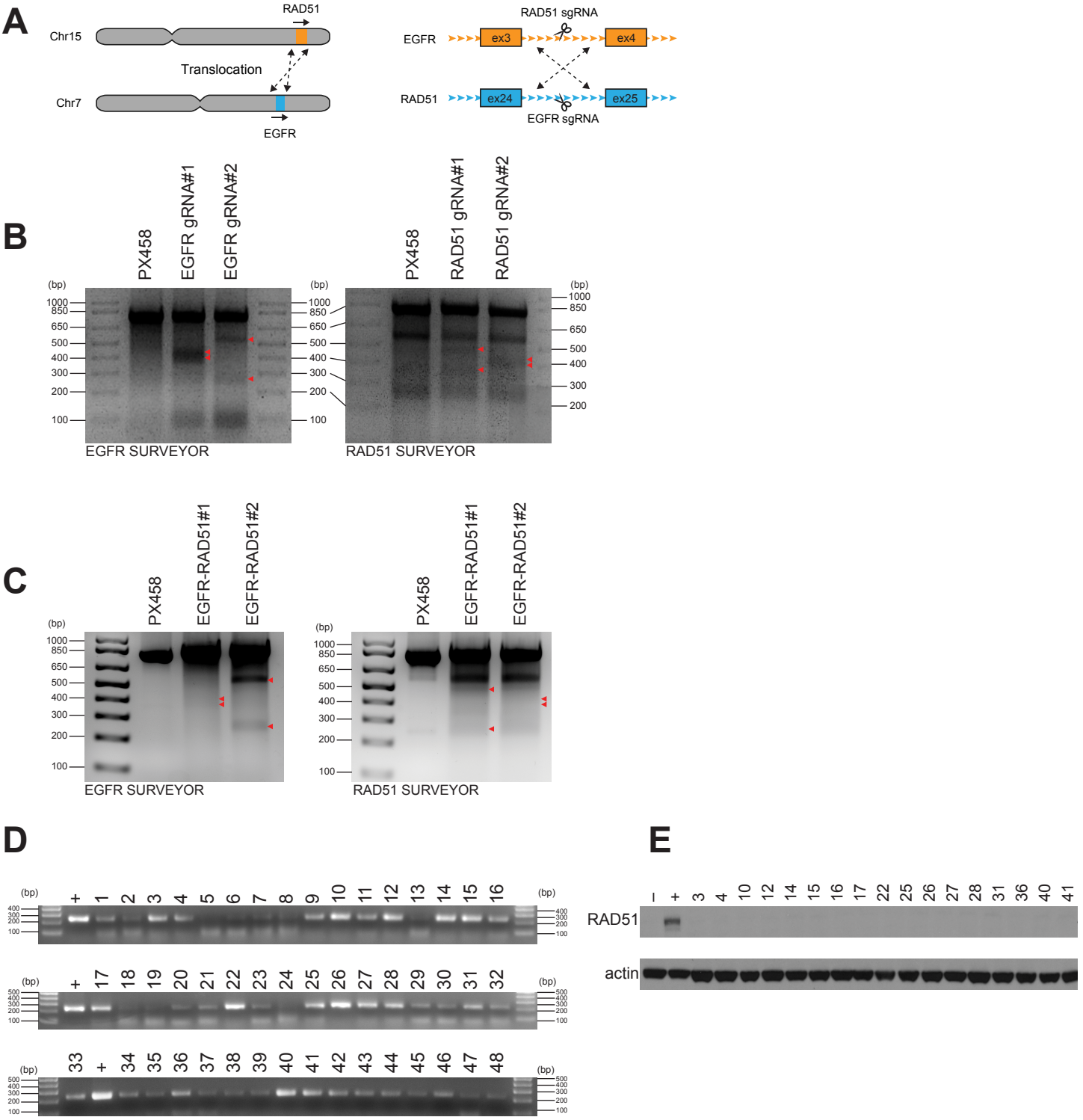


Figure 4.1: Utilizing the CRISPR/Cas9 system to promote EGFR-RAD51 formation in cells
A, schematic depicting the strategy for generating chromosomal rearrangements using Cas9 from S. pyogenes (SpCas9) that directs DNA cleavage at each targeted genomic site. The EGFR-RAD51 rearrangement results from a translocation between chromosomes 15 and 7. Shown are the intronic sites where Cas9 was targeted. Adapted from (19). **B,** SURVEYOR assay for single EGFR and RAD51 gRNAs. 293FT cells were transfected with constructs expressing Cas9 alone (PX458, vector) or with a targeting sgRNA (EGFR#1, #1, or RAD51#1, #2). For each sgRNA target site, the surrounding ~800 bp SURVEYOR assay to assess the formation of indels. **C,** SURVEYOR assay for multiplexed EGFR-RAD51 (#1, #2) PX458 constructs. Throughout: red triangles indicate predicted T7EI cleavage sites (31). D, RT-PCR of putative EGFR-RAD51 AALE clones. + = positive control = Ba/F3 EGFR-RAD51. **D,** western blot of putative EGFR-RAD51 AALE clones. + = positive control = Ba/F3 EGFR-RAD51; – = negative control = AALE. RAD51 blot is taken between 100 and 200 kD.

To test whether cancers driven by EGFR-RAD51 are sensitive to therapies aimed at the DNA damage response, we will treat our published IL-3 independent Ba/F3 cells with a variety of agents. Specifically, we will test whether the Ba/F3 cells harboring EGFR-RAD51 are hyper-sensitive to PARP inhibitors and relevant platinum based chemotherapy agents (cisplatin and carboplatin) compared to similar (Ba/F3) cells expressing pMSCV vector, EGFR-WT, or EGFR-L858R. We will also test commercially available inhibitors aimed at RAD51-mediated strand exchange, such as RI-1 and B02 (32). Hypersensitivity to these agents (which target the DNA damage response and/or RAD51 specifically) as opposed to EGFR inhibitors would point towards the RAD51 portion of EGFR-RAD51 impacting normal DNA repair.

Since endogenous RAD51 is known to traffic between the cytoplasm and the nucleus (33,34), we hypothesize that EGFR-RAD51 fusions could inhibit normal RAD51 function by trapping endogenous RAD51 in the cytoplasm, at the cell membrane[c]. Therefore, if Ba/F3 cells harboring EGFR-RAD51 are hyper-sensitive to DNA-damaging agents, then we will test to see whether this is due to RAD51 localization. We will test this hypothesis by immunofluorescently evaluating the subcellular localization of co-expressed RFP-tagged RAD51 and/or GFP-tagged EGFR-RAD51 upon treatment with DNA damaging agents. If cells harboring EGFR-RAD51 are hypersensitive to HR-targeted agents, then we would expect to see increased amounts of RAD51 in the cytoplasm (as opposed to the nucleus).

Determination of EGFR-RAD51 Mode ofActivation and Signaling

Our published computational modeling suggests that EGFR-RAD51 can form the asymmetric dimers required for EGFR kinase activation and that RAD51, a filamentous protein (36), could help induce EGFR-RAD51 multimerization (18). However, dimerization of EGFR-RAD51, as we modeled, could be driven by asymmetric dimerization of the EGFR tyrosine kinase domain, dimerization through the known RAD51 dimerization interface, or both **(Figure 4.2)**.

To test this hypothesis, we are repeating our experiments in Ba/F3 and NR6 cells, but with constructs deficient in EGFR-RAD51 dimerization. Specifically, we created EGFR asymmetric dimerization interface mutants (I706Q and V948R (37)), EGFR kinase dead mutants (K745M and D837N), RAD51 oligomerization interface mutants (F993E, A996W, A1097W, and A1099W (38-42)[d]); and combinations thereof **(Figure 4.2)**. We have created stable cell lines (Ba/F3 and NR6) with some of these various dimerization-impaired forms of EGFR-RAD51 and will test whether they still are able to transform cells. Regardless of their transforming potential, cells harboring these constructs will all be tested for downstream signaling, including signaling via the PI3K/AKT and MAPK pathways, in the absence or presence of EGF and EGFR-TKIs. If ablating the EGFR asymmetric dimer interface and/or RAD51 oligomerization prohibits transformation, we will confirm the interaction with immunoprecipitation (IP) experiments. Specifically, we will determine whether EGFR-RAD51 can co-IP with itself in the presence or absence of EGF. To accomplish this, we have created a series of dual tagged EGFR-RAD51 plasmids—pcDNA-EGFR-RAD51-V5/His and pTracer-EGFR-RAD51-FLAG/MYC—both of which are similar sized and driven by CMV promoters. We will add the same point mutations as indicated above, co/express these plasmids in EGFR-null CHO cells, and perform reciprocal IP experiments with or without EGF. Running the IP product out on a denaturing gel will allow us to determine what drives EGFR-RAD51 dimerization, and, by proxy, activation.

While these experiments could shed light on the nature of EGFR-RAD51 dimerization and activation, it still remains unclear as to how the fusion propagates downstream signaling. EGFR contains several (auto)phosphorylable tyrosines in the C-term tail of the receptor that positively regulate the transforming activity of EGFR by mediating downstream proliferative signaling. However, half of these tyrosine residues are absent in EGFR-RAD51 due to the fusion protein's lack of a C-term. Of the residues present in EGFR-RAD51, only tyrosines 869, 915, 944, and 978 have been experimentally verified as being able to bind adapters at physiologically relevant concentrations (43-45) **(Figure 4.3A)**. We will systematically transform these sites from tyrosines to phenylalanines (Y to F mutants) in order to test whether these residues are necessary for transformation. We plan to then generated Ba/F3 cells stably expressing these constructs to determine whether the mutants are able to support IL-3 independent growth. To support these results, we will similarly test anchorage independent growth in NR6 cells. As tyrosine 869 is the SRC phosphorylation site, in parallel, we will treat cells expressing EGFR-RAD51 'wild-type' (not containing any YF mutations) with SRC inhibitors, such as saracatinib (46), to determine if SRC

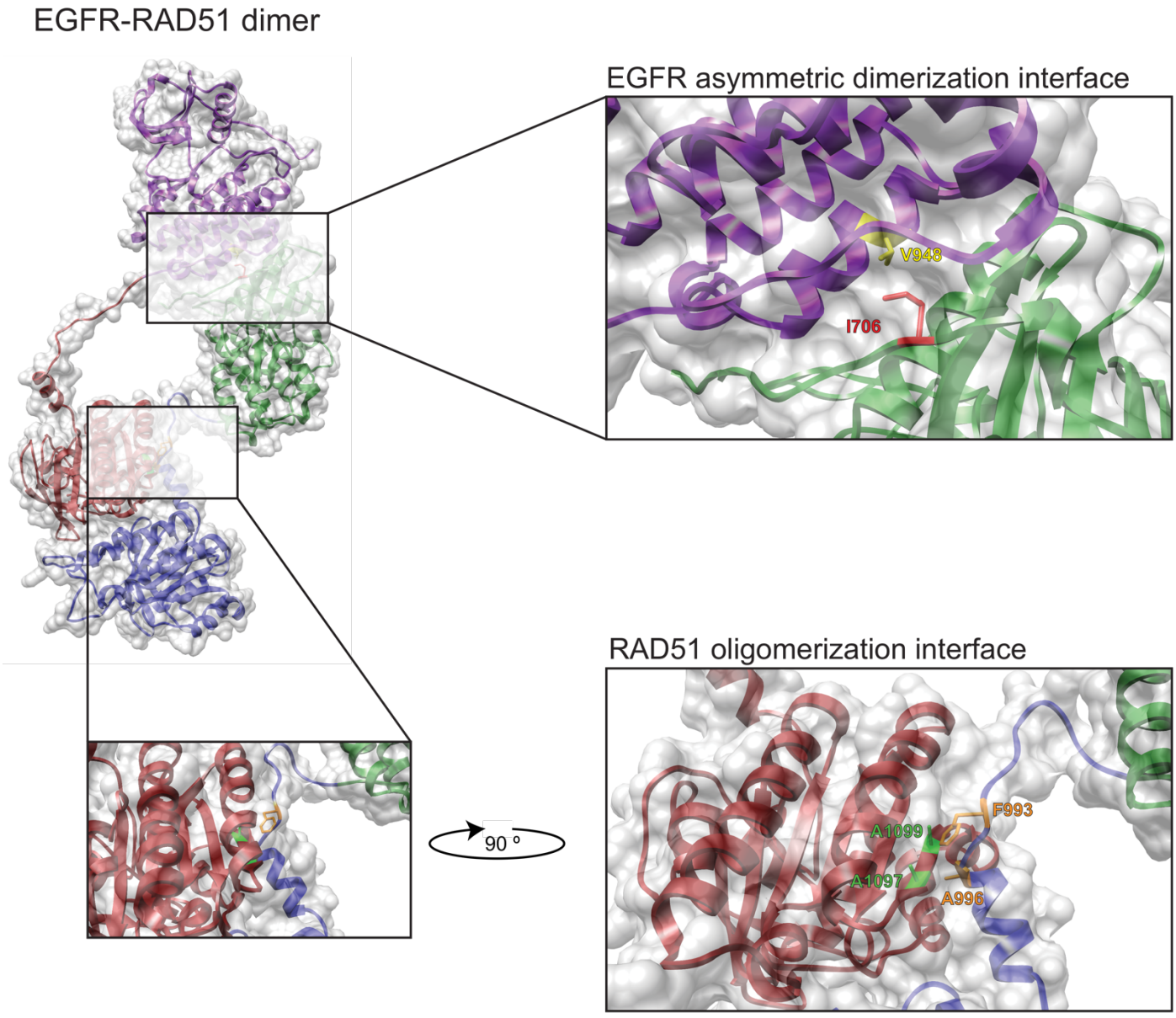
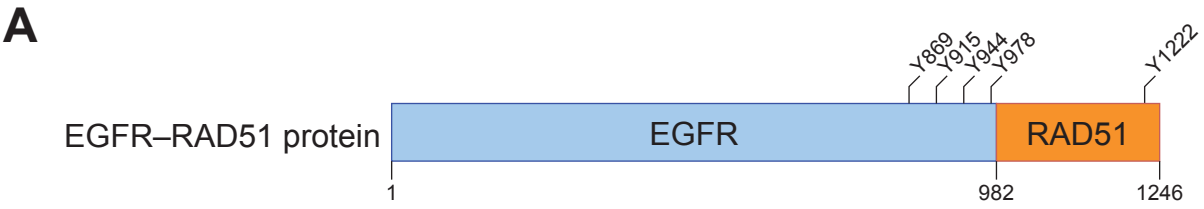
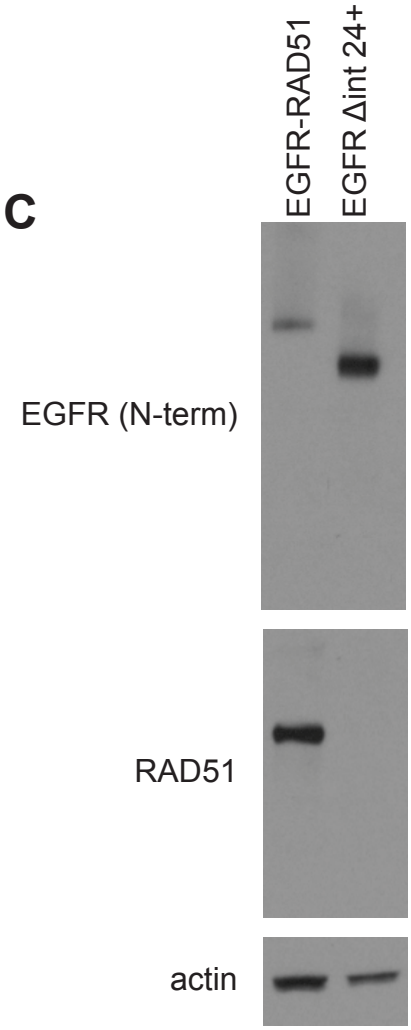


Figure 4.2: Proposed key residues for EGFR-RAD51 dimer and oligomer formation.
Left: ribbon diagram and space-filling model of the EGFR–RAD51 kinase domains illustrating the proposed mechanism of activation from (18). Purple, first EGFR kinase domain; green, second EGFR kinase domain; red, first RAD51 partner; blue, second RAD51 partner. Top insert: zoom of the EGFR asymmetric dimerization interface with key contact residues highlighted. Yellow, valine 948 (V948); red, isoleucine 706 (I706). Bottom insert: zoom of the RAD51 oligomerization interface with key contact residues highlighted. Orange, phenylalanine 993 (F993) and alanine 996 (A996); green, alanine 1097 (A1097) and alanine 1099 (A1099). Interaction between neighboring RAD51 molecules follows a 'key-lock' principle where F993 is intercalated between A1097 and A1099. A996 helps stabilize this key-lock and is evolutionarily conserved back to bacteria (38).





D

	-3	-2	-1	0	+1	+2	+3
EGFR-RAD51	C	K	I	Y	D	S	P
NCK1				pY	D	X	VP
ABL			IVL	Y	X	X	PF
SRC	E	E	I	Y	EG	X	F
EGFR	E	E	E	Y	F		

Figure 4.3: Determining how EGFR-RAD51 propagates downstream mitogenic signaling.
A, schematic representation of EGFR-RAD51 with physiologically relevant (auto)phosphorylable tyrosines marked. Numbers correspond to amino acid residues. Y, tyrosine residue. **B,** schematic representation of EGFR-RAD51 Δint24+ with physiologically relevant (auto)phosphorylable tyrosines marked. **C,** NR6 cells stably expressing EGFR-RAD51 or EGFR-RAD51 Δint24+ were subjected to western blot analysis with indicated antibodies. The two distinct EGFR variants were detected at the anticipated molecular weight of 150 and 110 kD. Both constructs can be detected with the N-terminal EGFR antibody, but only the fusion can be detected with the RAD51 antibody. **D,** Alignment of EGFR-RAD51 with high scoring effectors, adaptors, and/or kinases (47-50). Residues surrounding EGFR-RAD51 Y1222 are indicated in the table relative to Y1222 at position 0. The consensus binding sequence of NCK1 (51) and consensus phosphorylation site of several kinases are indicated below (52,53). In the case of multiple amino acids being selected as substrate at a particular position, they are listed in order of relative selectivity (54).

inhibition interrupts signaling and proliferation of cells expressing EGFR-RAD51. Of note, SRC and Y869 have been deemed critical for EGFR-mediated transformation (55).

Still, because EGFR C-term deletions have been reported in lung (56) and other cancers (57) it could be that EGFR-RAD51 is able to transduce signal independent of any phosphotyrosines or other C-term adapter moities. To test this hypothesis, we have created an EGFR construct that spans EGFR exons 1-24, which we call EGFR Δint24+ **(Figures 4.3A–B)**. We have created stable Ba/F3 and NR6 cells using this construct and will test the ability of this EGFR-RAD51 variant to support downstream mitogenic signaling **(Figure 4.3C)**. We are also planning on creating combinations of Δint24+ with tyrosines 869, 915, 955, and 998 substituted with phenylalanines (singly, and in combination). If the variants are unable to support oncogenic signaling and transformation, the data would point towards a key role for the RAD51 portion of EGFR-RAD51 and would bring into question previous findings claiming EGFR Δint24+ is oncogenic (56,57).

As such, an alternative and intriguing hypothesis has to do with the RAD51 portion of the fusion serving as the docking site for downstream adaptors. Although tyrosine phosphorylation of RAD51 is a contentious topic[e], there is mounting evidence that RAD51 is regulated by tyrosine phosphorylation in humans (62). One of RAD51's phosphorylation sites (Y315) is present in EGFR-RAD51 as Y1222 **(Figure 4.3A)**. Stringent computational prediction of cell signaling motif interactions using a variety of approaches (47,48) predicts that NCK1, a known oncogenic adaptor protein (63), can bind to pY1222. Indeed, NCK1's SH2 binding motif YDx(V/p) (51) is found as YDSP in EGFR-RAD51. Similar computational prediction using a variety of approaches (49,50) predicts that Y1222 could be phosphorylated by ABL (which has been experimentally confirmed (62)), SRC family kinases, and/or EGFR—each with decreasing likelihood **(Figure 4.3D)**. Therefore, we hypothesize that EGFR-RAD51 is phosphorylated at Y1222, perhaps as a result of autophosphorylation. To test this hypothesis, we have created a series of mutants: EGFR-RAD51 D837N (which ablates the EGFR kinase activity); Y1222F (which removes RAD51's tyrosine phosphorylation site); and combinations of the two. We will create stable cell lines (Ba/F3 and NR6) with EGFR-RAD51 Y1222F to see whether the mutant is still able to transform cells. If ablating the RAD51 tyrosine phosphorylation site prevents cellular transformation, then we will perform co-expression experiments in CHO cells to test for autophosphorylation. Additionally, since NCK1 is the most likely adaptor, we will perform co-IP experiments with/out Y1222F. Together these studies will should reveal the mechanism by which EGFR-RAD51 is activated and propagates downstream signal.

EGFR-KDD: Ongoing Work and Future Directions

Creating a Cell Line Model of EGFR-KDD

As discussed above with regards to our EGFR-RAD51 work, the main limitation with our current EGFR-KDD work has to do with the cell lines (Ba/F3 and NR6) that we use as our models. To circumvent the limitations of an overexpression/murine model, we attempted to develop a cell line from the index patient whose tumor harbored EGFR-KDD. At the time of acquired resistance to afatinib, the patient underwent a series of endobronchial ultrasound biopsies. After routine pathological assessment, remaining tumor sample was sent to Foundation Medicine (a company which performs targeted NGS on tumor samples (64)) and to our laboratory. In the lab, the ~0.5 mm^3 biopsy was minced with a sterile scalpel. Then, the biopsy was placed in media with feeder cells for conditional cell re-programming (25,27) **(Figure 4.4A)**. Under these conditions, the biopsy expanded and maintained its original glandular architecture for 8 passages. **(Figure 4.4B)** However, the conditionally reprogrammed cells (CRCs) are difficult to use for cellular-based assays (such as the MTT cell viability assay) because of the feeder cells. Therefore, we tried several different methods to wean the CRCs off the feeder cells, including culture on different sized vessels, on different surfaces (plastic, collagen, agar), with different media (such as the lung epithelial cell specific LHC9 (65)), and at different densities. No matter what method we used (including a variety of media and culture surfaces [see **Methods**]), it appeared (via light microscopy) that the cells changed morphology with each of these conditions, never retaining the glandular morphology of the initial biopsy **(Figure 4.4C)**. Moreover, while the cells tested positive for EGFR-KDD during the first passage as CRCs, they seemed to lose expression of the EGFR-KDD with repeated passage and in different conditions **(Figures 4.4D–E)**. While we have retained several pieces of the original biopsy, early passage CRCs, and differentiated cells for further optimization and testing, the index patient has unfortunately passed away, preventing further tissue collection.

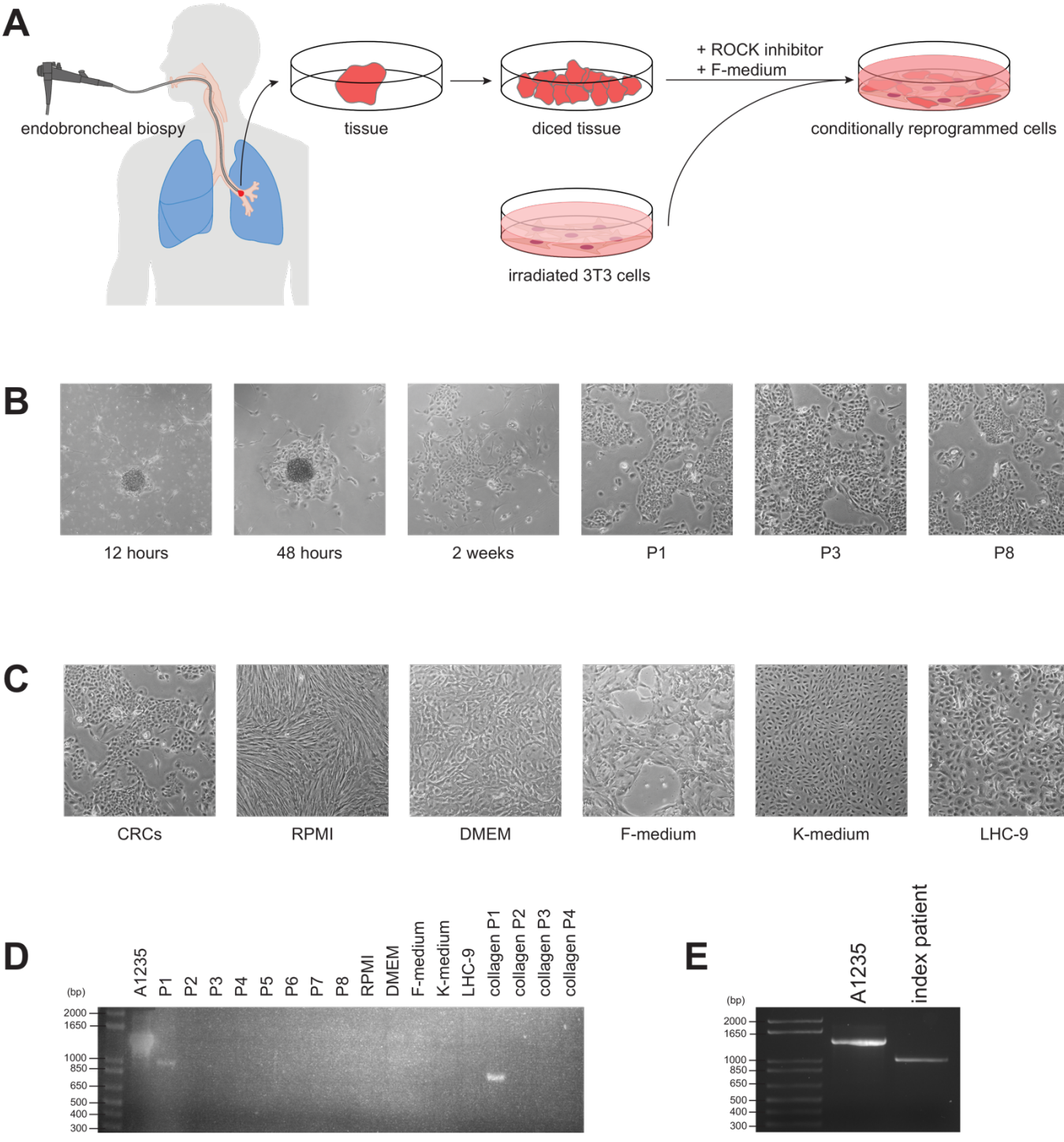


Figure 4.4: Development of a conditionally reprogrammed cell line from the index EGFR-KDD patient.
A, overview of the development of conditionally reprogrammed cells (CRCs) from an endobronchial biopsy of the index patient with a lung adenocarcinoma harboring EGFR-KDD. For details, see methods (27). **B,** 4 x microphotographs of CRCs showing the development from day 1 at 12 hours, day 3 at 48 hours, 2 weeks, and over the course of 8 passages. P, passage. **C,** 4 x microphotographs of parental CRCs and derivatives cultured in various media, at confluence. **D,** PCR amplification of the EGFR-KDD exon 25-18 breakpoint in A1235 cells (glioma cell line harboring EGFR-KDD), at various passages of the CRCs, from the cells in varying media (at P1), and from the cells grown as CRCs on collagen-coated plates. **E,** PCR amplification of the EGFR-KDD exon 25-18 breakpoint in A1235 cells (glioma cell line harboring EGFR-KDD) and in the index patient's primary tumor biopsy.

Maximizing Inhibition of EGFR-KDD

As shown in our published data, erlotinib, afatinib, and/or osimertinib can inhibit the proliferation of cells harboring EGFR-KDD; however, whether these agents are the best way to inhibit the proliferation of cells driven by EGFR-KDD remains an open question. This is an important question, with regards to the clinical treatment of tumors harboring EGFR-KDD and the biology of EGFR-KDD: various activation states of EGFR can best be treated by structurally different EGFR-TKIs (66). Therefore, we ran MTT assays against Ba/F3 cells harboring EGFR-KDD and tested the ability of all available EGFR-TKIs (erlotinib, gefitinib, afatinib, dacomitinib, neratinib, lapatinib, osimertinib, and rociletinib) to inhibit EGFR-KDD proliferation over a 4-log, 20-dose range **(Figure 4.5A)**. The irreversible TKIs afatinib and dacomitinib proved to be the most potent TKIs against EGFR-KDD, even when adjusting for clinically attainable plasma concentrations (67-74) **(Figure 4.5B)**. The anti-EGFR antibody cetuximab was also tested against EGFR-KDD and EGFR-L858R using MTT assays and proved to have little to no effect **(data not shown)**[f]. The clinically useful combination of cetuximab and afatinib (77,78) was also tested in soft agar assays and proved to be no more potent than either agent alone against the EGFR-KDD **(Figures 4.5C–D)**.

Several therapeutic avenues remain to be explored and could prove effective against the EGFR-KDD. Agents targeting MAPK (such as the MEK inhibitors trametinib and selumetinib) and agents targeting the PI3K/AKT pathway (such as MK-2206, dactolisib, and/or pictilisib) alone and in combination (dual vertical inhibition) with EGFR-TKIs have yet to be tested against the EGFR-KDD. We also plan on treating our cell line models with other agents that target EGFR signaling: anti-EGFR therapeutic antibodies with different targeting epitopes (panitumumab, Sym004, and necitumumab), EGFR-TKIs that may sample different activation states of the kinase (canertinib, icotinib, and vandetanib), and, because mutant EGFR is reliant upon the chaperone protein HSP90 for proper folding (79), HSP90 inhibitors (17AAG, ganetespib, and luminespib).

Acquired Resistance to EGFR-TKIs in the Setting of EGFR-KDD

Acquired resistance to afatinib in a patient whose tumor harbored the EGFR-KDD was described in our initial publication (80). For this patient, the putative mechanism of resistance involved genomic amplification of the whole EGFR-KDD allele, without any additional mutations (such as T790M). Since our initial publication, a report of acquired resistance in another patient with EGFR-KDD has been published (81). This patient was treated with EGFR-TKIs (erlotinib and gefitinib) on and off for 10 years before developing EGFR-KDD 'T790M'. However, this publication did not differentiate which of the EGFR-KDD gatekeeper residues was mutated (T790M or T1141M)[g]. It is important to understand where the TM mutation is located as it will information EGFR biology and, potentially, future treatment.

In order to model resistance is the setting of EGFR-KDD we initially planned on developing isogenic pairs of EGFR-TKI sensitive and resistant A1235 cell lines (which harbor endogenous EGFR-KDD). However, because these cell lines are glioblastoma lines, and because of the biological differences between brain and lung epithelia, this approach was abandoned. Instead, we focused on performing an ENU mutagenesis screen on Ba/F3 cell lines, which is a validated means of modeling resistance to TKIs (82,83). BA/F3 cells (EGFR-L858R and -KDD) were exposed to 50 μg/mL N-ENU for 48 hours, expanded, and grown in the presence of drug. EGFR-L858R served as a positive control in this experiment. The drugs used (based on clinically achievable doses (67,70,74)) were erlotinib (at 500, 1000, 2000 nM), afatinib (at 10, 100, 1000 nM), and osimertinib (at 200, 500, 1000 nM). The experiment was repeated twice (N=2), and, each time ≥ 72 resistant clones of EGFR-L858R were identified—at all drug / dose ranges. However, both times, EGFR-KDD failed to develop any resistant clones—even after 1 month of culture. We hypothesize that the lack of emergent resistance, in this assay, in EGFR-KDD cells may have to do with the fact that ENU works by creating points mutations / indels and not large genomic rearrangements; it is unlikely that ENU can induce amplification of the whole EGFR-KDD allele as was seen in the initial patient resistant to afatinib. To test whether the clinically observed T790M mutation can mediate resistance when located in EGFR-KDD, we have created Ba/F3 cells harboring EGFR-KDD-T790M and EGFR-KDD-T1141M. While we hypothesize that T1141M would mediate resistance (in the acceptor positon of the asymmetric dimmer (84)), both will be tested against EGFR-TKIs and compared to EGFR-KDD. In parallel, EGFR-T790M will be compared with EGFR-L858R. These findings should hint at the mechanism of activation of EGFR-KDD, especially if EGFR-KDD-T1141M mediates EGFR-TKI resistance.

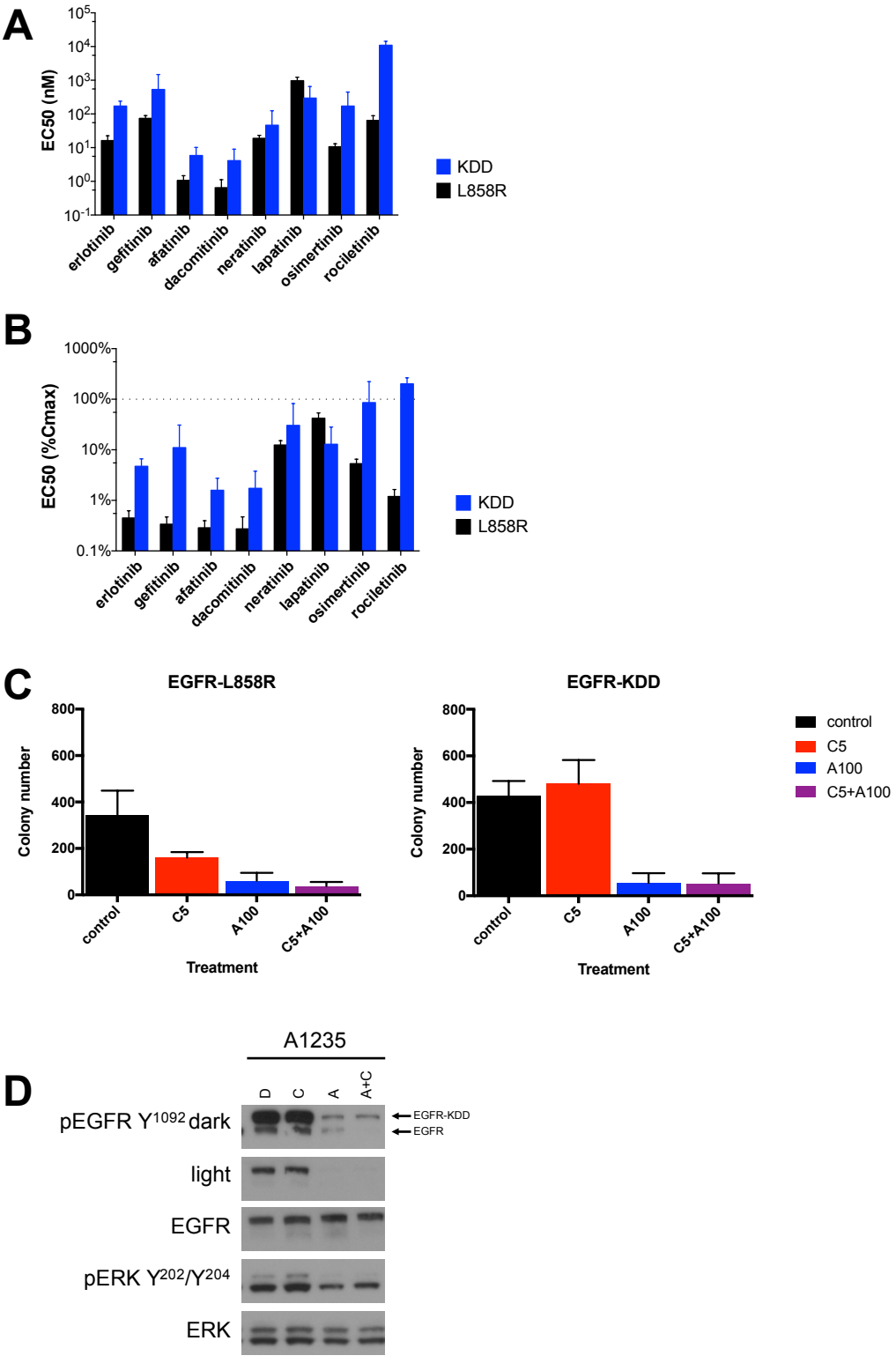


Figure 4.5: Maximizing inhibition of the EGFR-KDD
A, Ba/F3 cell lines stably expressing EGFR-L858R (black) or EGFR-KDD (blue) were treated with increasing doses of erlotinib, gefitinib, afatinib, dacomitinib, neratinib, lapatinib, osimertinib, and rociletinib for 72 hours. CellTiter Blue assays were performed to assess viability at endpoint, and an EC50 derived from curve fitting in Prism. Each bar represents the average result from triplicate experiments ± SD. **B,** EC50 results from Figure 4.5A adjusted for clinically attainable plasma concentrations (67-74). **C,** NR6 cells stably expressing EGFR-L858R or EGFR-KDD were grown in soft agar for three weeks in presence of the indicated drugs and quantified for colony formation . C5 = 5 µg/mL cetuximab; A100 = 100 nM afatinib; C5+A100 = 5 µg/mL cetuximab + 100 nM afatinib. **D,** A1235 cells (glioma cell line harboring EGFR-KDD) were treated with cetuximab, afatinib, or the combination, lysed, and analyzes for western blot analysis with the indicated antibodies. C = 5 µg/mL cetuximab; A = 100 nM afatinib; A+C = 5 µg/mL cetuximab + 100 nM afatinib.

Adding Experimental Support to the Intra-Molecular Asymmetric Dimer Model

As discussed in the introductory chapter, the activation of EGFR involves the formation of an asymmetric dimer between two EGFR TKDs, in which one TKD serves as an allosteric activator of the other (37). This mechanism of activation for EGFR does not require activation loop phosphorylation by a kinase (85); instead, this allosteric activation involves the C-lobe of one EGFR TKD (the first kinase of 'activator') forming a tight interface with the N-lobe of another (the second kinase or 'receiver'), which stabilizes the second kinase / receiver in the active conformation **(Figure 1.4B and Figure 4.6A)**. Given the presence of two tandem kinase domains within the EGFR-KDD, we hypothesized that EGFR-KDD could form an intra-molecular asymmetric dimer. To test this hypothesis, we modeled the EGFR-KDD based on the available experimental structure of the active asymmetric EGFR dimer. Conformational loop sampling with Rosetta demonstrated that the linker between the tandem tyrosine kinase domains allows for the proper positioning of the two domains necessary for asymmetric dimerization and intra-molecular activation **(Figure 3.1F and Figure 4.6B)** (80). Although this structural modeling demonstrates that the EGFR-KDD is geometrically capable of forming active intra-molecular dimers, further experimental data are needed to confirm this mechanism.

To provide experimental support for our structural model of EGFR-KDD, we created mutant variants of the protein which would be unable to propagate downstream signal if the EGFR-KDD was forming an intra-molecular dimer **(Figures 4.7A–B)**. Based on our model and on the documented existence of higher ordered EGFR oligomers (86,87), we hypothesize that EGFR-KDD can form intra- and inter- molecular dimers **(Figure 4.8A)**. Consistent with this hypothesis, mutating the contact residues at the C-lobe within the first kinase (V948R) or at the N-lobe within the second kinase (I1057Q) should ablate EGFR-KDD intra-molecular dimerization and auto-activation **(Figure 4.8B)**. Similarly, since the second (receiver) kinase is active in an intra-molecular dimer, mutating the catalytic aspartate of the second TKD (D1188N) should impair EGFR-KDD enzymatic activity **(Figure 4.8B).** Mutations that do not interfere with the intra-molecular dimerization interface—I706Q (N-lobe interface disruption of the first kinase) and V1299R (C-lobe interface disruption of the second kinase)—should allow for constitute activation of the EGFR-KDD **(Figure 4.8C)**. Moreover, ablating the activity of the first kinase (D837N) should have no effect on EGFR-KDD auto-activation and intra-molecular activation as this first kinase only serves to activate the second by physical contact and not phosphorylation **(Figure 4.8C)**. Finally, any mutation that ablates intra-molecular dimerization or auto-activation should be rescued by the addition of EGF—which should allow for inter-molecular dimerization by creating chains of active kinase domains. These mutations and hypotheses are summarized in **Figures 4.8**.

In order to create these EGFR-KDD mutants, we first constructed a new EGFR-KDD cDNA in which we used the degeneracy of the genetic code to create two distinct kinase domains, at the nucleotide level, that encoded for the same amino acid sequence **(Figure S4.1)**. With two differently-coded kinase domains, we were able to use site-directed mutagenesis to insert the previously mentioned amino acid mutations at the N-lobe, C-lobe, and catalytic interfaces. Had we used our previously published EGFR-KDD cDNA, any attempt to insert a single amino acid mutation with site-directed mutagenesis would have forcibly inserted two such mutations (because of the identically-coded tandem kinase domains). We successfully created all the aforementioned EGFR-KDD mutants and confirmed all variants by bi-directional Sanger sequencing. Unfortunately, despite repeated attempts, we could not reproducibly re-create the western blot from **Figure 4.8** or create reliable EGFR-KDD cell line models. In fact, we could barely produce any EGFR-KDD (and variant) protein—in any system we tested (**Figures S4.2A–C**).

A series of troubleshooting experiments led us to conclude that the low level of EGFR-KDD expression had to do with the new, degenerate, EGFR-KDD constructs (with degenerate kinase domains). Specifically, in creating the two degenerate kinase domains, we produced an EGFR-KDD cDNA that could not express at levels suitable for cellular transformation likely due to several translation-related factors (88). For one, because the degenerate kinase domain was encoded using rarer codons, we had a transcript that had a poor codon adaptation index (88,89) **(Figure S4.2D)**. Simultaneously, these rare codons were, at times, in tandem, which can predispose to ribosomal frameshift and improper translation (88,90) **(Figure S4.2E)**. We therefore devised a new cloning and mutational scheme to insert mutations in the EGFR-KDD constructs (two wild-type EGFR kinase domains) that we had previously published (80). Using multi-site directed mutagenesis, a repeating restriction site, and a series of sub-cloning reactions, we were able to generate all EGFR-KDD constructs in a non-degenerate construct (see **methods and Figure S4.3A**). We proceeded to create new retroviral EGFR-KDD plasmids and stable Ba/F3 cells harboring the variants **(Figure S4.3B)**.

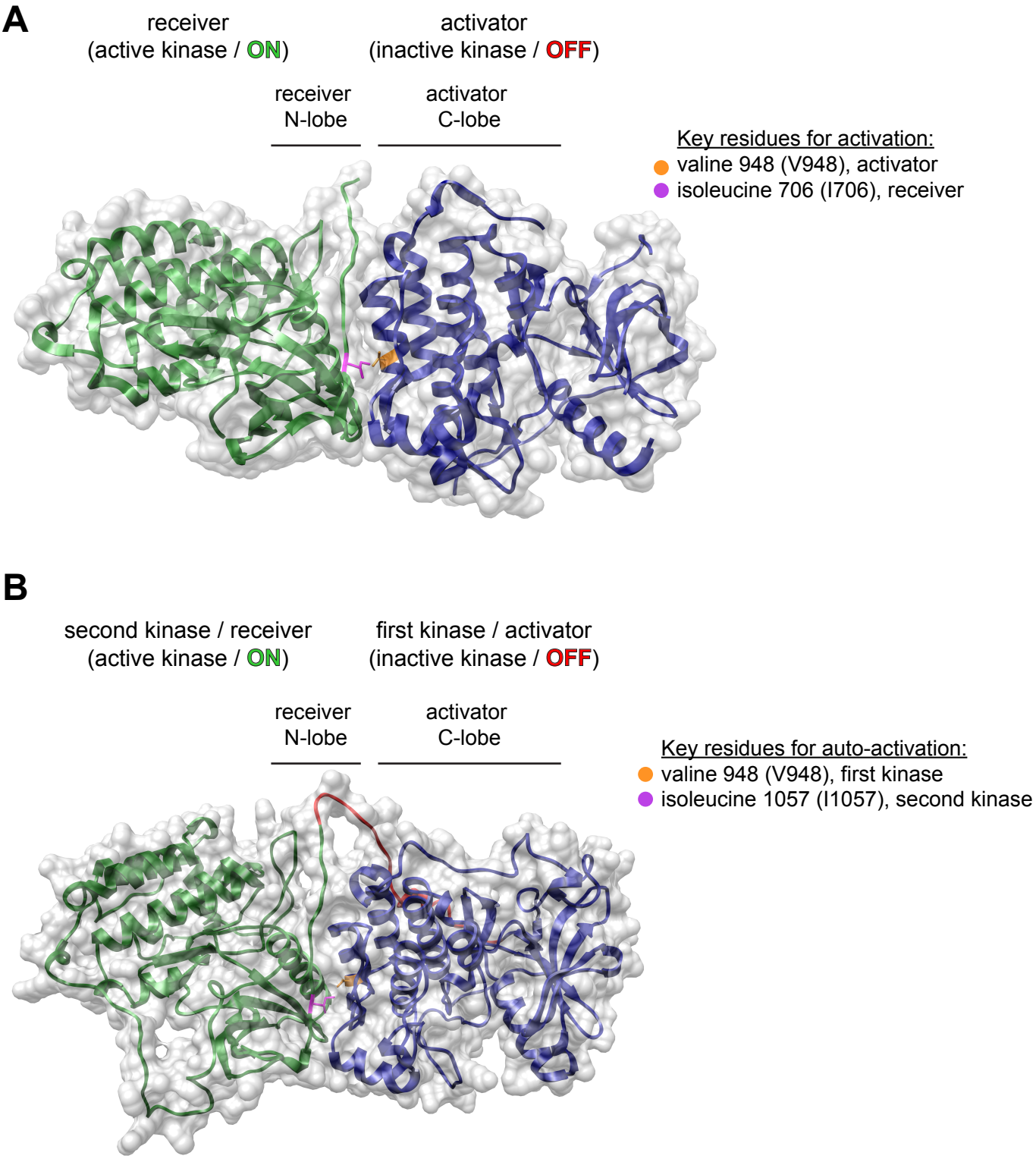
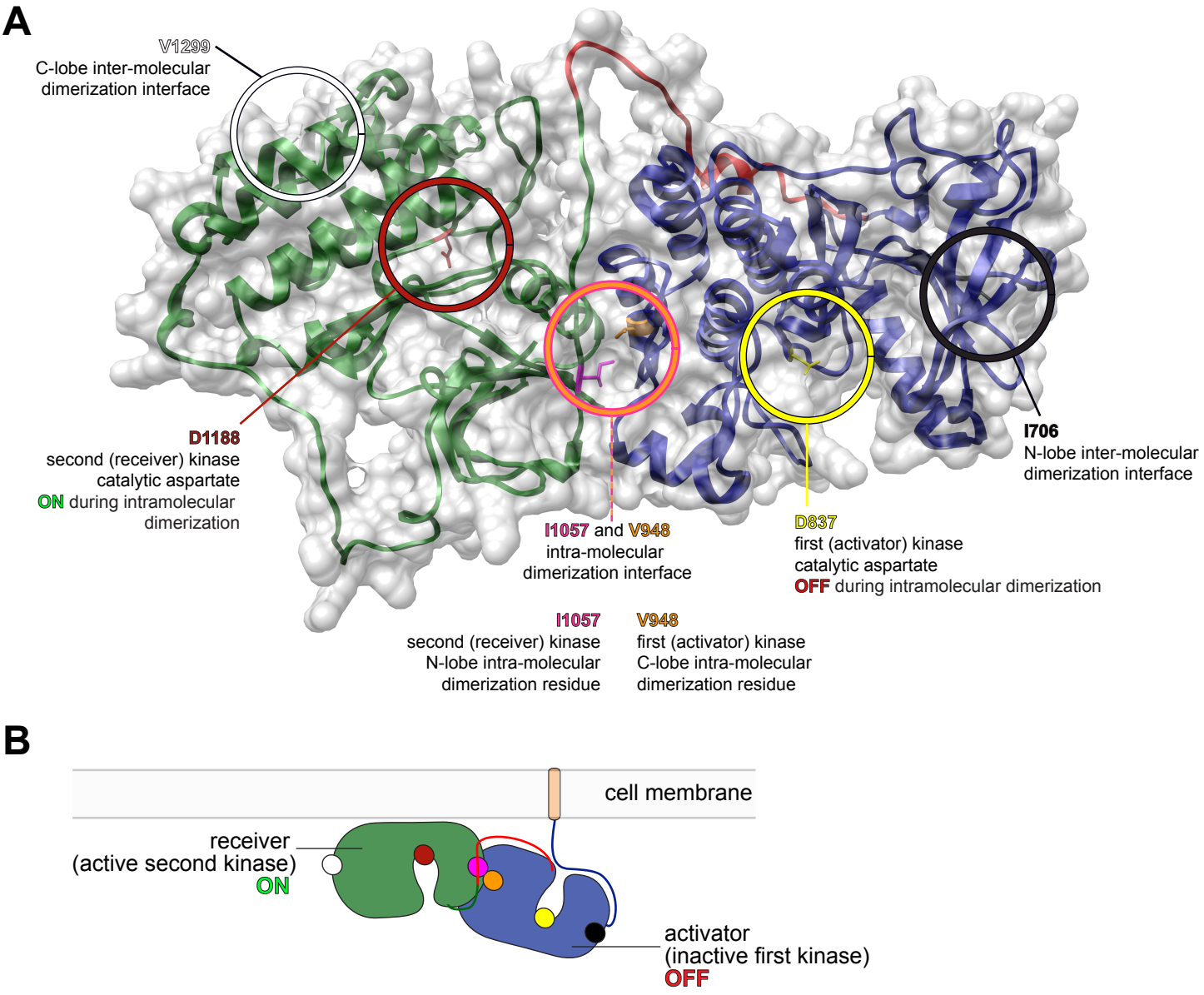


Figure 4.6: Comparison of WT-EGFR activation and putative EGFR-KDD auto-activation.
A, also **Figure 1.4B**. Ribbon diagram and space filling model of the asymmetric dimer of the EGFR TKD (PDB 2GS6 (37)) generated using Chimera (91). The activator kinase is colored navy, and the receiver (enzymatically active) kinase is colored dark green. Key contact residues at the asymmetric interface are highlighted in orange (valine 948) and magenta (isoleucine 706). **B,** ribbon diagram and space-filling model of the EGFR-KDD kinase domains illustrating the proposed mechanism of autoactivation (80), generated using Chimera (91). The first kinase, or activator kinase, is colored navy, and the second kinase, or receiver (enzymatically active) kinase, is colored dark green. The linker is colored red. Key contact residues at the asymmetric (intra-molecular) interface are highlighted in orange (valine 948) and magenta (isoleucine 1057).



	residue	location	function	
●	I706	N-lobe, first kinase	inter-molecular activation (receives)	●
●	D837	catalytic site, first kinase	inter-molecular ATP transfer	●
●	V948	C-lobe, first kinase	intra-molecular activation (activates)	●
●	I1057	N-lobe, second kinase	intra-molecular activation (receives)	●
●	D1188	catalytic site, second kinase	intra-molecular ATP transfer	●
○	V1299	C-lobe, second kinase	inter-molecular activation (activates)	○

Figure 4.7: Key residues for EGFR-KDD intra- and inter- molecular dimerization and function
A, ribbon diagram and space-filling model of the EGFR-KDD kinase domains illustrating the proposed mechanism of autoactivation (80), generated using Chimera (91). The first kinase, or activator kinase, is colored navy, and the second kinase, or receiver (enzymatically active) kinase, is colored dark green. The linker is colored red. Key residues are displayed, colored, and highlighted: I706 (isoleucine 706, black); D837 (aspartate 837, yellow); V948 (valine 948, orange); I1057 (isoleucine 1057, magenta); D1188 (aspartate 1188, dark red); V1299 (valine 1299, white). **B,** model of EGFR-KDD auto-activation; adapted from (92). The first kinase, or activator kinase, is colored navy, and the second kinase, or receiver (enzymatically active) kinase, is colored dark green. The linker is colored red. Key residues are pictured as circles: I706 (isoleucine 706, black); D837 (aspartate 837, yellow); V948 (valine 948, orange); I1057 (isoleucine 1057, magenta); D1188 (aspartate 1188, dark red); V1299 (valine 1299, white).

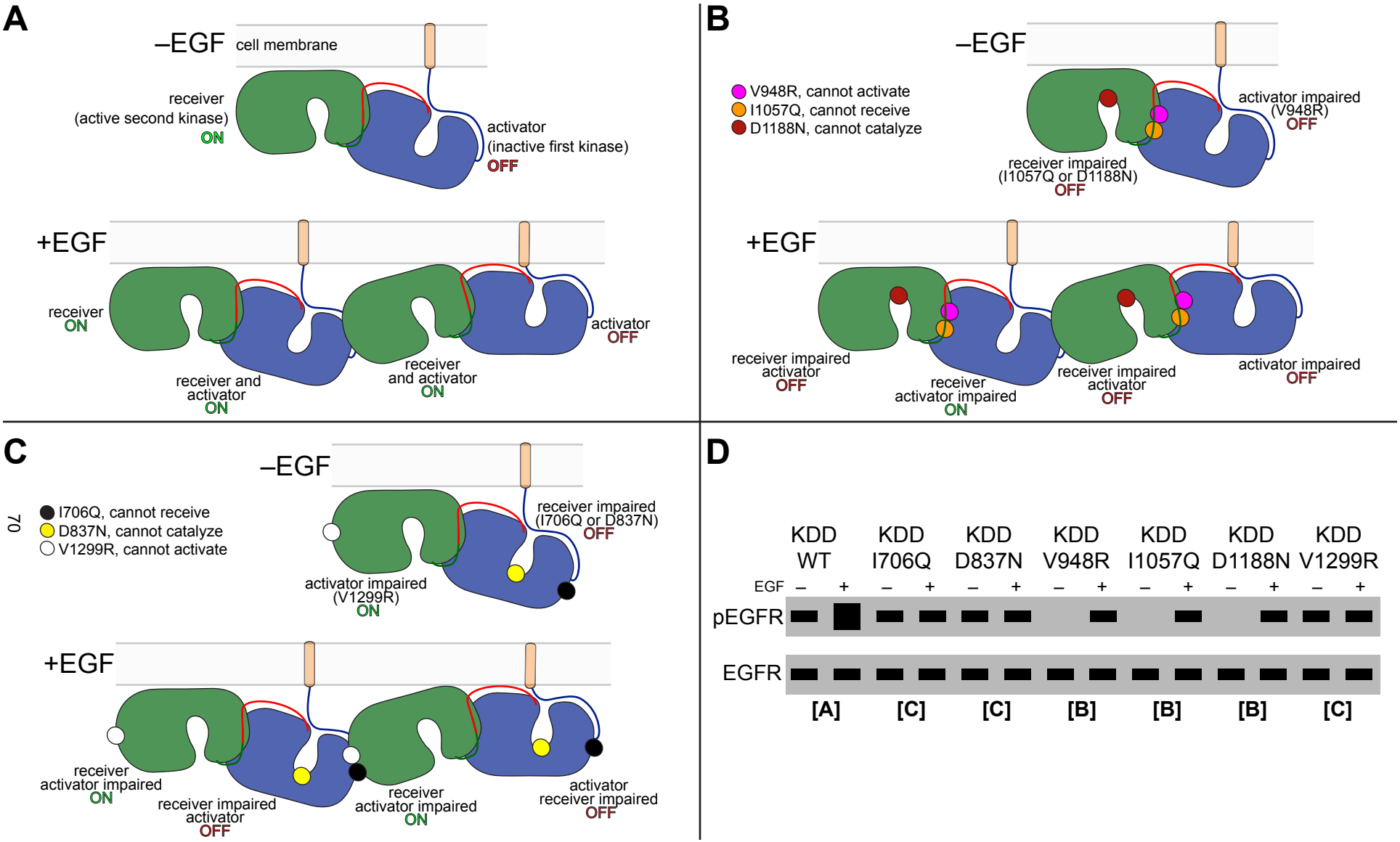


Figure 4.8: Expected effects of EGFR-KDD dimerization-inhibiting point mutations
A, expected effect of EGF withdrawal (top) and addition (bottom) on EGFR-KDD. Removing EGF should have no effect on EGFR-KDD kinase activity; adding EGF promotes EGFR-KDD multimerization and mass activation. **B,** expected effect of intra-molecular dimerization inhibiting (V948R or I1057Q) and receiver kinase dead (D1188N) mutations. Any of these mutations, alone or in combination, would ablate intra-molecular auto-activation in the absence of EGF (top), but this would be rescued with addition of EGF (bottom). **C,** expected effect of inter-molecular dimerization inhibiting (I706Q or V1299R) and activator kinase dead (D837N) mutations. Any of these mutations, alone or in combination, would ablate inter-molecular activation. The molecule could still auto-activate in the absence of EGF (top), but this would activity would not be greater with the addition of EGF (bottom). **D,** expected western blot if individual EGFR-KDD mutants were to be expressed in EGFR-null cells ± EGF and probed with the indicated antibodies. Brackets indicate the scenarios from the mutants in the panels of this figure.

We will now use these cell line models (and others) in a series of experiments to determine how the variants effect activation of the EGFR-KDD protein. Specifically, we will test whether the EGFR-KDD variants can overcome IL-3 dependence of Ba/F3 cells and anchorage-dependence of NR6 cells and YAMC EGFR$^{-/-}$ cells—EGFR null cells that contain a heat labile large T antigen (93). Additionally, we will repeat transient transfection westerns and perform drug sensitivity experiments to gain insight into the structural mechanism of activation of the EGFR-KDD. We are confident that the degenerate KDD construct could not express well due to several contextual translation issues **(Figure S4.2)** and believe that our new system will allow us to tease out the mechanism of activation of the EGFR-KDD—providing experimental support for the structural model we proposed and a new way to study EGFR biology.

Methods

Cell Culture

A1235 cells were a kind gift from Drs. Fenstermaker and Ciesielski (Roswell Park Cancer Institute, Buffalo, NY) (94). 293F/T cells were purchased from ATCC. Ba/F3 cells were purchased from DSMZ. Plat-GP cells were purchased from CellBioLabs. NR6 cells were a kind gift from Dr. William Pao (Roche, CH) (77). Ba/F3 cells were maintained in RPMI 1640 medium (Mediatech). YAMC EGFR$^{-/-}$ cells were a kind gift from Dr. Robert Whitehead (Vanderbilt University, TN) (95). A1235, 293F/T, and NR6 cells were maintained in DMEM (Gibco). Media was supplemented with 10% heat inactivated fetal bovine serum (Atlanta Biologicals) and penicillin-streptomycin (Mediatech) to final concentrations of 100 U/ml and 100 µg/ml, respectively. The Ba/F3 cell line was supplemented with 1 ng/mL murine IL-3 (Gibco). The Plat-GP cell line was cultured in the presence of 1 µg/mL blasticidin (Gibco). YAMC EGFR$^{-/-}$ cells were cultured in DMEM containing 5 % FBS, 1 µg/mL insulin (Sigma), 10 µM α-thioglycerol Sigma), 1 µM (or ~0.5 µg/mL) hydrocortisone (Sigma), and 5 U/mL murine γ-interferon (Peprotech). Due to their heat-labile large T antigen, YAMC EGFR$^{-/-}$ cells were cultured at 33 ºC. All other cell lines were maintained in a humidified incubator with 5% CO2 at 37°C and routinely evaluated for mycoplasma contamination.

gRNA Target Selection and Cloning

Suitable gRNAs were created using the online CRISPR Design Tool (96) with Streptococcus pyogenes Cas9 (SpCas9) as the chosen nuclease and with the most recent human genome reference sequence (GRCh37/hg19). Oligos were synthesized (Eurofins MWG Operon), annealed, and ligated into the BbsI sites of pSpCas9(BB)-2A-GFP (PX458, Addgene) (20). gRNA sequences are listed in **Table S4.1**. EGFR and RAD51 gRNAs were multiplexed into PX458 using the PrecisionX Multiplex gRNA Cloning Kit (SBI), an H1 promoter amplicon (97), and overlapping PCR primers **(Table S4.1)**. The resulting PX458 constructs (EGFR-RAD51#1 and #2 [each having the corresponding EGFR and RAD51 gRNAs]) have an EGFR gRNA driven by a H1 promoter and a RAD51 gRNA driven by a U6 promoter. The multiplexed gRNAs were also subcloned from PX458 into the BsmBI sites of plentiCRISPRv2 (98).

gRNA Testing in 293FT Cells

293FT cells (Thermo) were plated the day before transfection at 1×10^6 cells per well of a six-well plate. On the day of transfection, the medium was replaced with 2 mL of antibiotic-free DMEM (Gibco) with 10% heat inactivated fetal bovine serum (FBS, Atlanta Biologicals). Cells were transfected with 3 µg of PX458 and a 2:1 ratio (µL L2K: µg DNA) of Lipofectamine 2000 (Life Technologies) via 500 mL of OPTI-MEM (Life). 48 hours later, transfection efficiencies were estimated by detection of the green fluorescent protein (GFP) reporter; transfection efficiencies averaged > 90% **(data not shown)**. After visualization of GFP, cells were manually detached from plates, washed with PBS, snap frozen, and their DNA extracted with the DNEasy kit (Qiagen). sgRNA functional testing was then performed with the SURVEYOR nuclease assay (99) per the manufacturer's instructions (IDT). 200 ng of DNA was amplified by HotStarTaq Mastermix PCR (Qiagen) using primers that created 800 bp amplicons surrounding the sgRNA-targeted sites **(Table S4.1)**. SURVEYOR products were the run on 2% TBE-agarose gels and visually inspected for banding **(Figures 4.1B–C)**.

Transfecting and Testing AALE Cells with Multiplexed EGFR-RAD51 Constructs

5×10^6 AALE cells (a kind gift of Dr. Meyerson, DFCI) were plated in a 100 mm dish the day before transfection. On the day of transfection, the medium was replaced with 6 mL of antibiotic-free DMEM with 10% heat inactivated fetal bovine serum. Cells were transfected with 10 µg of PX458 and a 2:1 ratio (µL L2K: µg DNA) of Lipofectamine 2000 via 500 mL of OPTI-MEM. Media was replaced after 24 hours and cells split into

4×96-well plates by limiting dilution 72 hours after transfection. Single clones were marked after 24 hours, transferred to 12-well plates after two weeks, and then expanded into 100 mm dishes. After reaching confluence, cells were trypsinized and split 1:1 into pellets for DNA extraction or into supplemented DMSO media for long term storage.

To test for the presence of EGFR-RAD51 fusions, DNA was extracted from ½ of the cell pellets with the DNEasy kit and RT-PCR performed with SuperScript III One-Step RT-PCR System with Platinum Taq (Life). RT-PCR primers are listed in **Table S4.1**. Ba/F3 cells stably expressing EGFR-RAD51 were used as a positive control (18). Samples that had signal nearly equivalent signal as compared to the positive control subsequently had protein extracted from the other ½ of the cell pellets and were subjected to western blot analysis. Cells were lysed in RIPA buffer (150 mM NaCl, 1% Triton-X-100, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris·HCl, pH 8.0) with freshly added 40 mM NaF, 1 mM Na-orthovanadate, and protease inhibitor mini tablets (Thermo). Protein was quantified using protein assay reagent and a SmartSpec plus spectrophotometer (Bio-Rad) per the manufacturer's protocol. Lysates were subjected to SDS-PAGE followed by blotting with the indicated antibodies (C-term RAD51 clone C20 #6862, 1:1000 dilution from Santa-Cruz and actin #A2066 1:5000 dilution from Sigma) and detection by Western Lightning ECL reagent (Perkin Elmer).

Creation of EGFR-RAD51 Dimerization-Impaired Mutants

Mutations were introduced into pMa-EGFR-RAD51 (18) with the QuickChange II kit (Stratagene). Site-directed mutagenesis was carried out per the manufacturer's recommendations with the exception of extension time being set at 1.5 mins / kb. Mutagenesis primers are listed in **Table S4.2**. After bi-directional dideoxy sequencing, EGFR-RAD51 mutants were subcloned into the BglII/EcoRI sites of pMSCV-puro (Clontech) or into the BamHI/EcoRI sites of pcDNA3.1 (Life). Plasmids were then re-sequenced to ensure proper insertion and that no additional mutations were introduced.

Creation of Tagged EGFR-RAD51 Constructs

New stop codons with restriction sites were introduced into pMa-EGFR-RAD51 (18) with the QuickChange II kit (Stratagene). Site-directed mutagenesis was carried out per the manufacturer's recommendations with the exception of extension time being set at 1.5 mins / kb. Mutagenesis primers are listed in **Table S4.2**. After bi-directional dideoxy sequencing, EGFR-RAD51 mutants were subcloned into the HindIII/EcoRI sites of pTracer-CMV2ΔGFP-3xFLAG (Clontech) or into the HindIII/EcoRI sites of pcDNA6/V5His-B. Plasmids— pTracer-EGFRRAD51-3XFLAG and pcDNA-EGFRRAD51-V5/His—were then re-sequenced to ensure proper insertion and that no additional mutations were introduced.

Derivation of the VP-34 Cell Line

VP34 was derived from the index patient with a lung adenocarcinoma harboring EGFR-KDD as a conditionally reprogrammed cell line, per protocol (27). Briefly, fresh endobronchial biopsy tissue was taken from the patient to the lab and mechanically diced/separated. This tissue was laid scrapped into a 100 mm dish containing F-medium and 5×10^4 lethally irradiated (3000 rad via Cesium irradiation) 3T3 cells. Cells were thereafter maintain like normal cell with the difference being that every passage was onto a new 100 mm dish with freshly irradiated 3T3 cells.

F-medium: 370 mL of DMEM, 125 mL of F12 nutrient mix, 0.5 mL of hydrocortisone/EGF mix (25 µg/ml hydrocortisone in EtOH; 0.125 µg/ml EGF), 0.5 mL of 10 µg/mL insulin, 0.5 mL of amphotericin B (Gibco), 5 mL pf P/S, and 4.3 µL of cholera toxin (Sigma). After adding the ROCK inhibitor Y-27632 (Enzo) to a final concentration of 10 µM, the medium was kept at 4 °C for up to 2 months.

Creation of Stable Cell Lines Harboring EGFR-RAD51 and EGFR-KDD Mutants

Empty pMSCV-puro retroviral vector or pMSCV-puro vectors encoding EGFR-KDD or RAD51 variants were transfected, along with the envelope plasmid pCMV-VSV-G (CellBioLabs), into cells Plat-GP packaging cells (CellBioLabs). Viral media was harvested 48 hours after transfection, spun down to remove debris, and supplemented with 2 µg/mL polybrene (Santa Cruz). 2.5×10^6 Ba/F3 cells, 1×10^6 NR6 cells, or 1×10^6 YAMC $EGFR^{-/-}$ cells were re-suspended in 10 mL viral media. Transduced cells were selected for 1 week in 2 µg/mL puromycin (Invitrogen) and Ba/F3 cells were selected for an additional week in the absence of IL-3. Stable polyclonal populations were used for experiments and routinely tested for expression of EGFR constructs.

Soft Agar and MTT Assays

Cell viability assays, including soft agar and CellTiter Blue assays were performed as previously described (18,80).

ENU Mutagenesis Resistance Screen

ENU mutagenesis was performed per established protocols (82,83). Briefly, ENU (Sigma) was dissolved in DMSO at 50 mg/mL and stored in aliquots at 80°C. ENU was added to Ba/F3 EGFR-L858R or EGFR-KDD cells (5×10^6 cells) to a concentration of 50 µg/mL, followed by culture for 12 to 24 hours. This dose of ENU has been established as minimally cytotoxic (82). The cells were then washed 3 times, re-plated in complete medium, and allowed to expand over 1 week under exponential growth conditions.

ENU-exposed Ba/F3 cells were the plated and cultured in 96-well plates at 1×10^5 cells/well in 200 µL complete media supplemented with graded concentrations of the respective inhibitors. Inhibitor concentrations were as follows: erlotinib (at 500, 1000, 2000 nM), afatinib (at 10, 100, 1000 nM), and osimertinib (at 200, 500, 1000 nM). Non-mutagenized Ba/F3 EGFR-L858R cells, incubated with the maximum concentration of each drug, were included as controls. Wells were observed for cell growth by visual inspection under an inverted microscope and media color change every 2 to 3 days for at least 4 weeks. When growth in a well occurred, cells were transferred to 24-well plates and expanded in the presence of the corresponding inhibitor concentration used in the screen. If growth was simultaneously observed in all 96 wells of a given condition, only 24 wells were expanded for further analysis. Cells were enventually expanded to 100 mm dishes and frozen down.

Creation of EGFR-KDD Sequences, Mutations, and Constructs

The degenerate EGFR-KDD sequence was designed by replacing the first repeat of exons 18–25 with a degenerate sequence. This degenerate sequence was created by copying exons 18–25 of EGFR (CCDS5514.1) into GeneDesign (100) and using the most different RCSU codon juggling algorithm. The complete cDNA **(Figure S4.1)** was synthesized by Life Technologies.

Mutations were introduced into pMa-EGFR-KDD (degenerate) with the QuickChange II kit (Stratagene). Site-directed mutagenesis was carried out per the manufacturer's recommendations with the exception of extension time being set at 1.5 mins / kb. Mutagenesis primers are listed in **Table S4.3**. After bi-directional dideoxy sequencing, EGFR-KDD (degenerate) mutants were subcloned into the BglII/EcoRI sites of pMSCV-puro (Clontech) or into the BamHI/EcoRI sites of pcDNA3.1 (Life). Plasmids were then re-sequenced to ensure proper insertion and that no additional mutations were introduced.

Mutations were introduced into pMa-EGFR-KDD (tandem identical sequences) (80)with the Multi-Side Direction Mutagenesis kit (Stratagene). Site-directed mutagenesis was carried out per the manufacturer's recommendations with the exception of extension time being set at 1.5 mins / kb. Mutagenesis primers are listed in **Table S4.3**. After bi-directional dideoxy sequencing, pMa-EGFR-KDD (tandem) mutants were digested with ClaI and re-combined per **Figure S4.3** to create all 10 mutants. Plasmids were then re-sequenced to ensure proper insertion and that no additional mutations were introduced. pMA-EGFR-KDD (tandem) single mutants were then blunt ligated into the HpaI site of pMSCV-puro (Clontech). Plasmids were finally re-sequenced to ensure proper insertion and that no additional mutations were introduced.

References

1. Palacios, R. & Steinmetz, M. Il-3-dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes in vivo. *Cell* **41,** 727–734 (1985).

2. Daley, G. Q. & Baltimore, D. Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 9312–9316 (1988).

3. Warmuth, M., Kim, S., Gu, X.-J., Xia, G. & Adrián, F. Ba/F3 cells and their use in kinase drug discovery. *Curr Opin Oncol* **19,** 55–60 (2007).

4. Cifone, M. A. & Fidler, I. J. Correlation of patterns of anchorage-independent growth with in vivo behavior of cells from a murine fibrosarcoma. *Proc. Natl. Acad. Sci. U.S.A.* **77,** 1039–1043 (1980).

5. Holliday, R. Neoplastic transformation: the contrasting stability of human and mouse cells. *Cancer Surv.* **28,** 103–115 (1996).

6. Weinberg, R. A. *The biology of cancer*. (2014).

7. Hawley, R. G., Lieu, F. H., Fong, A. Z. & Hawley, T. S. Versatile retroviral vectors for potential use in gene therapy. *Gene Ther.* **1,** 136–138 (1994).

8. Miller, A. D. & Rosman, G. J. Improved retroviral vectors for gene transfer and expression. *Biotech.* **7,** 980–2– 984–6– 989–90 (1989).

9. Frenkel-Morgenstern, M. *et al.* Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. *Genome Research* **22,** 1231–1242 (2012).

10. Kim, T. M. *et al.* RAD51 Mutants Cause Replication Defects and Chromosomal Instability. *Mol. Cell. Biol.* **32,** 3663–3680 (2012).

11. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–307 (2013).

12. Shaver, T. M. *et al.* Diverse, Biologically Relevant, and Targetable Gene Rearrangements in Triple-Negative Breast Cancer and Other Malignancies. *Cancer Research* **76,** 4850–4860 (2016).

13. Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research* **41,** D955–61 (2013).

14. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* 1–44 (2016).

15. Gazdar, A. F., Girard, L., Lockwood, W. W., Lam, W. L. & Minna, J. D. Lung Cancer Cell Lines as Tools for Biomedical Discovery and Research. *JNCI Journal of the National Cancer Institute* **102,** 1310–1321 (2010).

16. Desai, T. J., Brownfield, D. G. & Krasnow, M. A. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* **507,** 190–194 (2014).

17. Latysheva, N. S. & Babu, M. M. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Research* **44,** 4487–4503 (2016).

18. Konduri, K. *et al.* EGFR Fusions as Novel Therapeutic Targets in Lung Cancer. *Cancer Discovery* **6,** 601–611 (2016).

19. Choi, P. S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nature Communications* **5,** 3728 (2014).

20. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8,** 2281–2308 (2013).

21. Maddalo, D. *et al.* In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516,** 423–427 (2015).

22. Lundberg, A. S. *et al.* Immortalization and transformation of primary human airway epithelial cells by gene transfer. *Oncogene* **21,** 4577–4586 (2002).

23. Lovly, C. M. *et al.* Rationale for co-targeting IGF-1R and ALK in ALK fusion-positive lung cancer. *Nat Med* **20,** 1027–1034 (2014).

24. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 4283–4288 (2008).

25. Liu, X. *et al.* ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am. J. Pathol.* **180,** 599–607 (2012).

26. Crystal, A. S. *et al.* Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science* **346,** 1480–1486 (2014).

27. Liu, X. *et al.* Conditional reprogramming and long-term expansion of normal and tumor cells from human biospecimens. *Nature Protocols* **12,** 439–451 (2017).

28. Walsh, A. J., Cook, R. S., Sanders, M. E., Arteaga, C. L. & Skala, M. C. Drug response in organoids generated from frozen primary tumor tissues. *Sci. Rep.* 1–11 (2015).

29. Ramirez, R. D. *et al.* Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer Research* **64,** 9027–9034 (2004).

30. Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Research* **40,** 5795–5818 (2012).

31. Vouillot, L., Thélie, A. & Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes*| *Genomes*| *Genetics* (2015).

32. Carvalho, J. O. F. & Kanaar, R. Targeting homologous recombination-mediated DNA repair in cancer. *Expert Opin. Ther. Targets* **18,** 000–000 (2014).

33. Davies, A. A. *et al.* Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Molecular Cell* **7,** 273–282 (2001).

34. Jeyasekharan, A. D. *et al.* A cancer-associated BRCA2 mutation reveals masked nuclear export signals controlling localization. *Nature Publishing Group* **20,** 1191–1198 (2013).

35. Rajendram, M. *et al.* Anionic Phospholipids Stabilize RecA Filament Bundles in Escherichia coli. *Molecular Cell* **60,** 374–384 (2015).

36. Bell, J. C. & Kowalczykowski, S. C. RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends Biochem. Sci.* **41,** 491–507 (2016).

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

38. Pellegrini, L. *et al.* Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* **420,** 287–293 (2002).

39. Shin, D. S. *et al.* Full-length archaeal Rad51 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2. *EMBO J.* **22,** 4566–4576 (2003).

40. Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nat. Struct. Mol. Biol.* **11,** 791–796 (2004).
41. Davies, O. R. & Pellegrini, L. Interaction with the BRCA2 C terminus protects RAD51-DNA filaments from disassembly by BRC repeats. *Nat. Struct. Mol. Biol.* **14,** 475–483 (2007).

42. Reymer, A., Frykholm, K., Morimatsu, K., Takahashi, M. & Nordén, B. Structure of human Rad51 protein filament from molecular modeling and site-specific linear dichroism spectroscopy. *Proceedings of the National Academy of Sciences* **106,** 13248–13253 (2009).

43. Schulze, W. X., Deng, L. & Mann, M. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* **1,** E1–E13 (2005).

44. Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439,** 168–174 (2006).

45. Kaushansky, A., Gordus, A., Chang, B., Rush, J. & MacBeath, G. A quantitative study of the recruitment potential of all intracellular tyrosine residues on EGFR, FGFR1 and IGF1R. *Mol. BioSyst.* **4,** 643–12 (2008).

46. Laurent F Hennequin *et al.* N-(5-Chloro-1,3-benzodioxol-4-yl)-7-[2-(4-methylpiperazin-1-yl)ethoxy]-5-(tetrahydro-2H-pyran-4-yloxy)quinazolin-4-amine, a Novel, Highly Selective, Orally Available, Dual-Specific c-Src/Abl Kinase Inhibitor†. *J. Med. Chem.* **49,** 6465–6488 (2006).

47. Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Research* **31,** 3635–3641 (2003).

48. Kundu, K., Mann, M., Costa, F. & Backofen, R. MoDPepInt: an interactive web server for prediction of modular domain-peptide interactions. *Bioinformatics* **30,** 2668–2669 (2014).

49. Xue, Y. *et al.* GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Research* **33,** W184–W187 (2005).

50. Wong, Y. H. *et al.* KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Research* **35,** W588–W594 (2007).

51. Frese, S. *et al.* The phosphotyrosine peptide binding specificity of Nck1 and Nck2 Src homology 2 domains. *J. Biol. Chem.* **281,** 18236–18245 (2006).

52. Pinna, L. A. & Ruzzene, M. How do protein kinases recognize their substrates? *Biochim. Biophys. Acta* **1314,** 191–225 (1996).

53. Ubersax, J. A. & Ferrell, J. E., Jr. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8,** 530–541 (2007).

54. Songyang, Z. *et al.* Catalytic Specificity of Protein-Tyrosine Kinases Is Critical for Selective Signaling. *Nature* **373,** 536–539 (1995).

55. Chung, B. M. *et al.* The role of cooperativity with Src in oncogenic transformation mediated by non-small cell lung cancer-associated EGF receptor mutants. *Oncogene* **28,** 1821–1832 (2009).

56. Cho, J. *et al.* Glioblastoma-Derived Epidermal Growth Factor Receptor Carboxyl-Terminal Deletion

Mutants Are Transforming and Are Sensitive to EGFR-Directed Therapies. *Cancer Research* **71,** 7587–7596 (2011).

57. Park, A. K. J., Francis, J. M., Park, W.-Y., Park, J.-O. & Cho, J. Constitutive asymmetric dimerization drives oncogenic activation of epidermal growth factor receptor carboxyl-terminal deletion mutants. *Oncotarget* **6,** 8839–8850 (2015).

58. Hunter, T. The Genesis of Tyrosine Phosphorylation. *Cold Spring Harbor Perspectives in Biology* **6,** a020644–a020644 (2014).

59. Chi, A. *et al.* Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 2193–2198 (2007).

60. Gould, K. L. & Nurse, P. Tyrosine phosphorylation of the fission yeast cdc2+ protein kinase regulates entry into mitosis. *Nature* **342,** 39–45 (1989).

61. Parker, L. L., Atherton-Fessler, S. & Piwnica-Worms, H. p107wee1 is a dual-specificity kinase that phosphorylates p34cdc2 on tyrosine 15. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 2917–2921 (1992).

62. Subramanyam, S., Ismail, M., Bhattacharya, I. & Spies, M. Tyrosine phosphorylation stimulates activity of human RAD51 recombinase through altered nucleoprotein filament dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **113,** E6045–E6054 (2016).

63. Li, W., Hu, P., Skolnik, E. Y., Ullrich, A. & Schlessinger, J. The SH2 and SH3 domain-containing Nck protein is oncogenic and a common target for phosphorylation by different surface receptors. *Mol. Cell. Biol.* **12,** 5824–5833 (1992).

64. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology* **31,** 1023–1031 (2013).

65. Lechner, J. F. & LaVeck, M. A. A serum-free method for culturing normal human bronchial epithelial cells at clonal density. *Journal of tissue culture methods* **9,** 43–48 (1985).

66. Vivanco, I. *et al.* Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discovery* **2,** 458–471 (2012).

67. Hidalgo, M. *et al.* Phase I and pharmacologic study of OSI-774, an epidermal growth factor receptor tyrosine kinase inhibitor, in patients with advanced solid malignancies. *J. Clin. Oncol.* **19,** 3267–3279 (2001).

68. Baselga, J. *et al.* Phase I safety, pharmacokinetic, and pharmacodynamic trial of ZD1839, a selective oral epidermal growth factor receptor tyrosine kinase inhibitor, in patients with five selected solid tumor types. *J. Clin. Oncol.* **20,** 4292–4302 (2002).

69. Burris, H. A. Phase I Safety, Pharmacokinetics, and Clinical Activity Study of Lapatinib (GW572016), a Reversible Dual Inhibitor of Epidermal Growth Factor Receptor Tyrosine Kinases, in Heavily Pretreated Patients With Metastatic Carcinomas. *Journal of Clinical Oncology* **23,** 5305–5313 (2005).

70. Eskens, F. A. L. M. *et al.* A phase I dose escalation study of BIBW 2992, an irreversible dual inhibitor of epidermal growth factor receptor 1 (EGFR) and 2 (HER2) tyrosine kinase in a 2-week on, 2-week off schedule in patients with advanced solid tumours. *Br. J. Cancer* **98,** 80–85 (2008).

71. Wong, K. K. *et al.* A Phase I Study with Neratinib (HKI-272), an Irreversible Pan ErbB Receptor Tyrosine Kinase Inhibitor, in Patients with Solid Tumors. *Clinical Cancer Research* **15,** 2552–2558 (2009).

72. Janne, P. A. *et al.* Phase I dose-escalation study of the pan-HER inhibitor, PF299804, in patients with advanced malignant solid tumors. *Clin. Cancer Res.* **17,** 1131–1139 (2011).

73. Sequist, L. V. *et al.* Rociletinib in EGFR-mutated non-small-cell lung cancer. *N Engl J Med* **372,** 1700–1709 (2015).

74. Jänne, P. A. *et al.* AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med* **372,** 1689–1699 (2015).

75. Li, S. *et al.* Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **7,** 301–311 (2005).

76. Meador, C. B. *et al.* Optimizing the sequence of anti-EGFR-targeted therapy in EGFR-mutant lung cancer. *Molecular Cancer Therapeutics* **14,** 542–552 (2015).

77. Regales, L. *et al.* Dual targeting of EGFR can overcome a major drug resistance mutation in mouse models of EGFR mutant lung cancer. *J. Clin. Invest.* **119,** 3000–3010 (2009).

78. Janjigian, Y. Y. *et al.* Dual inhibition of EGFR with afatinib and cetuximab in kinase inhibitor-resistant EGFR-mutant lung cancer with and without T790M mutations. *Cancer Discovery* **4,** 1036–1045 (2014).

79. Ahsan, A. *et al.* Wild-type EGFR is stabilized by direct interaction with HSP90 in cancer cells and tumors. *Neoplasia* **14,** 670–677 (2012).

80. Gallant, J. N. *et al.* EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib. *Cancer Discovery* **5,** 1155–1163 (2015).

81. Baik, C. S., Wu, D., Smith, C., Martins, R. G. & Pritchard, C. C. Durable Response to Tyrosine Kinase Inhibitor Therapy in a Lung Cancer Patient Harboring Epidermal Growth Factor Receptor Tandem Kinase Domain Duplication. *Journal of Thoracic Oncology* **10,** e97–e99 (2015).

82. Bradeen, H. A. Comparison of imatinib mesylate, dasatinib (BMS-354825), and nilotinib (AMN107) in an N-ethyl-N-nitrosourea (ENU)-based mutagenesis screen: high efficacy of drug combinations. *Blood* **108,** 2332–2338 (2006).

83. Ercan, D. *et al.* EGFR Mutations and Resistance to Irreversible Pyrimidine-Based EGFR Inhibitors. *Clin. Cancer Res.* **21,** 3913–3923 (2015).

84. Red Brewer, M. *et al.* Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proceedings of the National Academy of Sciences* **110,** E3595–604 (2013).

85. Lemmon, M. A. & Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141,** 1117–1134 (2010).

86. Huang, Y. *et al.* Molecular basis for multimerization in the activation of the epidermal growth factor receptor. *Elife* **5,** 18756 (2016).

87. Needham, S. R. *et al.* EGFR oligomerization organizes kinase-active dimers into competent signalling platforms. *Nature Communications* **7,** 13307 (2016).

88. Brar, G. A. Beyond the Triplet Code: Context Cues Transform Translation. *Cell* **167,** 1681–1692 (2016).

89. Sharp, P. M. & Li, W. H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15,** 1281–1295 (1987).

90. Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55,** 447–458 (1988).

91. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

92. Kovacs, E., Zorn, J. A., Huang, Y., Barros, T. & Kuriyan, J. A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor. *Annu. Rev. Biochem.* **84,** 739–764 (2015).

93. Dise, R. S., Frey, M. R., Whitehead, R. H. & Polk, D. B. Epidermal growth factor stimulates Rac activation through Src and phosphatidylinositol 3-kinase to promote colonic epithelial cell migration. *AJP: Gastrointestinal and Liver Physiology* **294,** G276–G285 (2007).

94. Fenstermaker, R. A., Ciesielski, M. J. & Castiglia, G. J. Tandem duplication of the epidermal growth factor receptor tyrosine kinase and calcium internalization domains in A-172 glioma cells. *Oncogene* **16,** 3435–3443 (1998).

95. Whitehead, R. H., VanEeden, P. E., Noble, M. D., Ataliotis, P. & Jat, P. S. Establishment of conditionally immortalized epithelial cell lines from both colon and small intestine of adult H-2Kb-tsA58 transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 587–591 (1993).

96. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31,** 827–832 (2013).

97. Ranganathan, V., Wahlin, K., Maruotti, J. & Zack, D. J. Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. *Nature Communications* **5,** 4516 (2014).

98. Sanjana, N. E., Shalem, O. & Zhang, F. correspondence. *Nat Meth* **11,** 783–784 (2014).

99. Guschin, D. Y. *et al.* in *Engineered Zinc Finger Proteins* (eds. Mackay, J. P. & Segal, D. J.) **649,** 247–256 (Humana Press, 2010).

100. Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D. & Bader, J. S. GeneDesign 3.0 is an updated synthetic biology toolkit. *Nucleic Acids Research* **38,** 2603–2606 (2010).

Notes

[a] Interestingly, while the majority of these clones (92) were derived from AALE cells transfected with the PX458-EGFR-RAD51#2 construct (which appears to be more efficient at inducing indels in EGFR and RAD51 **(Figure 4.1C**), only clones from PX458-EGFR-RAD51#1 construct were tested for presence of the fusion.

[b] Peter S. Choi, PhD, Meyerson lab, personal communication: “for creating the fusions, I had a very tough time generating clones with anything beyond intrachromosomal inversions/deletions. Even for those events, it was usually in the single-digit percentages… I think the efficiency is still too low to make it practical (although screening by PCR from genomic DNA is not too bad).”

[c] Adding weight to this hypothesis is the observation that, in yeast, RECA (the RAD51 homolog) polymerizes at and is stabilized by the cell membrane (35).

[d] EGFR-RAD51 residues corresponding to the oligomerization residues of RAD51 were determined by aligning the RAD51 CCDS (10062.1) with the EGFR-RAD51 ORF translation.

[e] This is argued because RAD51 homologs, such as RECA in yeast (36), could be said to predate the evolution of tyrosine phosphorylation (58). While there exist no know protein tyrosine kinases in yeast (59), tyrosine phosphorylation appears to play a key part in yeast signaling (60)—even if such phosphorylation is performed by dual specificity kinases, such as WEE1 (61).

[f] Cetuximab binds to the extracellular domain of EGFR and inhibits ligand binding (75). However, MTT assays are not the best method of testing for cetuximab efficacy; soft-agar assays are better for reasons that aren't quite clear (76).

[g] Based on the sequencing platform and alignments, the authors could not differentiate between the two mutations. Personal communication with Christina Baik, MD, MPH and Colin Pritchard, MD, PhD: "We didn't specifically investigate if the T790M mutation was in only one or both kinase domains, so we can't say for sure." NB: it could also be the case that the patient's other EGFR was mutated to T790M.

CHAPTER V: DISCUSSION

Summary of Findings

These studies have established two EGFR rearrangements, EGFR-KDD and EGFR-RAD51, as bona fide oncogenes and therapeutic targets in lung adenocarcinoma. Prior to these studies, EGFR-KDD (1-5) and EGFR fusions (6,7) had only been detected in glioma. Now, we have data that these alterations are recurrent in lung cancer and targetable with available anti-EGFR therapies[a]. Our work is already proving clinically relevant as these alterations have been identified in additional lung cancer patients—who may now have more targeted treatment options. Still, further translational studies are needed to understand how best to detect these alterations, treat them with available anti-EGFR (and other) agents, and manage the acquired resistance that companies treatment with anti-EGFR agents (8). In terms of basic science, these alterations offer interesting models to study EGFR biology because of the unique ways in which they may be activating the receptor, recruiting adaptors, and signaling downstream.

EGFR Rearrangements in Lung Adenocarcinoma

Current knowledge about EGFR-mutant lung adenocarcinoma is largely derived from the initial and most prevalent mutations detected in patients: EGFR-L858R and EGFR-ex19del (9-11). As both these mutations activate the EGFR TKD and sensitize cancers harboring these mutations to EGFR-TKIs, they have come to define EGFR-mutant lung adenocarcinoma. Indeed, most phase III trials investigating first-line treatment of patients with EGFR-TKIs defined EGFR-mutant lung adenocarcinoma as those cancers harboring these 'common' mutations (EGFR-L858R or EGFR-ex19del)[b]. As a result, all first and second generation EGFR-TKIs are FDA approved to treat these specific common EGFR mutations. This is problematic because there is significantly less evidence to support the treatment of tumors harboring uncommon EGFR mutations (defined as anything but EGFR-L858R or EGFR-ex19del) and because it presumes that EGFR-L858R and EGFR-ex19del are equivalent mutations. However, a wealth of pre-clinical and clinical data has shown that not all EGFR mutations are created equal. For one, the kinase activity of common EGFR mutants, their sensitivity to EGFR TKIs, and structures are all slightly, but significantly, different (20,21). Clinically, this translates to patients with EGFR-L858R and EGFR-ex19del tumors having significantly different progression free and overall survivals (22)[c.] Together, these data point towards a need for better understanding of individual and less common EGFR mutations present in lung adenocarcinoma.

While L858R and ex19del do account for the majority of EGFR-mutant lung adenocarcinoma, their reported prevalence (up to 90% of EGFR-mutant lung adenocarcinoma (24)) is deceivingly high due to clinical laboratory detection methods. Because the first-identified TKI-sensitizing EGFR mutations in lung adenocarcinoma (G719X, ex19del, L858R, and L861Q (9-11)) were all single nucleotide variants (SNVs) or small insertions/deletions (indels), PCR followed by dideoxynucleotide sequencing was initially used for detection of variants. While these techniques have evolved, leading to more sensitive multiplexed assays that query mutations in several genes at once (25-27), PCR-based methods remain the standard for detection of EGFR mutations in lung adenocarcinoma (28)[d]. Because they rely on panels of pre-determined primers, PCR-based tests are biased towards known mutations in exon 18–21 of EGFR, which encode the majority of the EGFR TKD. As a result of this bias, some currently available clinical diagnostic tests are specific—for EGFR L858R, ex19del, and a few other mutations—and not very sensitive for EGFR alterations; it is possible that they over-report EGFR SNVs and indels (L858R or ex19del) while under-reporting larger scale rearrangements that maintain the wild type sequence of EGFR (such the KDD and fusion events).

To overcome sensitivity limits in PCR-based methods, next generation sequencing (NGS) has recently made its way into the clinical realm for tumor mutational testing[e]. While NGS allows for comprehensive whole genome sequencing (WGS), throughput and turnaround time have prevented this method from becoming part of routine clinical testing. Instead, hybrid-capture-based methods, somewhat of a middle ground between WGS and PCR-based methods, are becoming the standard for identifying actionable targets in lung adenocarcinoma. These methods pair capture of transcripts using biotinylated primers, NGS, and large scale bioinformatics to enable detection of known mutations and discovery of new alterations. While this approach (specifically, the FoundationOne® test (31))[f] allowed us to detect EGFR-KDD and EGFR fusions in several patients, it does have it's drawbacks: namely, the panels of probes used in these assays interrogate only specific regions of cancer-related genes and may therefore miss detection of certain novel mutations in other locations of the genome.

As a result of these diagnostic limitations, we believe that our reported prevalence of EGFR-KDD and EGFR fusions is smaller than the true proportion of these rearrangements in lung adenocarcinoma. PCR-

based methods would have missed (and in some patients did) these alterations because they maintain an intact EGFR TKD sequence. At the same time, NGS methods may have missed EGFR-KDD and EGFR fusions because genomic breakpoints for these rearrangements occur in introns. While available NGS diagnostics usually sequence the entirety of known cancer driver coding sequences, they only sequence specific introns in select genes known to commonly undergo rearrangements[g]. It is our understanding that EGFR-KDD and EGFR fusions were identified as a result of exonic EGFR sequencing reads extending into introns and serendipitously identifying the rearrangements' breakpoints. Clearly, the identification of EGFR-KDD and EGFR fusions point towards a need for hybrid-capture-based probes located in the introns of EGFR and RAD51. Such probes would be more likely to identify these rearrangements, provide patients with the opportunity to receive targeted therapy, and increase the relative prevalence of EGFR-KDD and RAD51. Further refinements of these assays, combined with the ever-increasing speed and ever-decreasing cost of WGS, may help discover more clinically relevant EGFR fusions or alterations in the future.

The ability to detect an increasing number of EGFR (and other) alterations led to the discovery of EGFR-KDD and EGFR fusions; but, as with other uncommon EGFR alterations, led to questions with regards to how best to treat these alterations. While evidence from our laboratory models and from the treated index patients suggested that these alterations are actionable with EGFR-TKIs, it isn't clear which EGFR-TKI is most potent against these alterations—due to likely subtle structural differences amongst agents and rearrangements (32)—and/or whether combinations of agents may be better suited for therapy. Even if optimally treated with EGFR-TKIs, tumors harboring uncommon EGFR mutations acquire resistance (33). It is also notable that patients whose tumors harbored EGFR re-arrangements developed the T790M resistance mutations after treatment with EGFR-TKIs. It is unclear whether these patients could benefit from treatment with the T790M-selective EGFR-TKI osimertinib (34), from dual EGFR blockade with afatinib (an EGFR-TKI) and cetuximab (the anti-EGFR antibody) (35,36), or from combination treatment with a MEK inhibitor (37). More work is needed to determine the optimal EGFR-TKIs used to treat EGFR rearrangements, the approach in the setting of acquired resistance, and, potentially, the sequence of therapies (38).

EGFR Rearrangements and the Biology of EGFR

Because our data suggests that the EGFR-KDD is activated via intramolecular dimerization, the alteration could be used to test the current model of physiological / WT- EGFR activation: ligand-induced, receptor-mediated dimerization of the ECD (39-41) that is propagated across the plasma membrane via interactions within the TM (42,43) and JM domains (44-46) and results in allosteric activation of the TKD via asymmetric dimerization (47,48). These various moving pieces contribute both activating and inhibiting forces to maintain EGFR mitogenic signaling strength (tone) and prevent inappropriate ligand-independent activation. However, it is currently unclear which of these counterbalancing regulatory forces/movements dominate when the receptor is activated. As our data point towards the EGFR-KDD being constitutively active in the absence of ligand but further activated in the presence of ligand (by virtue of asymmetric dimerization), they also indicate that EGFR activation may be less reliant on the ECD and membrane interactions than on asymmetric dimerization (and the JM latch). A crystal structure of the EGFR-KDD ICD[h] could support our data and reveal the most critical pieces for EGFR activation.

EGFR-RAD51, on the other hand, could be used to test models of RTK signaling. It is commonly held that the first response to autophosphorylation of RTKs is the recruitment and activation of a host of downstream signaling adaptors (49). However, EGFR-RAD51 does not contain a C-term beyond the kinase domain. As a result, this fusion lacks signaling adaptor binding sites. Interestingly, EGFR-RAD51 is able to activate downstream signaling via the MAPK and PI3K/AKT pathways. Because it has been demonstrated that EGFR can multimerize (50), and because a higher local density of phospho-tyrosines is able to recruit different adaptor molecules (51), we hypothesize that EGFR-RAD51 is able to signal downstream by formation of large concatemers. These large active signaling EGFR complexes could contain such a high density of phospho-tyrosine 869 that these residues could recruit MAPK and PI3K/AKT pathway molecules such as GRB2 and SHC in addition to SRC. Thus EGFR-RAD51 could help validate that promiscuity changes with local protein concentration may contribute to the oncogenic potential of RTKS and other signaling proteins (52).

Conclusion

Over the past decade, NSCLC, specifically, EGFR-mutant lung adenocarcinoma, has served as a model for personalized medicine. The field has taken a complex disease (NSCLC), traced the problem to one of its genetic origins (EGFR), and rationally treated the root cause (with EGFR-TKIs). Data in these studies demonstrate the diversity of modes through which EGFR protein is activated in lung adenocarcinoma. The

cancer is able to maintain EGFR tone through a variety of radical genomic alterations, including gene duplication and fusion[i]. The resultant gene products, the EGFR-KDD and EGFR-RAD51, are recurrent oncogenic alterations and therapeutically actionable targets, highlighting the importance of EGFR in lung adenocarcinoma. Similar to the field as a whole, these studies have taken a complex disease (index cases of patients with driver-negative lung adenocarcinoma), traced the problem to one of its genetic origins (EGFR re-arrangements), demonstrated the oncogenicity of these proteins, highlighted the actionability of EGFR-rearrangements, and yielded valuable models for future study of biology. There is still much to learn, including how EGFR re-arrangements are activated and how they propagate signal. Fortunately, in an age of precision medicine, there exist the knowledge, technologies, and systems to address these problems and readily study all aspects of EGFR—in the laboratory and in the clinic.

References

1. Steck, P. A. *et al.* Expression of epidermal growth factor receptor and associated glycoprotein on cultured human brain tumor cells. *J. Cell. Biochem.* **32,** 1–10 (1986).

2. Steck, P. A., Lee, P., Hung, M. C. & Yung, W. K. Expression of an altered epidermal growth factor receptor by human glioblastoma cells. *Cancer Research* **48,** 5433–5439 (1988).

3. Panneerselvam, K., Kanakaraj, P., Raj, S., Das, M. & Bishayee, S. Characterization of a novel epidermal-growth-factor-receptor-related 200-kDa tyrosine kinase in tumor cells. *Eur. J. Biochem.* **230,** 951–957 (1995).

4. Fenstermaker, R. A., Ciesielski, M. J. & Castiglia, G. J. Tandem duplication of the epidermal growth factor receptor tyrosine kinase and calcium internalization domains in A-172 glioma cells. *Oncogene* **16,** 3435–3443 (1998).

5. Ciesielski, M. J. & Fenstermaker, R. A. Oncogenic epidermal growth factor receptor mutants with tandem duplication: gene structure and effects on receptor function. *Oncogene* **19,** 810–820 (2000).

6. Frattini, V. *et al.* Frattubu.V_2013_Nature Genetics_Genomic landscape of GBM - EGFR fusions. *Nature Publishing Group* **45,** 1141–1149 (2013).

7. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155,** 462–477 (2013).
8. Camidge, D. R., Pao, W. & Sequist, L. V. Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nature Publishing Group* 1–9 (2014).

9. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 13306–13311 (2004).

10. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304,** 1497–1500 (2004).

11. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350,** 2129–2139 (2004).

12. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361,** 947–957 (2009).

13. Sequist, L. V. *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *Journal of Clinical Oncology* **31,** 3327–3334 (2013).

14. Wu, Y.-L. *et al.* Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol.* **15,** 213–222 (2014).

15. Maemondo, M. *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* **362,** 2380–2388 (2010).

16. Mitsudomi, T. *et al.* Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. *Lancet Oncol.* **11,** 121–128 (2010).

17. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13,** 239–246 (2012).

18. Zhou, C. *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol.* **12,** 735–742 (2011).

19. Wu, Y.-L. *et al.* First-line erlotinib versus gemcitabine/cisplatin in patients with advanced EGFR mutation-positive non-small-cell lung cancer: analyses from the phase III, randomized, open-label, ENSURE study. *Annals of Oncology* **26,** 1883–1889 (2015).

20. Carey, K. D. *et al.* Kinetic analysis of epidermal growth factor receptor somatic mutant proteins shows increased sensitivity to the epidermal growth factor receptor tyrosine kinase inhibitor, erlotinib. *Cancer Research* **66,** 8163–8171 (2006).

21. Yun, C.-H. *et al.* Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **11,** 217–227 (2007).

22. Kuan, F.-C. *et al.* Overall survival benefits of first-line EGFR tyrosine kinase inhibitors in EGFR-mutated non-small-cell lung cancers: a systematic review and meta-analysis. *Br. J. Cancer* **113,** 1519–1528 (2015).

23. Sebastian, M., Schmittel, A. & Reck, M. First-line treatment of EGFR-mutated nonsmall cell lung cancer: critical review on study methodology. *Eur Respir Rev* **23,** 92–105 (2014).

24. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat. Rev. Cancer* **7,** 169–181 (2007).

25. Pan, Q., Pao, W. & Ladanyi, M. Rapid polymerase chain reaction-based detection of epidermal growth factor receptor gene mutations in lung adenocarcinomas. *J Mol Diagn* **7,** 396–403 (2005).

26. Dias-Santagata, D. *et al.* Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. *EMBO Mol Med* **2,** 146–158 (2010).

27. Su, Z. *et al.* A platform for rapid detection of multiple oncogenic mutations with relevance to targeted therapy in non-small-cell lung cancer. *J Mol Diagn* **13,** 74–84 (2011).

28. Tan, D. S. W. *et al.* The International Association for the Study of Lung Cancer Consensus Statement on Optimizing Management of EGFR Mutation-Positive Non-Small Cell Lung Cancer: Status in 2016. *Journal of Thoracic Oncology* **11,** 946–963 (2016).

29. Weiss, G. J. *et al.* Evaluation and comparison of two commercially available targeted next-generation sequencing platforms to assist oncology decision making. *Onco Targets Ther* **8,** 959–967 (2015).

30. Kuderer, N. M. *et al.* Comparison of 2 Commercially Available Next-Generation Sequencing Platforms in Oncology. *JAMA Oncol* (2016).

31. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology* **31,** 1023–1031 (2013).

32. Vivanco, I. *et al.* Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discovery* **2,** 458–471 (2012).

33. Lohinai, Z. *et al.* Distinct Epidemiology and Clinical Consequence of Classic Versus Rare EGFR Mutations in Lung Adenocarcinoma. *J Thorac Oncol* **10,** 738–746 (2015).

34. Cross, D. A. E. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discovery* **4,** 1046–1061 (2014).

35. Regales, L. *et al.* Dual targeting of EGFR can overcome a major drug resistance mutation in mouse models of EGFR mutant lung cancer. *J. Clin. Invest.* **119,** 3000–3010 (2009).

36. Janjigian, Y. Y. *et al.* Dual inhibition of EGFR with afatinib and cetuximab in kinase inhibitor-resistant EGFR-mutant lung cancer with and without T790M mutations. *Cancer Discovery* **4,** 1036–1045 (2014).

37. Tricker, E. M. *et al.* Combined EGFR/MEK Inhibition Prevents the Emergence of Resistance in EGFR-Mutant Lung Cancer. *Cancer Discovery* **5,** 960–971 (2015).

38. Meador, C. B. *et al.* Optimizing the sequence of anti-EGFR-targeted therapy in EGFR-mutant lung cancer. *Molecular Cancer Therapeutics* **14,** 542–552 (2015).

39. Ogiso, H. *et al.* Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* **110,** 775–787 (2002).

40. Garrett, T. P. J. *et al.* Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* **110,** 763–773 (2002).

41. Schlessinger, J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110,** 669–672 (2002).

42. Arkhipov, A. *et al.* Architecture and membrane interactions of the EGF receptor. *Cell* **152,** 557–569 (2013).

43. Endres, N. F. *et al.* Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. *Cell* **152,** 543–556 (2013).

44. Thiel, K. W. & Carpenter, G. Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. *Proceedings of the National Academy of Sciences* **104,** 19238–19243 (2007).

45. Jura, N. *et al.* Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell* **137,** 1293–1307 (2009).

46. Red Brewer, M. *et al.* The juxtamembrane region of the EGF receptor functions as an activation domain. *Molecular Cell* **34,** 641–651 (2009).

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

48. Valley, C. C. *et al.* Enhanced dimerization drives ligand-independent activity of mutant epidermal growth factor receptor in lung cancer. *Molecular Biology of the Cell* **26,** 4087–4099 (2015).

49. Lemmon, M. A. & Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141,** 1117–1134 (2010).

50. Huang, Y. *et al.* Molecular basis for multimerization in the activation of the epidermal growth factor receptor. *Elife* **5,** 18756 (2016).

51. Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439,** 168–174 (2006).

52. Fan, Q.-W. *et al.* EGFR phosphorylates tumor-derived EGFRvIII driving STAT3/5 and progression in glioblastoma. *Cancer Cell* **24,** 438–449 (2013).

53. Stein, R. A. & Staros, J. V. Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis. *BMC Evol. Biol.* **6,** 79 (2006).

54. Kovacs, E., Zorn, J. A., Huang, Y., Barros, T. & Kuriyan, J. A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor. *Annu. Rev. Biochem.* **84,** 739–764 (2015).

55. Yarden, Y. & Sliwkowski, M. X. Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell Biol.* **2,** 127–137 (2001).

56. Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34,** 4845–4854 (2015).

Notes

[a] EGFR-KDD was also found in other cancers but additional data is needed to determine if EGFR-KDD is therapeutically actionable in non-lung cancers.

[b] The only phase III trials that tested tumors for EGFR mutations other than EGFR-L858R and EGFR-ex19del were the IPASS (12) and LUX-Lung trials (13,14). Both of these trials tested for EGFR-T790M, L861Q, G719X, S768I, and ex20ins in addition to the common mutations. The NEJ002 (15), WJTOG3405 (16), EURTRAC (17), OPTIMAL (18), and ENSURE (19) trials only tested for EGFR-L858R and EGFR-ex19del.

[c] Regardless of trial/treatment, patients with EGFR-ex19del tumors have significantly better PFS and OS than those with EGFR-L858R (22,23).

[d] In fact, the only FDA-approved diagnostic test, the cobas® EGFR Mutation Test, is a PCR-based test that only probes for 42 different SNVs and indels in exons 18-21.

[e] While these tests are increasingly becoming standard of care, they must be approached with care. Not every NGS assay is equal (often yielding differing results on the same samples (29,30)), and none are FDA approved.

[f] The test simultaneously sequences the coding region of 315 cancer-related genes plus introns from 28 genes often rearranged or altered in cancer.

[g] These tests do sequence regions where known fusion events do occur, such as the introns of EML4, ALK, CD74, and ROS1.

[h] As of this writing, the Lovly lab is working with the Kuriyan lab to obtain this structure.

[i] That being said, gene duplication and fusion events are not so rare—especially, and intriguingly so, with EGFR. In fact, three duplication events led to the four HER family members that are present in vertebrates (53-55). Moreover, oncogenic fusions can commonly be found in normal cells and tissues (56).

APPENDIX

Diagnosis and genotype	# cases
Total # of clinical samples genotyped	56,240
Total # of non-small cell lung cancer cases	10,097
# of cases harboring EGFR L858R mutation	472
# of cases harboring EGFR exon 19 deletion	668
# of cases harboring EGFR G719X mutation	92
# of cases harboring EGFR L861Q mutation	36
# of cases harboring EGFR S768I mutation	36
# of cases harboring EGFR exon 20 insertion	193
# of cases harboring EGFR exon 19 insertion	4
# of cases harboring EGFR exon 18-25 kinase domain duplication	4
# of cases harboring EGFR fusions	5

Table S2.1: Summary of *EGFR* alterations in NSCLC identified by FoundationOne.
The Foundation Medicine database of annotated clinical cases was interrogated for the presence of various *EGFR* genomic alterations known to recur in lung cancer. The 'X' in G719X denotes any of several amino acids which may be substituted for glycine in this variant. *EGFR* fusions events were defined by a genomic breakpoint in *EGFR* exons 23 through intron 25. This table corresponds to Supplementary Table 1 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

Patient No.	Gender	Age	Fusion Detected	Event Coverage	5'gene breakpoint	3'gene breakpoint	Exons	Other EGFR alterations	Other alterations
1	Female	35	EGFR-RAD51	345	EGFR intron 24 (chr7:55268583)	RAD51 intron 3 (chr15:40995548)	EGFR exons 1-24 RAD51 exons 4-9	None	• CDKN2A loss • CDKN2B loss • MYC amplification
2	Female	21	EGFR-RAD51	304	EGFR intron 24 (chr7:55268151)	RAD51 intron 3 (chr15:40995888)	EGFR exons 1-24 RAD51 exons 4-9	None	• CTNNB1 T41A • RANBP2 T951M • CDKN2B loss • TP53 loss • CDKN2A loss
3	Female	42	EGFR-PURB	21	EGFR exon 25 (chr7:55269019)	PURB 3' UTR (chr7:44919491)	EGFR exons 1-25 PURB 3' UTR	None	• CREBBP S893L • ARID1A Q1095*
4	Male	38	EGFR-RAD51	197	EGFR intron 24 (chr7:55268641)	RAD51 intron 3 (chr15:40993432)	EGFR exons 1-24 RAD51 exons 4-9	None	• RBM10 S570fs*133 • CHD4 D316H • MYC amplification • MCL1 amplification • IKBKE amplification • PIK3C2B amplification • MDM4 amplification
5	Female	60	EGFR-RAD51	223	EGFR intron 24 (chr7:55268412)	RAD51 intron 3 (chr15:40995404)	EGFR exons 1-24 RAD51 exons 4-9	EGFR amplification	• GRIN2A R1318W • ATR Q2408* • ARID1A P1484fs*10 • FGF3 amplification • FGF4 amplification • CDKN2A loss • PDCD1LG2 amplification • CCND1 amplification • CD274 amplification • FGF19 amplification • EMSY amplification • HGF amplification • JAK2 amplification • CDKN2B loss

Table S2.2: Summary of genomic coordinates for the kinase fusions identified in this study.
The sample number is depicted in the far left column. For tumor samples in which a kinase fusion was identified, the genomic coordinates for each of the fusion genes is indicated. Additional alterations are noted. N/A = Not Applicable, * = breakpoint not clear. This table corresponds to Supplementary Table 2 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

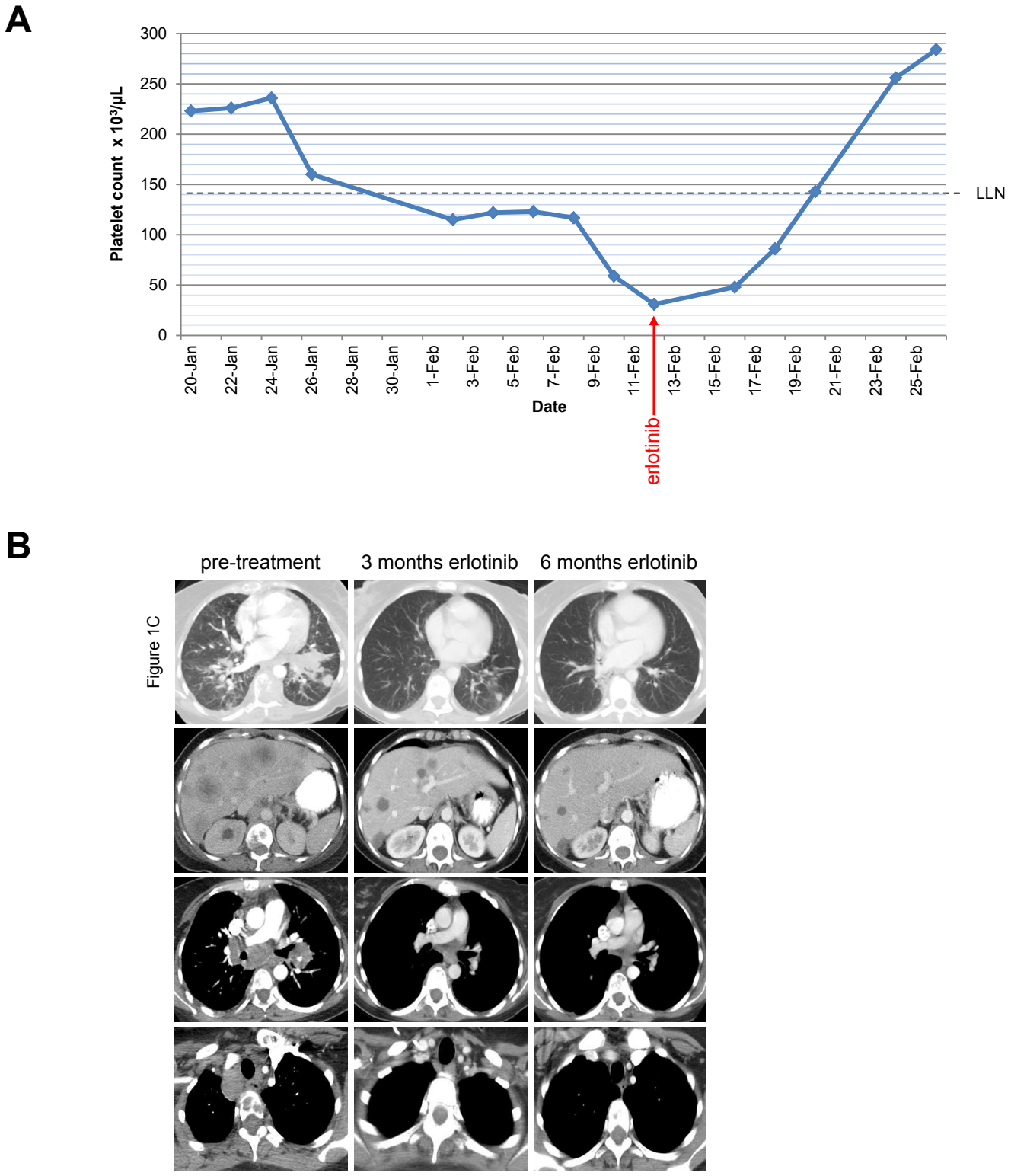
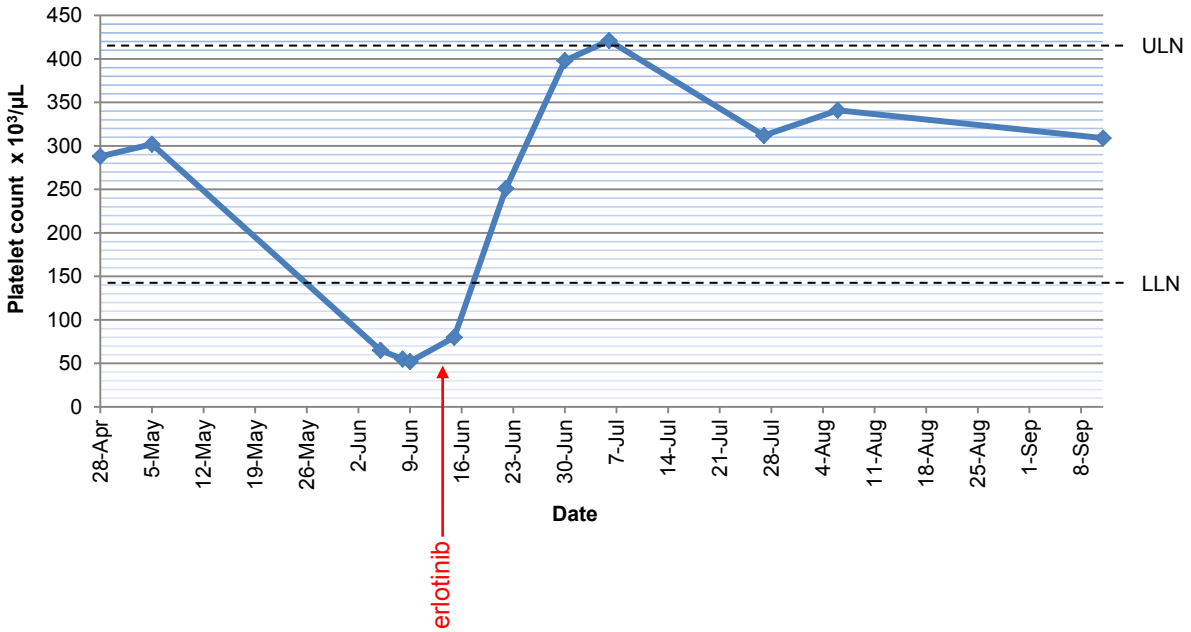


Figure S2.1: Additional information for Patient 1.
A, platelet counts in peripheral blood samples from patient 1 before and after the initiation of erlotinib therapy. ULN= upper limit of normal for platelet count. LLN= lower limit of normal for platelet count. **B,** Serial CT scans from patient 1 documenting response to the EGFR TKI, erlotinib. This figure includes the images from Figure 2.1C (as denoted on the left side of the image) and as well as additional images showing response to erlotinib at other disease sites. This figure corresponds to Supplementary Figure 1 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

A



B

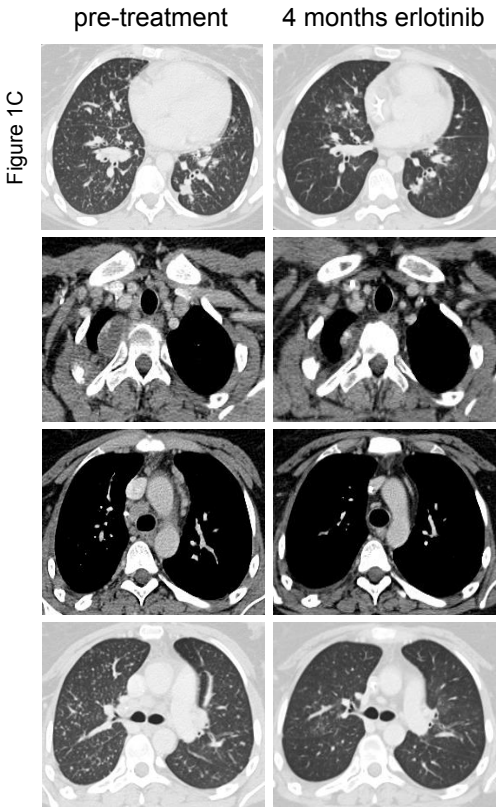


Figure S2.2: Additional information for Patient 2.
A, platelet counts in peripheral blood samples from patient 2 before and after the initiation of erlotinib therapy. ULN= upper limit of normal for platelet count. LLN= lower limit of normal for platelet count. **B,** Serial CT scans from patient 2 documenting response to the EGFR TKI, erlotinib. This figure includes the images from Figure 2.1C (as denoted on the left side of the image) and as well as additional images showing response to erlotinib at other disease sites. This figure corresponds to Supplementary Figure 2 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

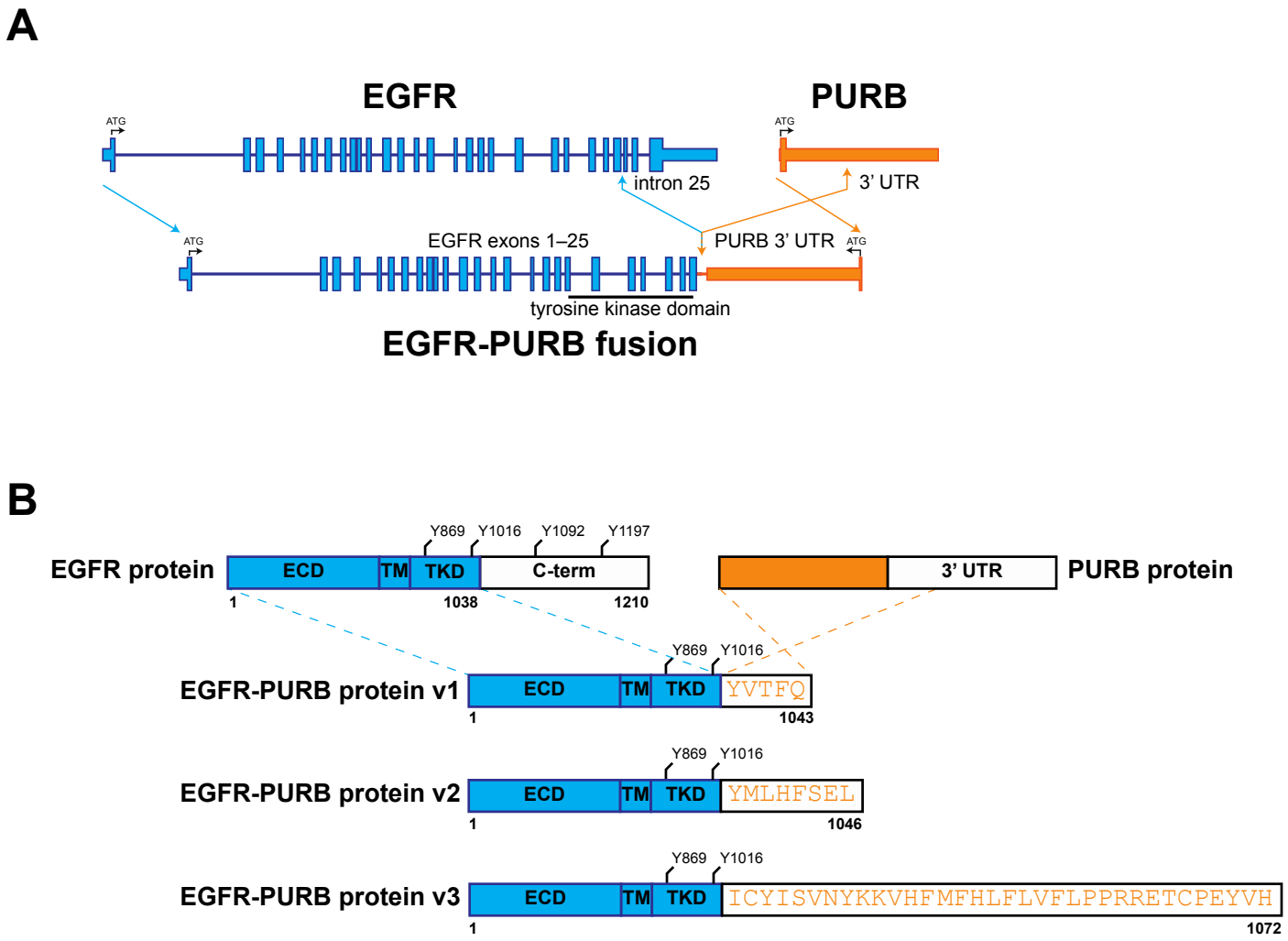


Figure S2.3: Additional information for Patient 3.
A, scaled representation of EGFR-PURB depicting the genomic structure of the fusion. ATG = translational start site. Blue = EGFR. Orange = PURB. **B,** putative representations of EGFR-PURB protein structures. Numbers correspond to amino acid residues. Y = tyrosine residue. ECD = extracellular domain. TM = transmembrane helix. TKD = tyrosine kinase domain. The three different EGFR-PURB protein variants (v1, v2, v3) are based on the different potential reading frames of the fusion sequence. **C,** Serial CT scans from patient 3 documenting response to the EGFR TKI, erlotinib. This figure includes the images from Figure 1C (as denoted on the left side of the image) and as well as additional images showing response to erlotinib at other disease sites. This figure corresponds to Supplementary Figure 3 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

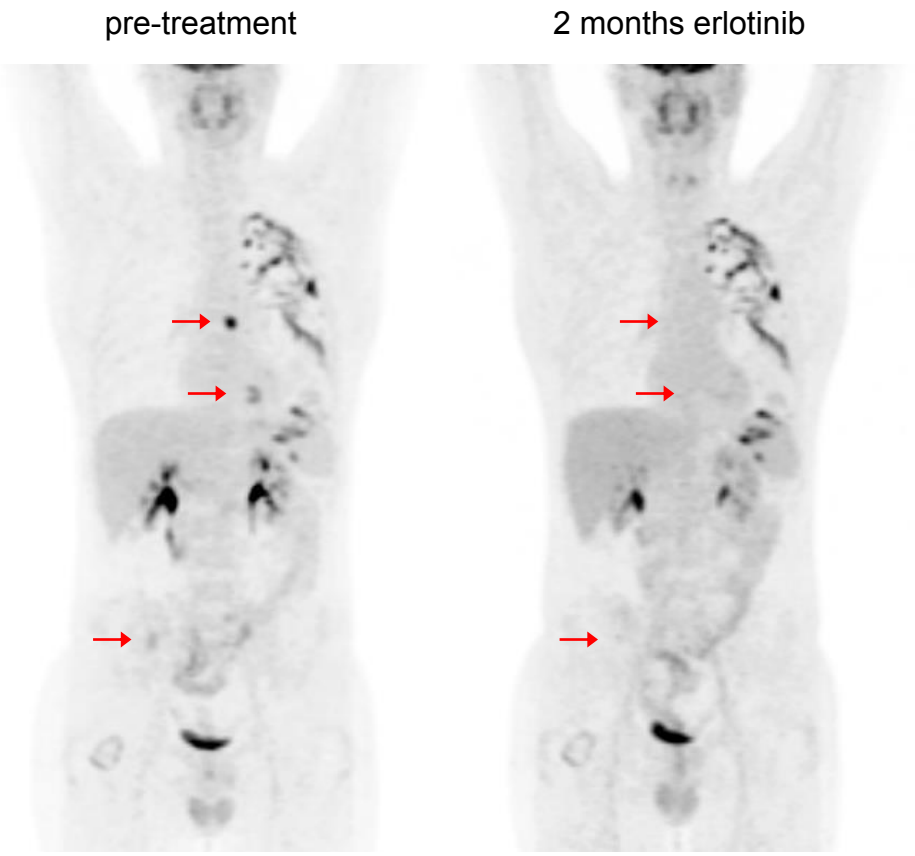


Figure S2.4: Additional clinical information for Patient 4.
Serial PET scans from patient 4 documenting response to the EGFR TKI, erlotinib. The images show response to erlotinib at disease sites other than those shown in Figure 1C. This figure corresponds to Supplementary Figure 4 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

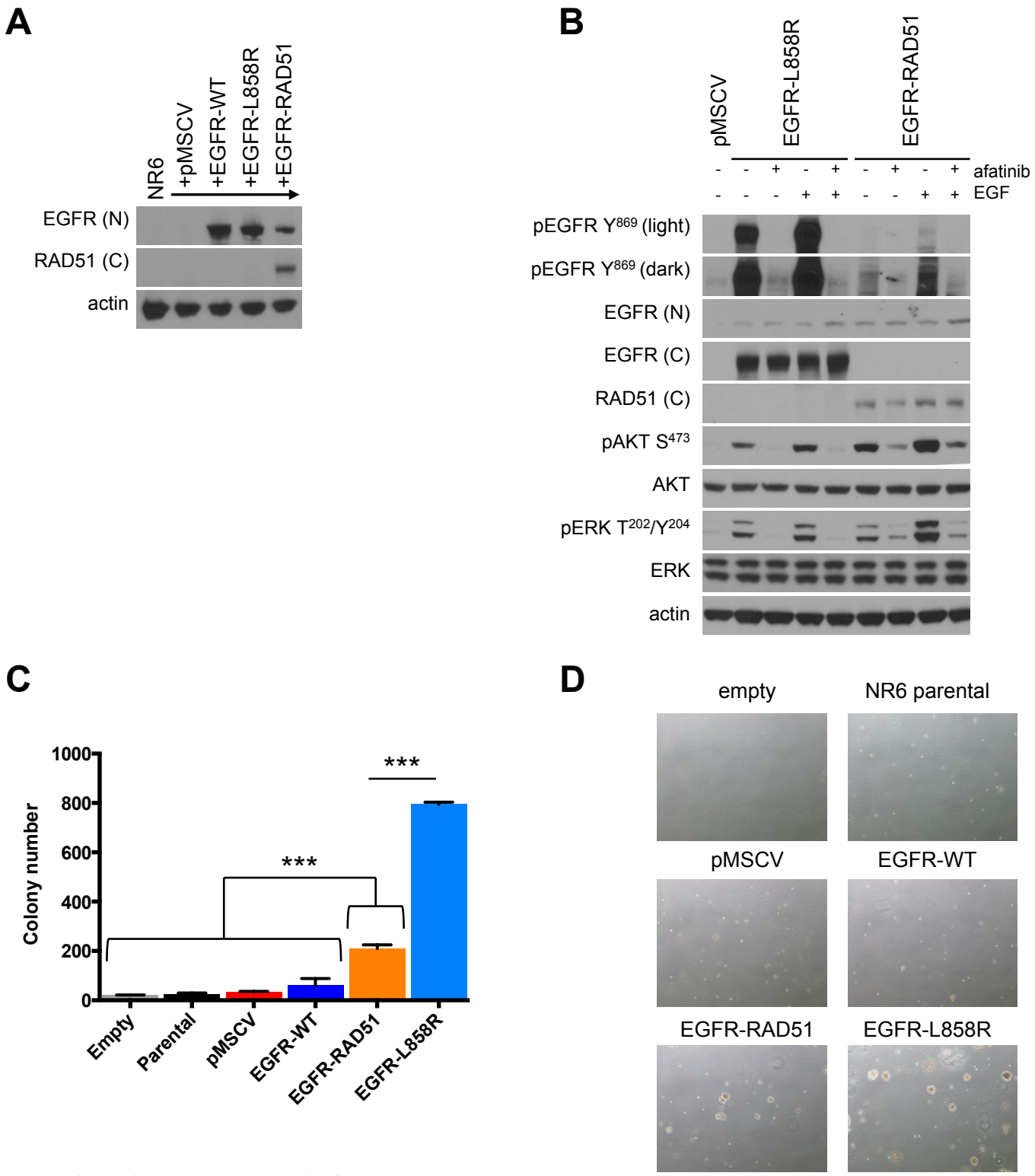
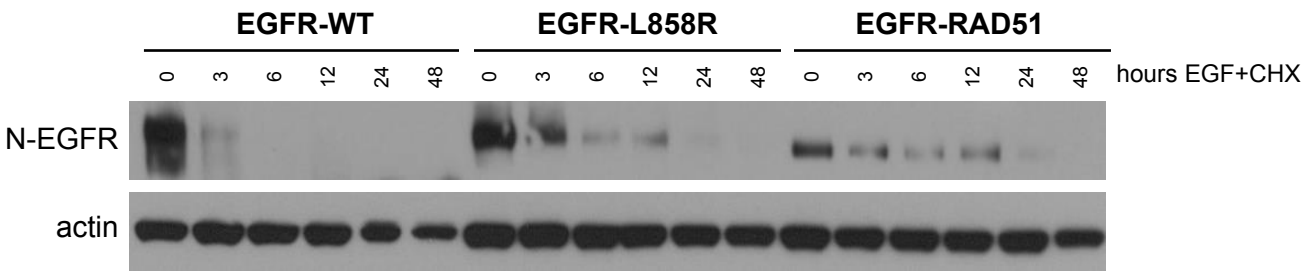
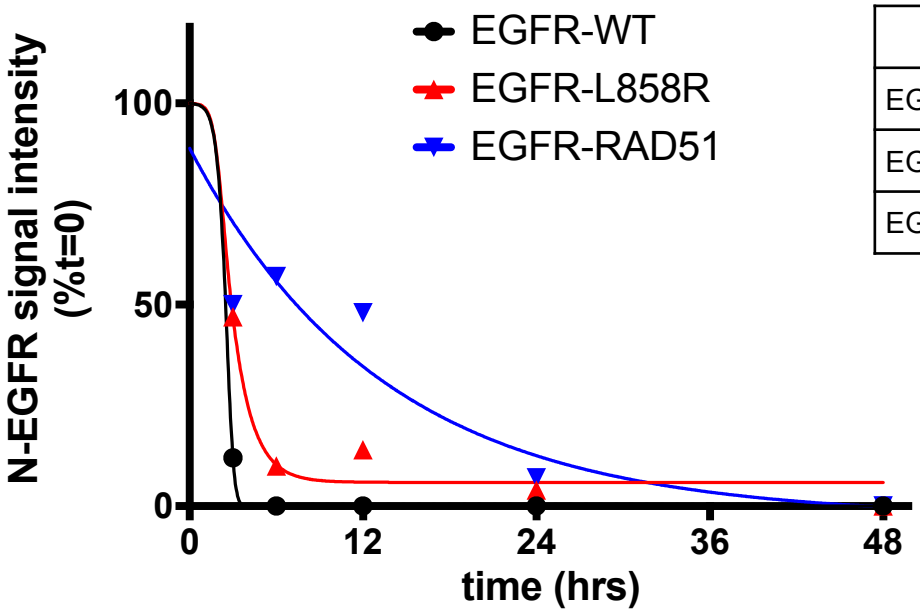


Figure S2.5: Characterization of EGFR-RAD51 in NR6 cells.
A, NR6 lines stably expressing pMSCV (vector only), EGFR-WT, EGFR-L858R or EGFR–RAD51 were subjected to western blot analysis with the indicated antibodies. EGFR-RAD51 fusion is detected with both the N-terminal EGFR antibody [EGFR(N)] and with the RAD51 antibody. There is no cross reactivity between wild-type RAD51 protein, which has MW ~35kD, and the EGFR-RAD51 fusion. **B,** NR6 expressing EGFR variants were serum starved for 16 hours, treated with 100 nM afatinib for 1 hour followed by 50 ng/mL EGF for 5 minutes, and subjected to western blot analysis with the indicated antibodies. **C,** NR6 cells stably expressing the indicated constructs (empty = no cells; pMSCV = vector only) were plated in triplicate in soft agar, grown for three weeks, and quantified for colony formation. ***, $P < 0.0001$. **D,** representative 4x microphotographs of soft agar assays showing empty wells (no cells) and NR6 cells transduced with pMSCV vector, EGFR-WT, EGFR-RAD51, and EGFR-L858R after three weeks of growth in soft agar. This figure corresponds to Supplementary Figure 5 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

A



B



Interpolated receptor half life	
EGFR-WT	2:29 hrs
EGFR-L858R	2:48 hrs
EGFR-RAD51	4:06 hrs

Figure S2.6: Relative stability of EGFR-WT, -L858R, and -RAD51.
A, Ba/F3 cells harboring indicated EGFR variants were serum starved overnight and then treated with the combination of 50 ng/mL EGF and 50 μg/mL cyclohexamide (CHX) at time = 0. Cells were lysed for western blot analysis with indicated antibodies at various timepoints. **B,** Quantification of N-EGFR western blot signal (using FiJi) relative to time = 0 (and normalized to actin) for each of the EGFR variants. Intensity was plotted into Prism and the receptor half-life determined by asymmetric sigmoidal interpolation. This figure corresponds to Supplementary Figure 6 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

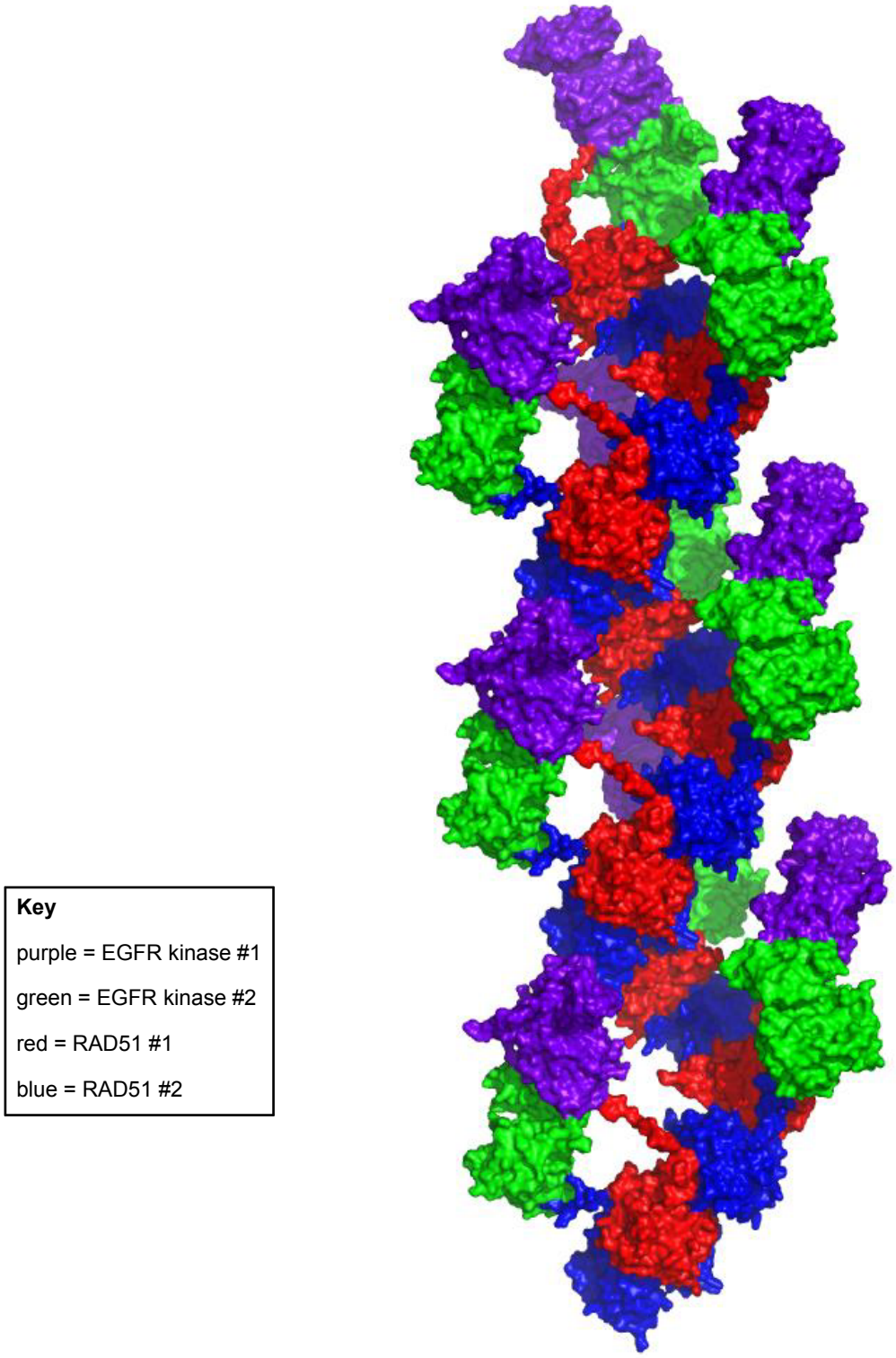


Figure S2.7: Structural model of an EGFR-RAD51 filament.
Space-filling model of the EGFR-RAD51 kinase domains illustrating hypothetical filament based on the RAD51 filament structure. Purple = first EGFR kinase domain; green = second EGFR kinase domain; red = first RAD51 partner; blue = second RAD51 partner. This figure corresponds to Supplementary Figure 7 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

		EGFR-L858R	EGFR-RAD51	difference
IC_{50} (nM)	erlotinib	5.11 ± 1.35	2.65 ± 1.91	*n.s. p*=0.23
	afatinib	≤ 1.5	8.47 ± 5.04	*n.s. p*=0.17
	osimertinib	3.51 ± 1.05	3.52 ± 1.49	*n.s. p*=0.08

Table S2.3 Results of MTT curve fitting from Prism.
These data correspond to the graphs shown in Figure 2.3A. This table corresponds to Supplementary Table 3 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

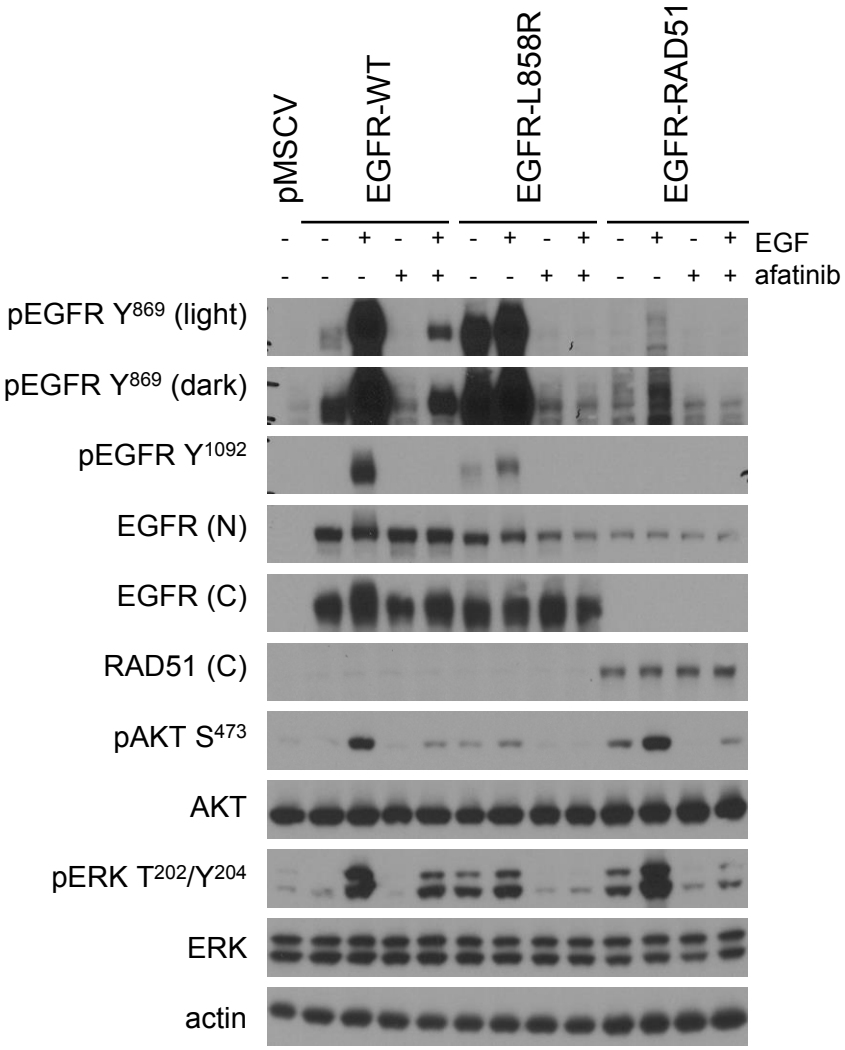


Figure S2.8: On-target inhibition of EGFR-RAD51 by EGFR-TKI.
Ba/F3 cells stably expressing EGFR-WT, EGFR-L858R and EGFR-RAD51 were serum starved for 16 hours, treated with 100 nM afatinib for 1 hour followed by 50 ng/mL EGF for 5 minutes, and subjected to western blot analysis with the indicated antibodies. This figure corresponds to Supplementary Figure 8 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

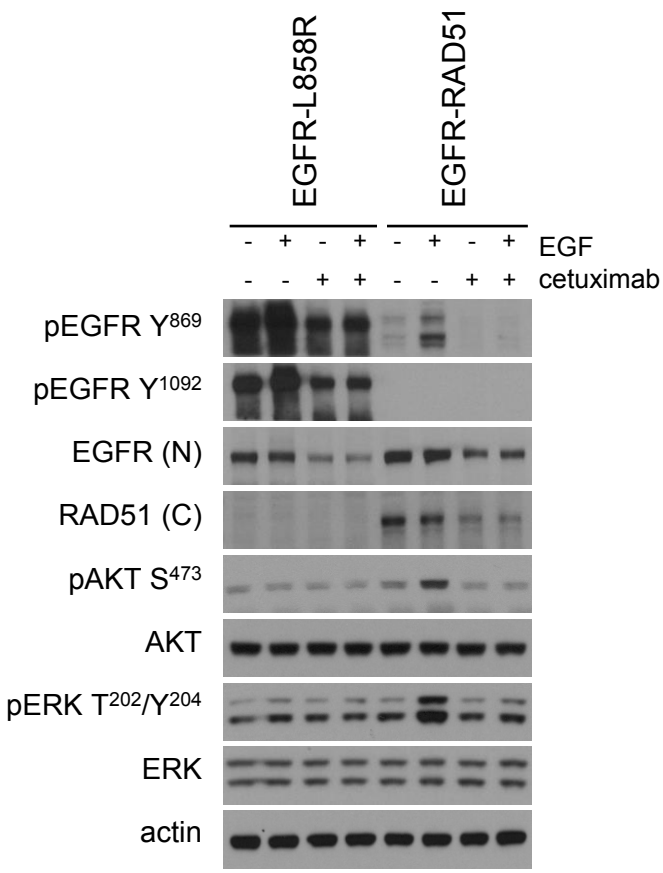


Figure S2.9: Cetuximab inhibits ligand-induced activation of downstream signaling pathways in cells expressing EGFR-RAD51.
Ba/F3 cells stably expressing EGFR-L858R and EGFR-RAD51 were serum starved for 16 hours, treated with 5 μg/mL cetuximab for 8 hours followed by 50 ng/mL EGF for 5 minutes, and subjected to western blot analysis with the indicated antibodies. This figure corresponds to Supplementary Figure 9 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

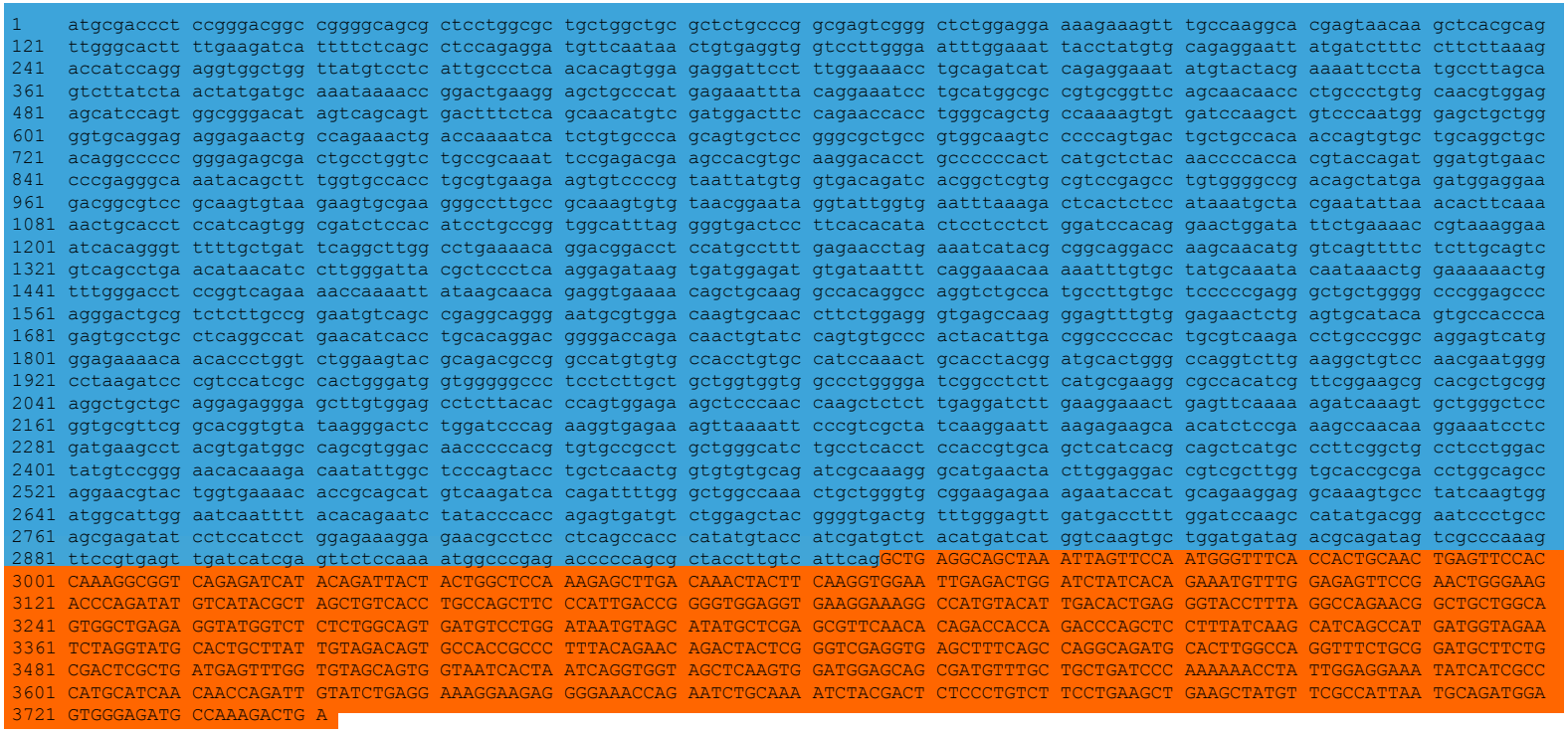
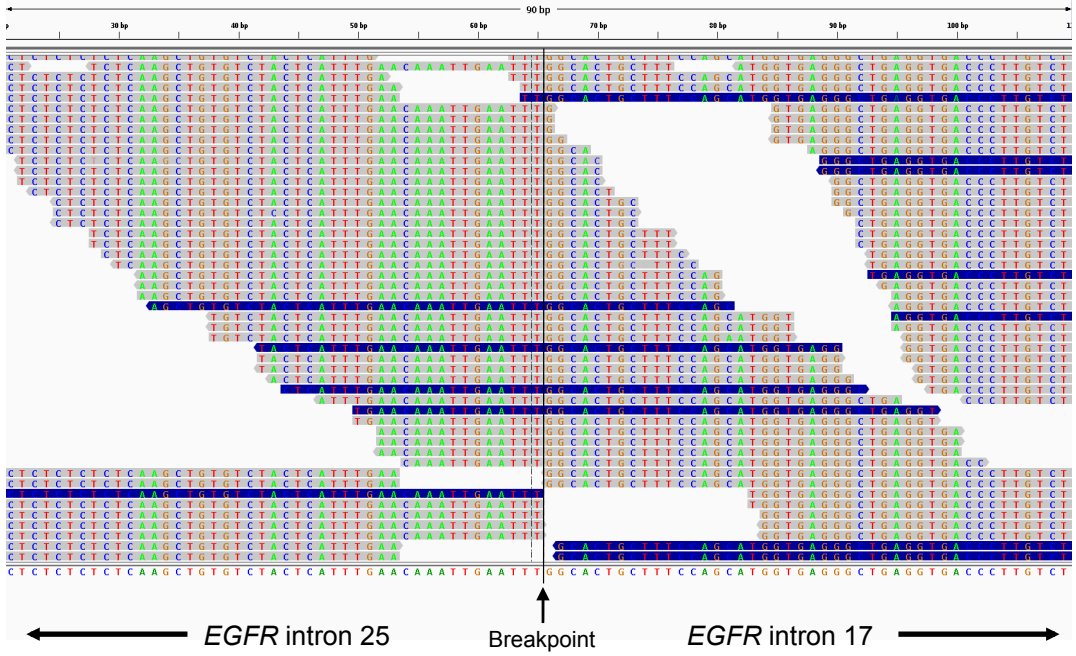


Figure S2.10: cDNA sequence of EGFR-RAD51.
Sequence of the EGFR-RAD51 fusion based on EGFR NM_005228 (ENST00000275493) and RAD51 NM_002875 (ENST00000267868). Exons 1-24 of EGFR are highlighted in blue and written in lower case. Exons 4-10 of RAD51 are highlighted in orange and written in capital letters. This figure corresponds to Supplementary Figure 10 from Konduri, Gallant et al., Cancer Discovery, 2016 (1).

A



B

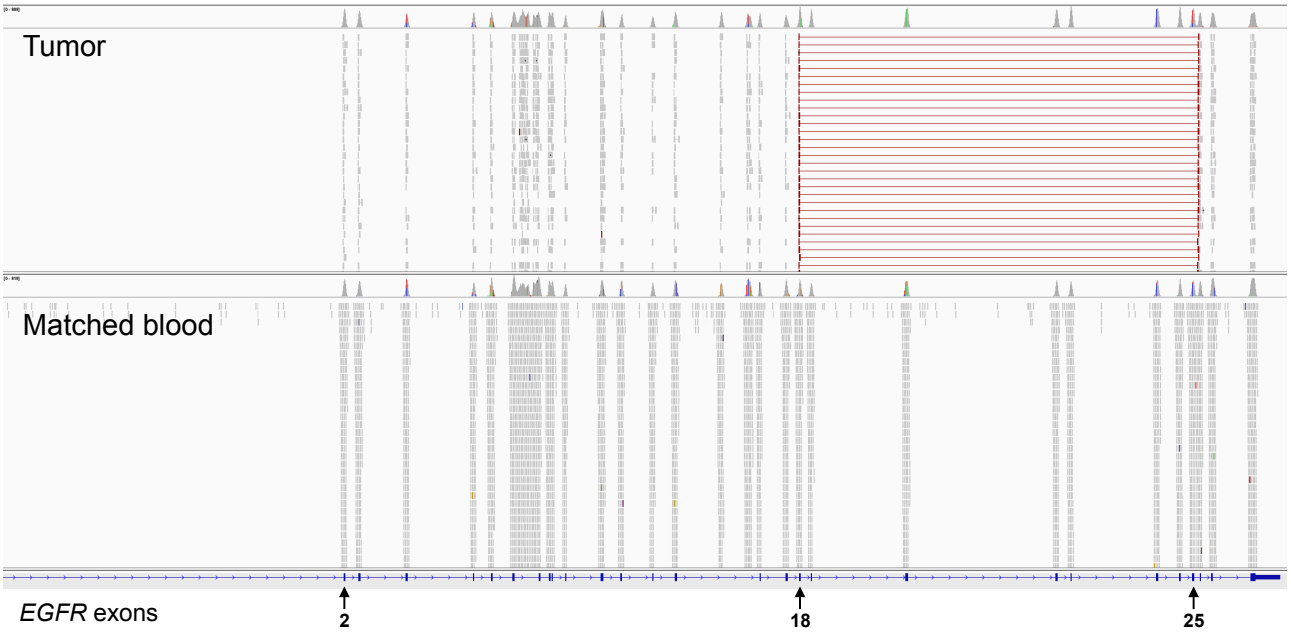


Figure S3.1: Sequencing reads of EGFR-KDD in index patient with lung adenocarcinoma.
A, NGS reads surrounding the genomic breakpoint of the EGFR-KDD as detected by the FoundationOne assay. **B,** paired NGS reads of tumor and matched whole blood from MSK-IMPACT™. This figure corresponds to Supplementary Figure 1 from Gallant et al. Cancer Discovery, 2015 (2).

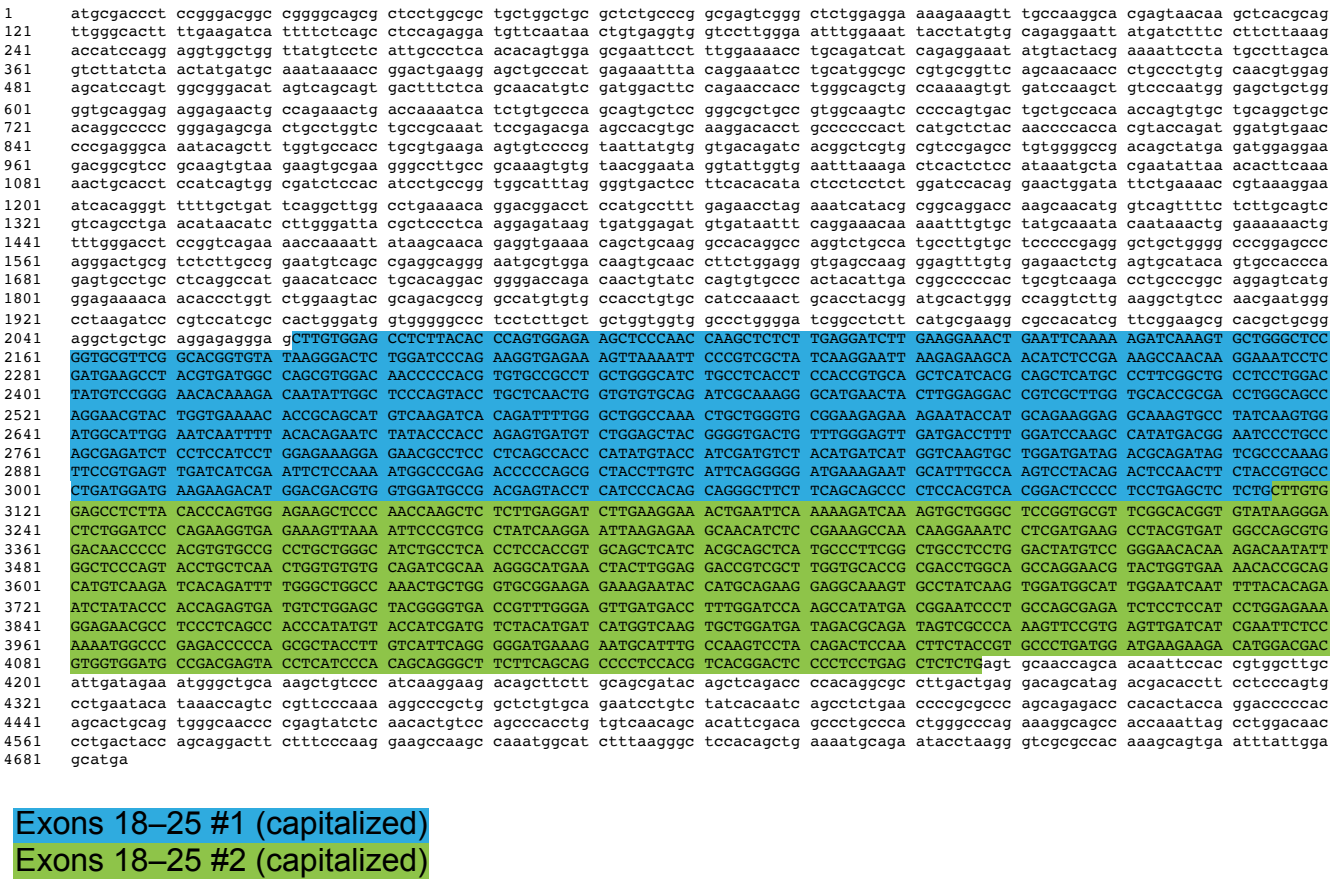


Figure S3.2: cDNA sequence of the EGFR-KDD.
Sequence of the EGFR exons 18-25 kinase domain duplication (EGFR-KDD). The first repeat of EGFR exons 18–25 is capitalized and highlighted with blue. The second (tandem) repeat of EGFR exons 18–25 is capitalized and highlighted with green. This figure corresponds to Supplementary Figure 2 from Gallant et al. Cancer Discovery, 2015 (2).

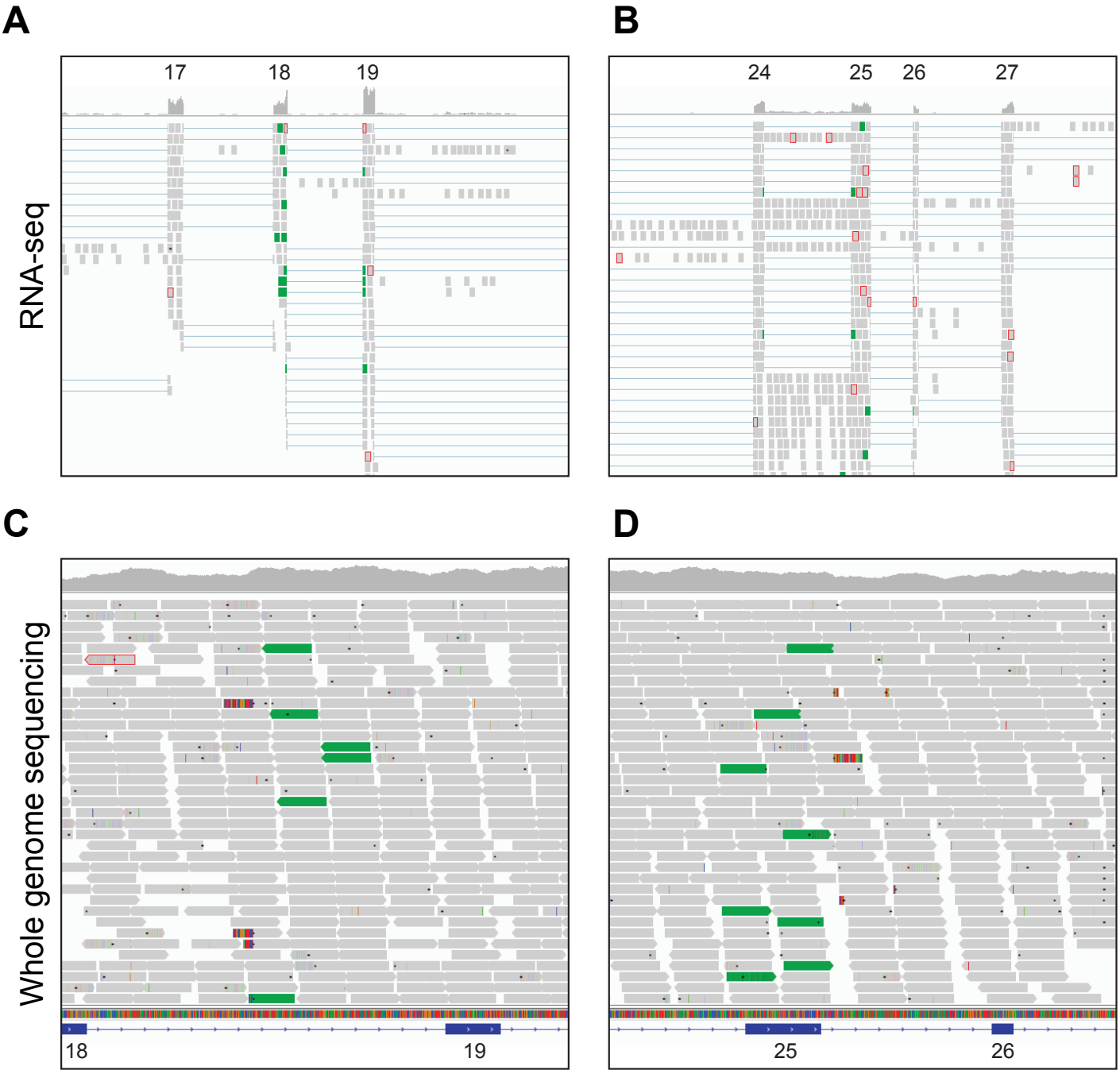


Figure S3.3: Sequencing reads identifying EGFR-KDD in a lung adenocarcinoma tumor from TCGA.
A, next-generation sequencing (NGS) reads of RNA-seq of EGFR exons 16–20 and **B,** 23–28 of TCGA-49-4512. **C,** whole-genome sequencing reads from exons 18–19 and **D,** 25–26 of TCGA-49-4512. Paired reads with a pattern of tandem duplication are pictured in green, and breakpoint-spanning (soft-clipped) reads are pictured as colored base pair mismatches. The boundaries of the EGFR-KDD are visible at the RNA level in EGFR exon 18 and exon 25 with corresponding intronic DNA breakpoints between exons 17 and 18 and exons 25 and 26. The data were visualized using the Integrative Genomics Viewer (3). This figure corresponds to Supplementary Figure 3 from Gallant et al. Cancer Discovery, 2015 (2).

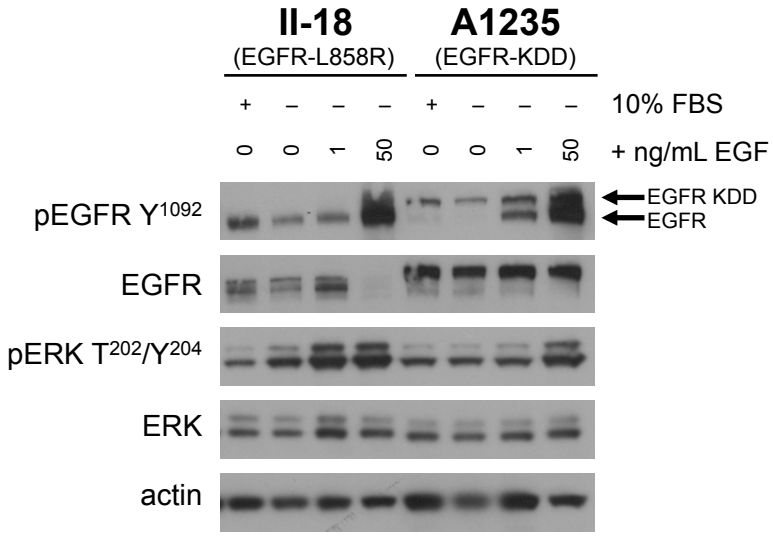


Figure S3.4: Autophosphorylation of endogenous EGFR-KDD in A1235 cells.
II-18 (EGFR-L858R) and A1235 (EGFR-KDD) cell lines were grown in the presence or absence of serum (FBS), or treated with 1 or 50 ng/mL EGF for 5 minutes after 16 hours of serum starvation, and then lysed for western blot analysis with the indicated antibodies. This figure corresponds to Supplementary Figure 4 from Gallant et al. Cancer Discovery, 2015 (2).

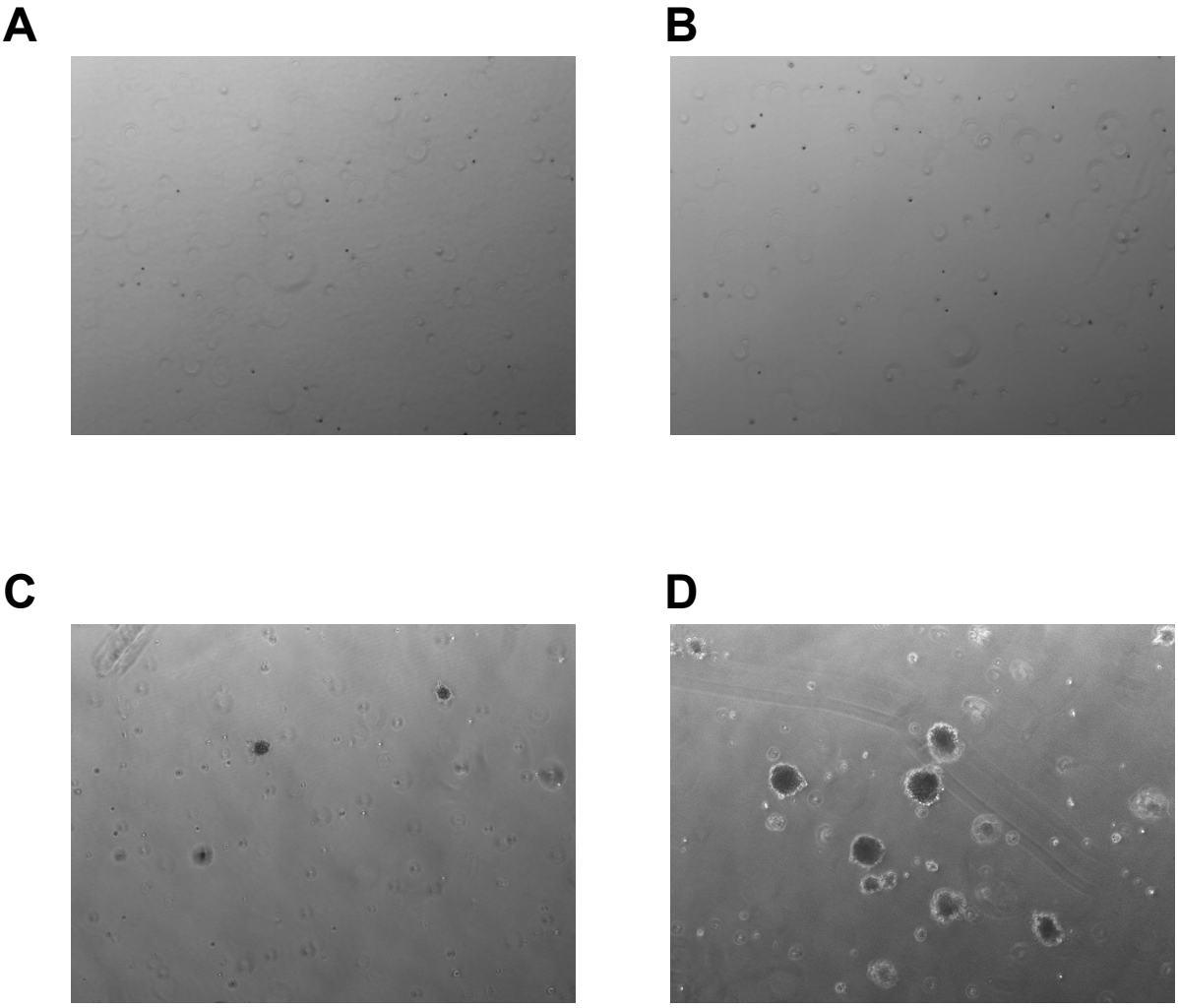


Figure S3.5: Colony formation of NR6 cells expressing EGFR variants.
A, representative 5x microscopy photographs of NR6 cells transduced with pMSCV vector, **B,** EGFR-WT, **C,** EGFR-L858R, or **D,** EGFR-KDD after 15 days of growth in soft agar. This figure corresponds to Supplementary Figure 5 from Gallant et al. Cancer Discovery, 2015 (2).

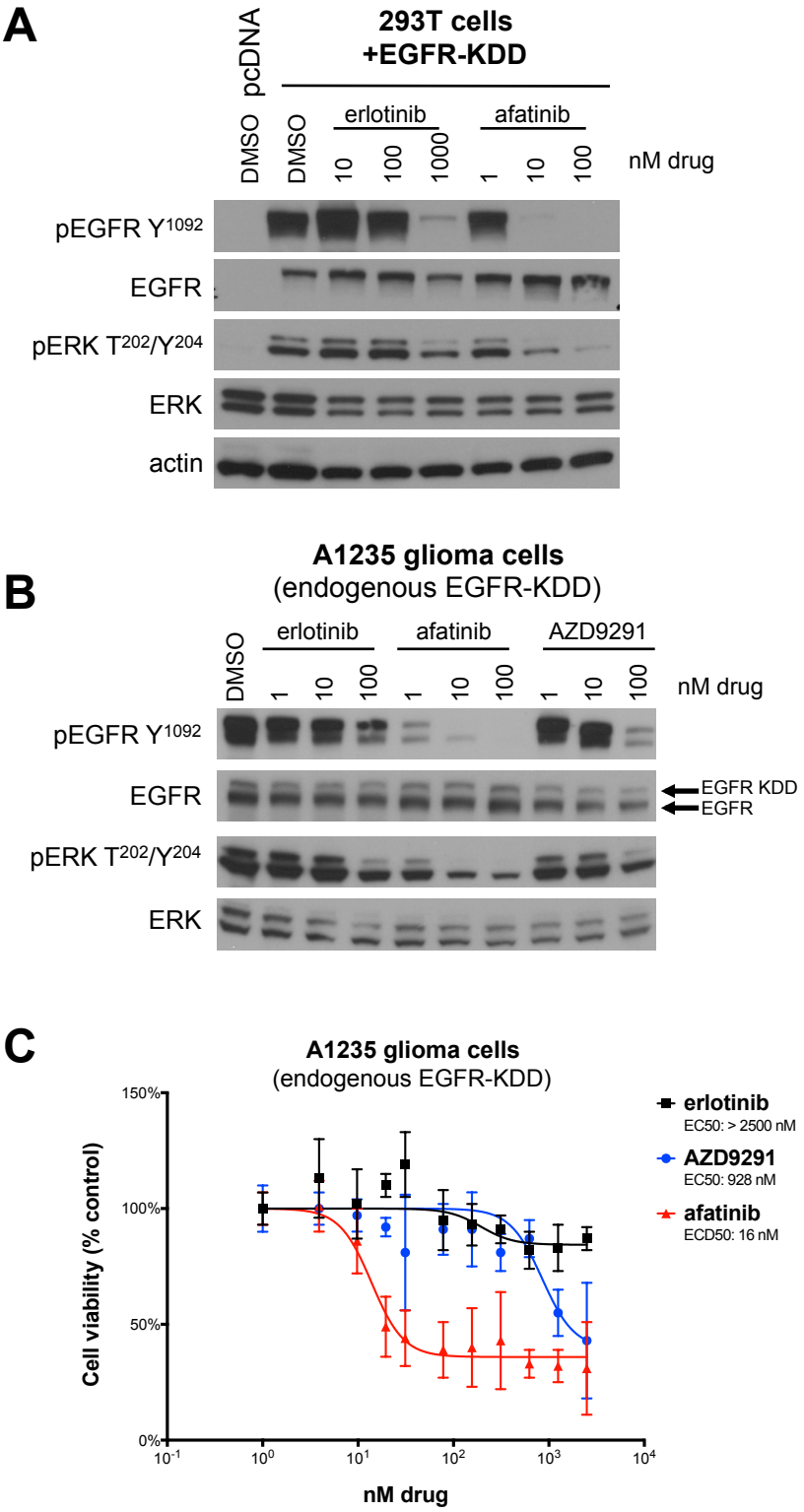


Figure S3.6: Efficacy of EGFR TKIs in endogenous and ectopic models of the EGFR-KDD.
A, 293T cells transfected with EGFR-KDD were treated with increasing doses of erlotinib or afatinib for 2 hours and lysed for western blot analysis. **B,** A1235 cells, which harbor endogenous EGFR-KDD, were treated with increasing doses of erlotinib, afatinib, or AZD9291 for 2 hours and lysed for western blot analysis. **C,** A1235 cells were treated with increasing doses of erlotinib, afatinib, or AZD9291 for 72 hours. Cell titer blue assays were performed to assess cell viability. Each point represents quadruple replicates. Data are presented as the percentage of viable cells compared to vehicle control. This figure corresponds to Supplementary Figure 6 from Gallant et al. Cancer Discovery, 2015 (2).

BA/F3 - EGFR-KDD

	erlotinib	afatinib	AZD9291
EC50 (nM)	109.5	3.1	28.5
Goodness of fit (R^2)	0.88	0.80	0.92

BA/F3 - EGFR-L858R

	erlotinib	afatinib	AZD9291
EC50 (nM)	6.1	0.02	4.0
Goodness of fit (R^2)	0.89	0.84	0.80

Table S3.1: Results of MTT curve fitting from Prism.
These data correspond to the graphs shown in Figure 3.2B. This table corresponds to Supplementary Table 1 from Gallant et al. Cancer Discovery, 2015 (2).

sgRNA cloning oligos		
Target and name	**Forward primer (5'→ 3')**	**Reverse primer (5'→ 3')**
EGFR gRNA#1	caccgGATAGGTGGATCGGCTGCCT	aaacAGGCAGCCGATCCACCTATCc
EGFR gRNA#2	caccgTGGCTAGCCGGACAGTGAGG	aaacCCACACTGTCCGGCTAGCCAc
RAD51 gRNA#1	caccgGACAATCACTTGAACCCGGA	aaacTCCGGGTTCAAGTGATTGTCc
RAD51 gRNA#2	caccgCAAAAATTAGCCACGTGTGG	aaacCCACACGTGGCTAATTTTTGc
EGFR-RAD51 gRNA#1	aaaggacgaaacaccgGATAGGTGGATCGGCTGCCTgttttagagctagaaatagcaag	ttctagctctaaaacTCCGGGTTCAAGTGATTGTCggatccaaggtgtctcatac
EGFR-RAD51 gRNA#2	aaaggacgaaacaccgTGGCTAGCCGGACAGTGAGGgttttagagctagaaatagcaag	ttctagctctaaaacCCACACGTGGCTAATTTTTGggatccaaggtgtctcatac
Primers to amplify targeted regions for SURVEYOR assay		
Target	**Forward primer (5'→ 3')**	**Reverse primer (5'→ 3')**
EGFR	GTTTTCCCAGTCACGACaattgcagtctgtgcttcca	CAGGAAACAGCTATGACctggtgcagagattttgagg
RAD51	GTTTTCCCAGTCACGACtctgttttcgctttttgggta	CAGGAAACAGCTATGACccatgggagctcaaggaata
Primers to amplify genomic breakpoint regions for RT-PCR		
EGFR-RAD51	GTTTTCCCAGTCACGACaattgcagtctgtgcttcca	CAGGAAACAGCTATGACccatgggagctcaaggaata
Primers to amplify genomic breakpoint regions for PCR and sequencing		
EGFR-RAD51	GTTTTCCCAGTCACGACgccacatcgtgaactaagca	CAGGAAACAGCTATGACactttgggaggctgaggtg

Table S4.1: Primers used for EGFR-RAD51 cell line engineering using the CRISPR-Cas9 system.
For sgRNA cloning oligos: cloning sites colored blue; U6 promoter enhancer colored green; sgRNA capitalized with PAM in red. For primers to amplify regions for SURVEYOR assay, RT/PCR, and sequencing: M13 sequencing primers are capitalized and the binding portion of the oligo is in lower case.

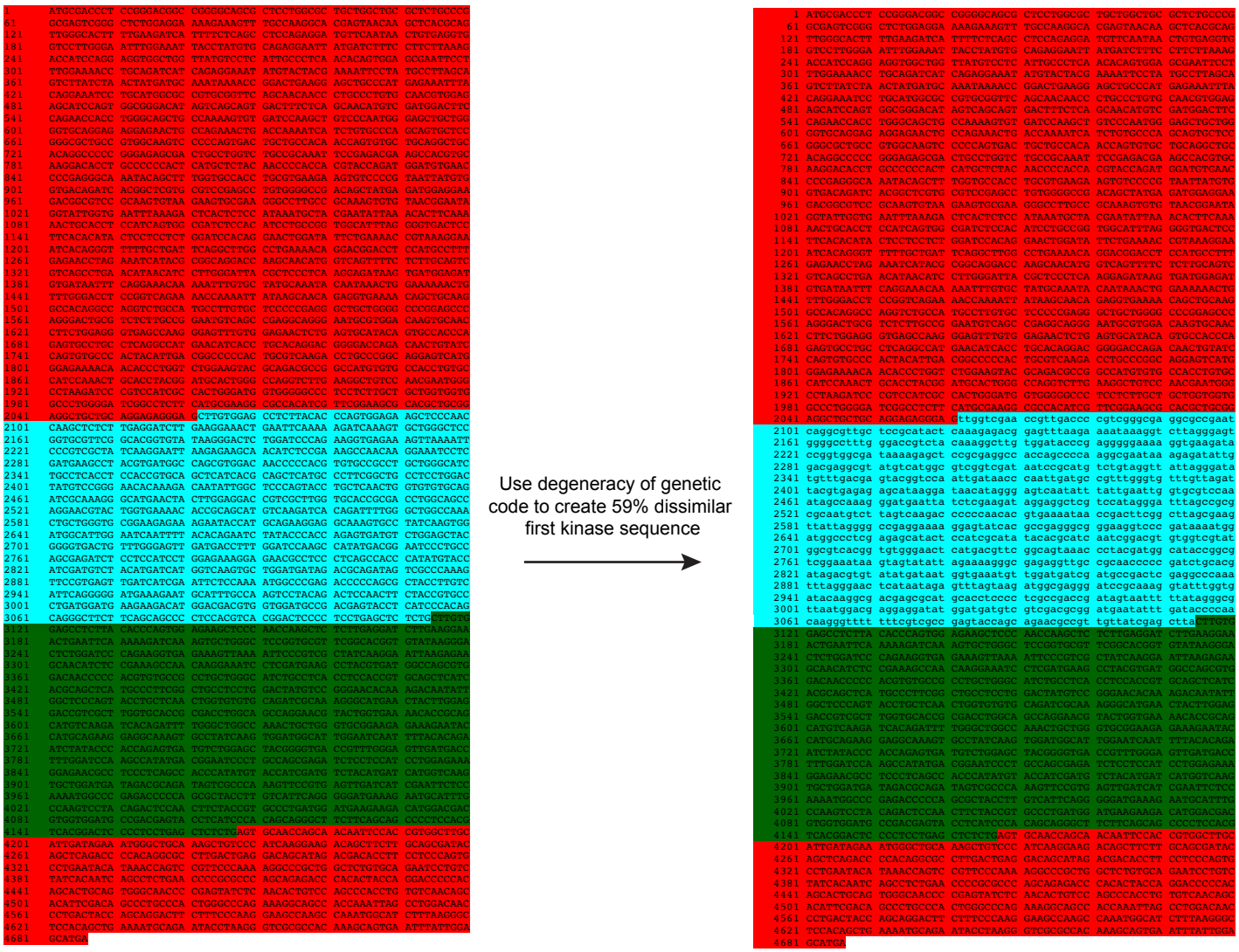


Figure S4.1: Comparison of EGFR-KDD sequences.
Left: published EGFR-KDD sequence (2). Right: EGFR-KDD with degenerate kinase sequences. Red: coding sequence outside of exons 18–25. Aqua: first exon 18–25 repeat; first kinase sequence. Dark green: second exons 18–25 repeat; second kinase sequence.

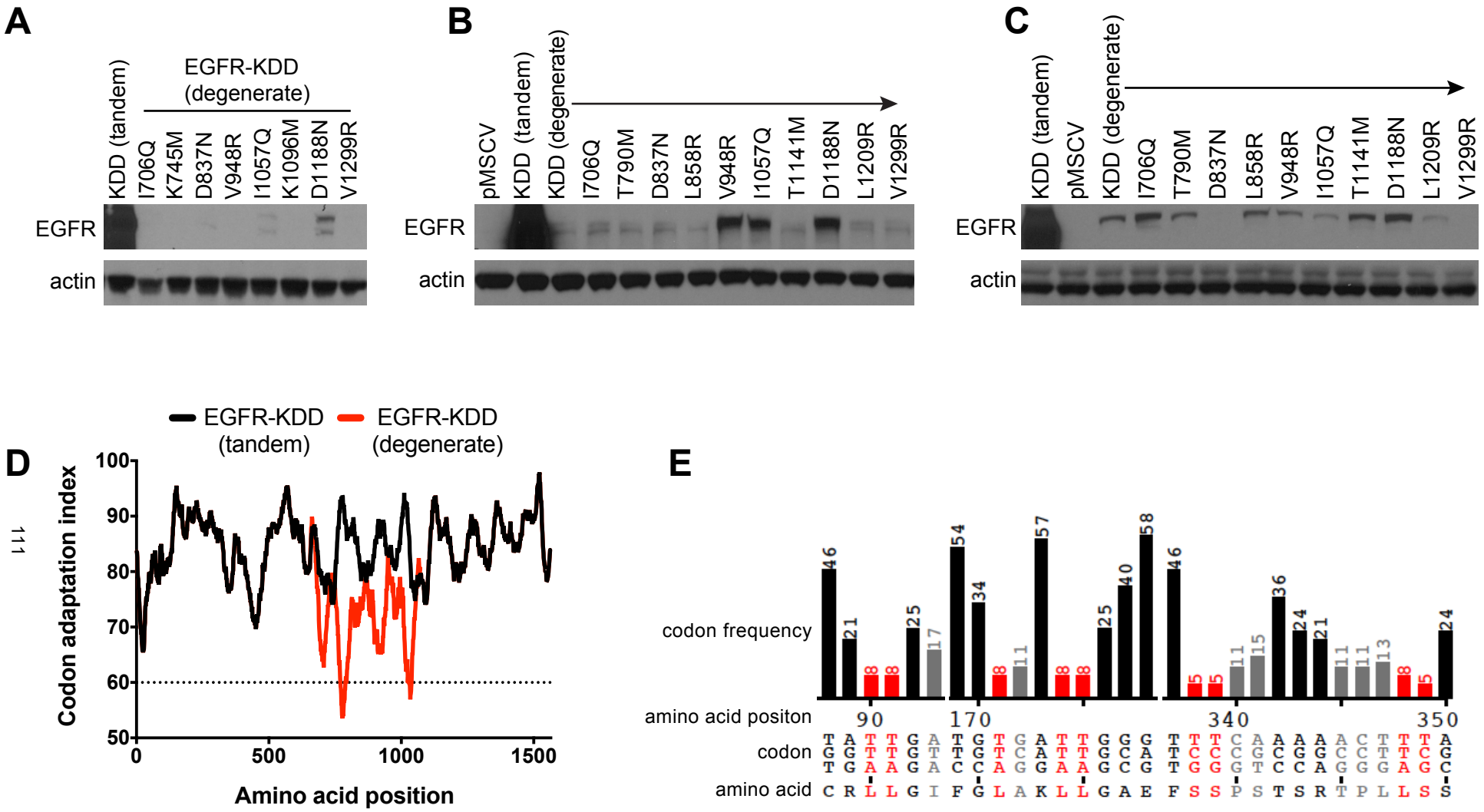


Figure S4.2: Lack of protein expression from degenerate EGFR-KDD constructs.
A, representative western blot (of N=4) of NR6 cells stably expressing EGFR-KDD and the indicated mutants. Tandem = identical kinase sequences (2); degenerate = two different kinase sequences. **B,** representative western blot (of N=4) of Ba/F3 cells stably expressing EGFR-KDD and the indicated mutants. Tandem = identical kinase sequences; degenerate = two different kinase sequences. **C,** western blot of CHO cells stably expressing EGFR-KDD and the indicated mutants. Tandem = identical kinase sequences; degenerate = two different kinase sequences. **D,** codon adaptation index of the tandem vs. degenerate EGFR-KDD sequences. Proteins with a codon adaptation > 80 should express; those with a codon adaptation below 60 (at any point) may not express (4). **E,** tandem rare codons through the sequence of the degenerate EGFR-KDD sequence. Tandem rare codons can cause ribosomal slippage and poor protein expression (5). The tandem EGFR-KDD had no tandem rare codons; the degenerate EGFR-KDD sequence had 4 pairs, as shown.

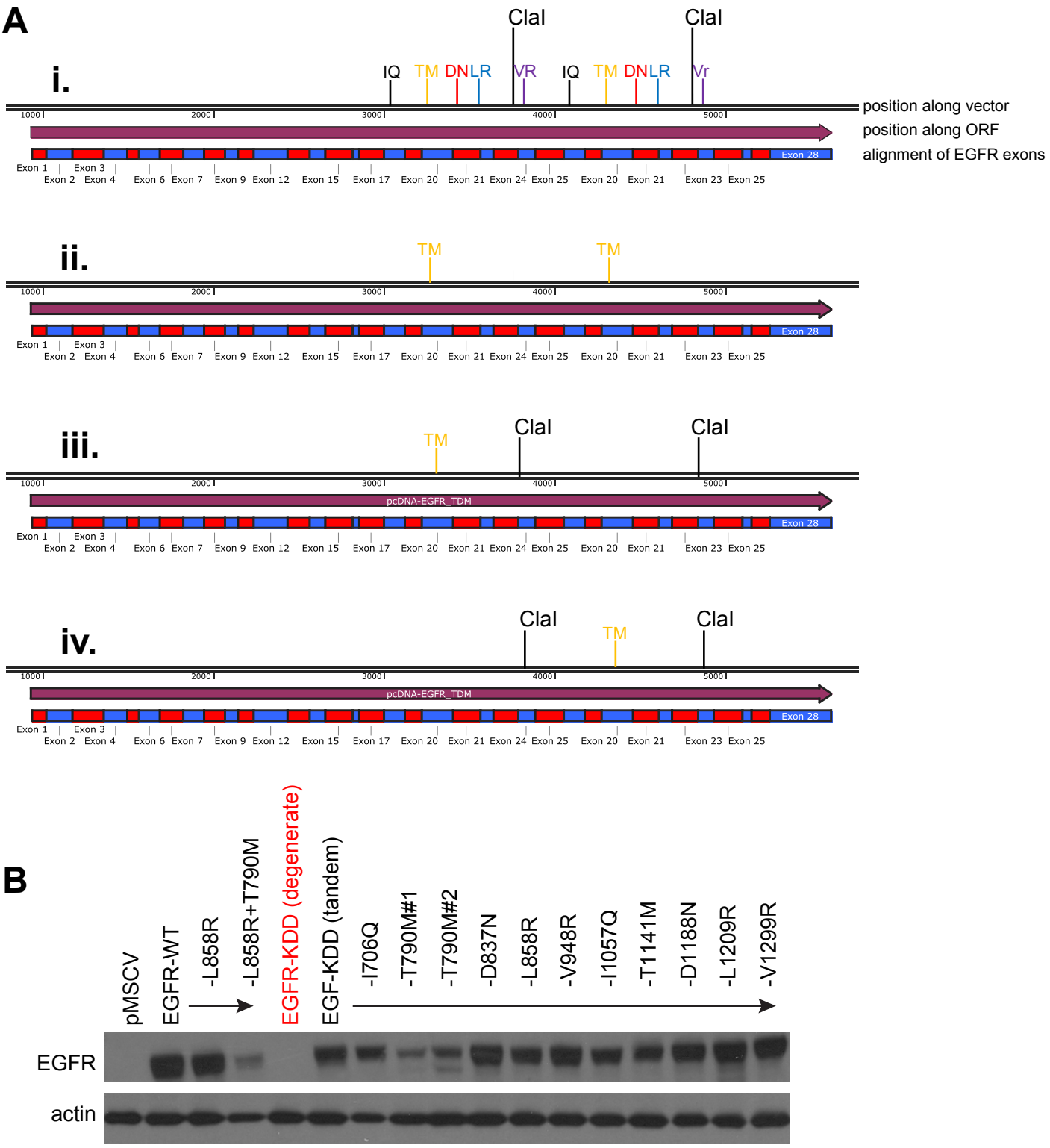


Figure S4.3: Creation of EGFR-KDD single mutants from a construct with two identical kinase domains. **A,** workflow for creation of single EGFR-KDD mutants. i, overview of the construct demonstrating the location of each mutation and the ClaI restriction sites. ii, sample multi-site directed mutagenesis, here for tandem TM (threonine to methionine; T790M and T1141M) mutations. iii, ClaI digest of TM mutated EGFR-KDD plasmid is recombined with a ClaI–ClaI segment from unmutated EGFR-KDD. iv, ClaI digest of unmutated EGFR-KDD plasmid is recombined with a ClaI–ClaI segment from TM mutated EGFR-KDD. **B,** Ba/F3 cells lines stably expressing EGFR constructs. Note the difference in expression levels between the EGFR-KDD (degenerate) construct and the EGFR-KDD tandem (two identical kinase domains) construct.

Mutation	Forward primer (5'→3')	Reverse primer (5'→3')
I706Q	cccaaccaagctctcttgaggcagttgaaggaaactgagttcaaa	tttgaactcagtttccttcaactgcctcaagagagcttggttggg
K745M	aaagttaaaattcccgtcgctatcatggaattaagagaagcaac	gttgcttctcttaattccatgatagcgacgggaattttaacttt
D837N	cttggtgcaccgcaacctggcagccag	ctggctgccaggttgcggtgcaccaag
Y869F	gctgggtgcggaagagaaagaattccatgcagaagg	ccttctgcatggaattctttctcttccgcacccagc
Y915F	tgacctttggatccaagccatttgacggaatccctg	cagggattccgtcaaatggcttggatccaaaggtca
Y944F	atatgtaccatcgatgtcttcatgatcatggtcaagtgc	gcacttgaccatgatcatgaagacatcgatggtacatat
V948R	cgatgtctacatgatcatgcgcaagtgctggatgatagac	gtctatcatccagcacttgcgcatgatcatgtagacatcg
Y978F	cgagacccccagcgcttccttgtcattcag	ctgaatgacaaggaagcgctgggggtctcg
F993E	ctgaggcagctaaattagttccaatgggtgagaccactgcaactgag	ctcagttgcagtggtctcacccattggaactaatttagctgcctcag
F993R	ctaaattagttccaatgggtcgcaccactgcaactgagttcc	ggaactcagttgcagtggtgcgacccattggaactaatttag
A996W	ttagttccaatgggtttcaccacttggactgagttccaccaaagg	cctttggtggaactcagtccaagtggtgaaacccattggaactaa
A996R	tccaatgggtttcaccactagaactgagttccaccaaagg	cctttggtggaactcagttctagtggtgaaacccattgga
F993R+A996R	gctaaattagttccaatgggtcgcaccactagaactgagttccaccaaaggcg	cgcctttggtggaactcagttctagtggtgcgacccattggaactaatttagc
A1097W	tggcagtgatgtcctggataatgtatggtatgctcgagcgttcaa	ttgaacgctcgagcataccatacattatccaggacatcactgcca
A1099W	gcagtgatgtcctggataatgtagcatattggcgagcgttcaacaca	tgtgttgaacgctcgccaatatgctacattatccaggacatcactgc
A1097W+A1099W	tctctctggcagtgatgtcctggataatgtatggtattggcgagcgttcaacacagaccac	gtggtctgtgttgaacgctcgccaataccatacattatccaggacatcactgccagagaga
Y1222F	ggaaaccagaatctgcaaaatcttcgactctccctg	cagggagagtcgaagattttgcagattctggtttcc
Δint24+	ccagcgctaccttgtcatttaggctgaggc	gcctcagcctaaatgacaaggtagcgctgg
stop/tag cloning	gagtgggagatgccaaagacagaattcattaattaactgg	ccagttattaatgaattctgtctttggcatctcccactc

Table S4.2: Site-directed mutagenesis primers for use with EGFR-RAD51.

Mutation	Forward primer (5'→3')	Reverse primer (5'→3')
Site-directed mutagenesis primers for use with multi-site directed mutagenesis		
IQ	cccaaccaagctctcttgaggcagttgaaggaaactgagttcaaa	tttgaactcagtttccttcaactgcctcaagagagcttggttggg
KM	aaagttaaaattcccgtcgctatcatggaattaagagaagcaac	gttgcttctcttaattccatgatagcgacgggaattttaacttt
TM	ccaccgtgcagctcatcatgcagctcatgc	gcatgagctgcatgatgagctgcacggtgg
DN	cttggtgcaccgcaacctggcagccag	ctggctgccaggttgcggtgcaccaag
LR	tcacagattttgggcgggccaaactgctggg	cccagcagtttggcccgcccaaaatctgtga
VR	cgatgtctacatgatcatgcgcaagtgctggatgatagac	gtctatcatccagcacttgcgcatgatcatgtagacatcg
Site-directed mutagenesis primers for use with EGFR-KDD degenerate construct		
I706Q	cgaatcaggcgttgctccgccagctcaaagagacggagtttaa	ttaaactccgtctctttgagctggcggagcaacgcctgattcg
K745M	gtgaagataccggtggcgataatggagctccgcgag	ctcgcggagctccattatcgccaccggtatcttcac
T790M	gagtacggtccaattgataatgcaattgatgccgtttgggtg	cacccaaacggcatcaattgcattatcaattggaccgtactc
D837N	ggctcgtccataggaatttagccgcgcgc	gcgcgcggctaaattcctatggacgagcc
L858R	gaaaataaccgacttcggcagagcgaagttattaggggcc	ggcccctaataacttcgctctgccgaagtcggttattttc
V948R	cgatagacgtgtatatgataatgaggaaatgttggatgatcgatgccg	cggcatcgatcatccaacatttcctcattatcatatacacgtctatcg
I1057Q	cccaaccaagctctcttgaggcagttgaaggaaactgagttcaaa	tttgaactcagtttccttcaactgcctcaagagagcttggttggg
K1096M	aaagttaaaattcccgtcgctatcatggaattaagagaagcaac	gttgcttctcttaattccatgatagcgacgggaattttaacttt
T1141M	ccaccgtgcagctcatcatgcagctcatgc	gcatgagctgcatgatgagctgcacggtgg
D1188N	cttggtgcaccgcaacctggcagccag	ctggctgccaggttgcggtgcaccaag
L1209R	tcacagattttgggcgggccaaactgctggg	cccagcagtttggcccgcccaaaatctgtga
V1299R	cgatgtctacatgatcatgcgcaagtgctggatgatagac	gtctatcatccagcacttgcgcatgatcatgtagacatcg

Table S4.3: Site-directed mutagenesis primers for use with EGFR-KDD constructs.

References

1. Konduri, K. *et al.* EGFR Fusions as Novel Therapeutic Targets in Lung Cancer. *Cancer Discovery* **6,** 601–611 (2016).

2. Gallant, J. N. *et al.* EGFR Kinase Domain Duplication (EGFR-KDD) Is a Novel Oncogenic Driver in Lung Cancer That Is Clinically Responsive to Afatinib. *Cancer Discovery* **5,** 1155–1163 (2015).

3. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29,** 24–26 (2011).

4. Sharp, P. M. & Li, W. H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15,** 1281–1295 (1987).

5. Brar, G. A. Beyond the Triplet Code: Context Cues Transform Translation. *Cell* **167,** 1681–1692 (2016).