JCI The Journal of Clinical Investigation

Somatic rearrangements causing oncogenic ectodomain deletions of FGFR1 in squamous cell lung cancer

Florian Malchers, … , Julie George, Roman K. Thomas

J Clin Invest. 2023. https://doi.org/10.1172/JCI170217.

Research In-Press Preview **Genetics** **Oncology**

The discovery of frequent 8p11-p12 amplifications in squamous cell lung cancer has fueled hopes that FGFR1, located inside this amplicon, might be a therapeutic target. In a clinical trial, only 11% of patients with 8p11 amplification (detected by FISH) responded to FGFR kinase inhibitor treatment. To understand the mechanism of FGFR1 dependency, we performed deep genomic characterization of 52 squamous cell lung carcinomas with 8p11-p12-amplification, including 10 tumors obtained from patients who had been treated with FGFR inhibitors. We discovered somatically altered variants of *FGFR1* with deletion of exons 1-8 that resulted from intragenic tail-to-tail rearrangements. These ectodomain-deficient FGFR1 variants (ΔEC-FGFR1) were expressed in the affected tumors and tumorigenic in in-vitro and in-vivo. Mechanistically, *Breakage-Fusion-Bridges* were the source of 8p11-p12 amplification, resulting from frequent head-to-head and tail-to-tail rearrangements. However, only tumors with tail-to-tail rearrangements within or in close proximity upstream of *FGFR1* exhibited FGFR1 dependency. Thus, the genomic events shaping the architecture of the 8p11-p12 amplicon provide a mechanistic explanation for the emergence of FGFR1-driven squamous cell lung cancer. Specifically, FGFR1 ectodomain deficient and *FGFR1*-centered amplifications caused by tail-to-tail rearrangements are novel somatic genomic event, which might be predictive of therapeutically relevant FGFR1 dependency.

Somatic rearrangements causing oncogenic ectodomain deletions of FGFR1 in squamous cell lung cancer

Authors: Florian Malchers[1, †], Lucia Nogova[2], Martijn H. A. van Attekum[1], Lukas Maas[1], Johannes Brägelmann[1,3,4], Christoph Bartenhagen[5], Luc Girard[6], Graziella Bosco[1], Ilona Dahmen[1], Sebastian Michels[2], Clare E. Weeden[7], Andreas Scheel[4], Lydia Meder[2], Kristina Golfmann[2], Philipp Schuldt[2], Janna Siemanowski[4], Jan Rehker[4], Sabine Merkelbach-Bruse[4], Roopika Menon[9], Oliver Gautschi[10], Johannes M. Heuckmann[9], Elisabeth Brambilla[11], Marie-Liesse Asselin-Labat[7], Thorsten Persigehl[12], John D. Minna[6], Henning Walczak[13, †], Roland T. Ullrich[2], Matthias Fischer[5], Hans Christian Reinhardt[14], Jürgen Wolf[2], Reinhard Büttner[4], Martin Peifer[1,15], Julie George[1,8], and Roman K. Thomas[1,4,16, ‡]

Affiliations:

[1] Department of Translational Genomics, Medical Faculty; University of Cologne, Weyertal 115b, 50931 Cologne, Germany

[†] present address: Institute of Biochemistry I, Medical Faculty; Joseph-Stelzmann-Str. 44, University of Cologne, 50931 Cologne, Germany & CECAD Cluster of Excellence, University of Cologne, 50931 Cologne, Germany

[2] Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn, Cologne Duesseldorf, University of Cologne, Germany

[3] Mildred Scheel School of Oncology, Cologne; University Hospital Cologne, Medical Faculty, 50931 Cologne, Germany

[4] Institute of Pathology, Medical Faculty; University of Cologne, Kerpener Str. 62, 50924 Cologne, Germany

[5] Department of Experimental Pediatric Oncology, University Children's Hospital of Cologne; University Hospital Cologne, Medical Faculty, 50931 Cologne, Germany

[6] University of Texas Southwestern Medical Center; 6000 Harry Hines Blvd, Dallas, TX 75390-8593, USA

[7] Personalized Oncology Division; Walter and Eliza Hall Institute of Medical Research, Department of Medical Biology, The University of Melbourne, Parkville, Australia

[8] Department of Head and Neck Surgery, Medical Faculty, University Hospital Cologne, 50924 Cologne, Germany

[9] DISCO Pharmaceuticals GmbH, Gottfried-Hagen Str. 60-62, 51105 Köln

[10] University of Berne and Cantonal Hospital of Lucerne; Cantonal Hospital of Lucerne, Switzerland

[11] Département d'Anatomie et Cytologie Pathologiques; 38043 Grenoble Cedex 09, France

[12] Institute for Diagnostic and Interventional Radiology, Faculty of Medicine and University Hospital Cologne; University Cologne, Kerpener Straße 62, 50937, Cologne, Germany

[13] Centre for Cell Death, Cancer, and Inflammation (CCCI), UCL Cancer Institute; University College London, 72 Huntley Street, London WC1E 6DD, UK

[14] Department of Hematology and Stem Cell Transplantation; University Hospital Essen, University Duisburg-Essen

[15] Center for Molecular Medicine Cologne (CMMC); Weyertal 115b, 50931 Cologne, Germany

[16] German Cancer Consortium (DKTK), partner site Heidelberg and German Cancer Research Center (DKFZ); Heidelberg, Germany

[‡] **Corresponding Author:** Roman K. Thomas, Department of Translational Genomics; University of Cologne, Weyertal 115b, 50931 Cologne, Germany, roman.thomas@uni-koeln.de

Conflict-of-interest: **LN** has received honoraria from Pfizer, Celgene, Novartis, Roche, Boehringer Ingelheim, Bristol-Myers Squibb, Takeda, Bayer, Janssen and **JS** has received honoraria from MDK, Biocartis and AstraZeneca. **SM-B** has received honoraria from AstraZeneca, Roche, Novartis, GSK, MSD, Targos, Molecular Health, Merck, personal fees and non-financial support from Janssen and BMS. **HCR** received consulting and lecture fees from Abbvie, AstraZeneca, Vertex and Merck. **H.C.R.** received research funding from Gilead Pharmaceuticals. **H.C.R.** is a co-founder of CDL Therapeutics GmbH. **RB** has received honoraria from AbbVie, Amgen, AstraZeneca, Bayer, BMS, Boehringer-Ingelheim, Illumina, Janssen, Lilly, Merck-Serono, MSD, Novartis, Qiagen, Pfizer, Roche, Targos MP Inc. **RKT** is a founder of NEO New Oncology, now part of Siemens Healthcare, a founder of PearlRiver Bio, now part of Centessa

Pharmaceuticals, a founder of Epiphanes Inc and a founder and CEO of DISCO Pharmaceuticals. **RKT** is a shareholder of Centessa, of Epiphanes and of DISCO Pharmaceuticals and has received consulting fees/ salary from NEO New Oncology, PearlRiver Bio and DISCO Pharmaceuticals. The remaining authors declare no competing financial interest.

ABSTRACT

The discovery of frequent 8p11-p12 amplifications in squamous cell lung cancer has fueled hopes that FGFR1, located inside this amplicon, might be a therapeutic target. In a clinical trial, only 11% of patients with 8p11 amplification (detected by FISH) responded to FGFR kinase inhibitor treatment. To understand the mechanism of FGFR1 dependency, we performed deep genomic characterization of 52 squamous cell lung carcinomas with 8p11-p12-amplification, including 10 tumors obtained from patients who had been treated with FGFR inhibitors. We discovered somatically altered variants of *FGFR1* with deletion of exons 1-8 that resulted from intragenic tail-to-tail rearrangements. These ectodomain-deficient FGFR1 variants (ΔEC-FGFR1) were expressed in the affected tumors and tumorigenic in both in-vitro and in-vivo models of lung cancer. Mechanistically, *Breakage-Fusion-Bridges* were the source of 8p11-p12 amplification, resulting from frequent head-to-head and tail-to-tail rearrangements. However, only tumors with tail-to-tail rearrangements within or in close proximity upstream of *FGFR1* exhibited FGFR1 dependency. Thus, the genomic events shaping the architecture of the 8p11-p12 amplicon provide a mechanistic explanation for the emergence of FGFR1-driven squamous cell lung cancer. Specifically, FGFR1 ectodomain deficient and *FGFR1*-centered amplifications caused by tail-to-tail rearrangements are a novel somatic genomic event that might be predictive of therapeutically relevant FGFR1 dependency.

INTRODUCTION

Squamous cell lung cancer is the second most common lung cancer subtype and among the cancers with the worst prognosis [1]. Unfortunately, genetically activated kinase targets that provide opportunities for effective treatment, such as those occurring in lung adenocarcinoma (e.g., *EGFR* mutations or *ALK* rearrangements) have so far not been identified in squamous cell lung cancer [2]. The discovery of recurrent *FGFR1* amplifications (located on 8p11.23) had raised hopes that patients with such amplification might benefit from FGFR inhibition in this hard-to-treat cancer [3, 4]. In a clinical trial testing this hypothesis by treating patients with 8p-amplified squamous-cell lung cancer with the FGFR inhibitor, BGJ398, approx. 11% of 8p11-p12-amplified tumors exhibited durable responses to single-agent FGFR inhibition [5]. Thus, while the majority of these tumors are not dependent on FGFR, a minority require the catalytical activity of the kinase for their survival. The heterogeneous 8p11-8p12 amplification pattern might explain this observation to some extent, and a recent publication highlighted the role of *NSD3* as an oncogenic driver in squamous cell lung cancer [6, 7]; however, a mechanistic reason for this phenotypic heterogeneity has so far been lacking. We therefore hypothesized that specific structural features of the 8p11 amplicon might explain FGFR1 dependency and performed an in-depth genomic and functional study of primary human squamous cell lung carcinomas, from both patients treated with FGFR inhibitors as well as untreated patients, cancer cell lines and patient-derived xenotransplants.

RESULTS

Tumors of lung cancer patients with a response to FGFR inhibition exhibit tail-to-tail rearrangements within *FGFR1*

We first collected 52 squamous cell lung carcinomas which were positive for *FGFR1*-amplification, tested by fluorescence in situ hybridization (FISH) or Affymetrix SNP 6.0 arrays, and performed deep genomic sequencing (**Fig. 1a**). We first analyzed treatment naive biopsy specimens obtained from 10 patients with *FGFR1*-amplified squamous cell lung cancer (**Fig. 1a, b**). Nine of these patients had been treated with the FGFR inhibitor, BGJ398 as part of a phase-I clinical trial (NCT01004224) and one patient had been treated with pazopanib, a multi-kinase inhibitor that also inhibits FGFR [5, 6]. Of these 10 patients, 4 had experienced a partial response (PR, defined as tumor shrinkage of at least 30% of the maximal tumor diameter) lasting for 9 to

17 months [5], and 6 patients had progressive disease (PD) under treatment (**Fig. 1a-e, Suppl. Fig. 1**). Given the low amount of available DNA obtained from these clinical specimens, we employed a hybrid capture-based sequencing assay tailored to cover much of the genomic 8p11-8p12 locus, as well as 226 additional genes [6]. We sequenced the 10 specimens with an average sequencing depth of 470x.

Although all specimens had been selected based on the presence of *FGFR1* amplification determined by FISH as part of the inclusion criteria (**Fig. 1f, Suppl. Fig. 2, 3**), we could not detect *FGFR1* amplification by this sequencing assay in all tumors (**Fig. 1f**). In particular, we did not observe *FGFR1* amplification in the specimen TUM009 (compare **Fig. 1f, Suppl. Fig. 3**). Thus, *FGFR1* amplification was insufficient to predict response to FGFR inhibition in this clinical cohort. However, to our surprise, in two specimens obtained from patients, who had responded to FGFR inhibition, we found highly covered tail-to-tail rearrangements within *FGFR1*, which induced an abrupt copy number change within the open reading frame (ORF) of *FGFR1* (TUM004 and TUM006 with 192 and 115 break detecting reads, respectively) (**Fig. 1g, h**). In both cases, rearrangements were caused by two chromosomal breaks in close proximity, which led to an intrachromosomal fusion in a tail-to-tail manner, and induced a loss of the first canonical *FGFR1* exons (**Fig. 1i, Suppl. Video 1**). Upstream-located enhancer- or silencer-sequences and the complete promotor region of *FGFR1*, including the TATA-box and transcription initiation region were deleted, suggesting they might cause a functional deletion within FGFR1. In detail, the rearrangement in the sample TUM004 deleted upstream DNA sequences including exon 1 of *FGFR1*, which is coding for the 5' untranslated region (UTR). The canonical ATG start codon, located in exon 2, was unaffected and allowed the translation of an in principle intact, full-length FGFR1 protein (**Fig. 1g and i**). The rearrangement in the sample TUM006 deleted upstream DNA sequences until exon 3 of *FGFR1*, and therefore lacked the canonical ATG start codon of *FGFR1*, providing a non-canonical start codon in exon 5 (**Fig. 1h and i**). These specific/intrachromosomal tail-to-tail rearrangements were exclusively identified in patients who had responded to FGFR inhibition. Furthermore, we observed robust staining for phosphorylated FGFR1 (pFGFR1) in immunohistochemistry analyses particular in the sample TUM006, indicating active transcription and catalytic activity of FGFR1 (**Fig. 1h right**). In this case, *FGFR1* can only be translated by using the first downstream in-frame ATG start codon located at exon 5, which would delete the immunoglobulin-like domain I and the acid box of FGFR1, which is known for its self-inhibitory

function [8]. On the contrary, tumors from patients without clinical response to FGFR inhibition never showed such rearrangements and instead exhibited three different *PIK3CA* mutations (G118S, E545K and H1047R, 3 out of 6 patients). In summary, half of patients with a response to FGFR inhibition had *FGFR1* amplification, which was caused by tail-to-tail rearrangements within the open reading frame (ORF) of *FGFR1,* leading to retained protein expression and catalytic function of FGFR1.

Intragenic somatic rearrangements causing ectodomain deletions of FGFR1 in primary squamous cell lung cancers

We next sought to validate our finding of ectodomain deletions in independent cohorts. Therefore, we performed an in-depth genomic characterization of 16 *FGFR1*-amplified cell lines and PDX models with known in vitro and in vivo response to FGFR-inhibition, and 26 additional *FGFR1*-amplified squamous cell lung carcinomas from patients with unknown response to FGFR-inhibition (**Fig. 1a**). To our disappointment, we could not detect similar rearrangements in any of the cell lines or PDX models to validate our findings from patients that had been treated with FGFR inhibitors. However, in the cohort of *FGFR1*-amplified squamous cell lung carcinomas with unknown response to FGFR-inhibition, we detected two patients with similar rearrangements (S00674 and A921). In the primary tumor S00674 we found a tail-to-tail rearrangement 150 kb upstream of the transcriptional start site of *FGFR1*. This rearrangement was again caused by two breaks in close proximity and led to an intrachromosomal fusion in a tail-to-tail manner (**Fig. 2a**). A second tail-to-tail rearrangement within the open reading frame (ORF) of *FGFR1* translocated closely to the transcriptional start site of *NSD3* (**Fig. 2a, b**), similar to the rearrangements observed in the two patients of the clinical trial who exhibited a response to FGFR inhibition (**Fig. 1g-i**). The tail-to-tail rearrangement upstream of *FGFR1* had a 3-fold higher coverage of reads across the breakpoint compared to the rearrangement within *FGFR1* (54 vs. 15 break-detecting reads), suggesting the second event had occurred later in cancer genome evolution. The rearrangement in S00674 might be suggestive of inactivation of *FGFR1* and might also have a particular relevance of *NSD3*. However, transcriptome sequencing revealed increased levels of *FGFR1* transcription compared to $NSD3_{long}$ (**Fig. 2c**)**.** Furthermore, we found active transcription across the breakpoint affecting the *FGFR1* gene (**Fig. 2b, d**). In detail, the tail-to-tail rearrangement was an intra-chromosomal event caused by two breaks within *FGFR1* (inside exon 6) and a non-coding region

close to the transcriptional start site of *NSD3* (**Fig. 2b**). The break led to deletion of *FGFR1* exons 1 to 5 and parts of exon 6, including the canonical ATG start codon and the genomic transcription initiation region. Although this finding was compatible with functional impairment of FGFR1, we confirmed transcription of the rearranged *FGFR1* gene by reverse transcriptase PCR, using two different primer pairs (primer pair 1, 2.268 bp and primer pair 2, 2.407 bp, **Suppl. Table 1**), covering the breakpoint - partly exon 6, and exons 7-18 (including the FGFR1 kinase domain) - followed by sequencing (**Fig. 2d**).

We speculated that the remaining open-reading frame of *FGFR1* might be intact and that FGFR kinase activity could be preserved by use of an alternative start codon. This hypothesis was confirmed by an additional *FGFR1*-amplified adenosquamous carcinoma from another patient with unknown response to FGFR inhibition. This patient also revealed a tail-to-tail rearrangement within *FGFR1* (case A921, **Fig. 2e**). The rearrangement was similar to the aforementioned rearrangements that we found in tumors from patients with a response to FGFR inhibition and led to deletion of exons 1 to 8 of *FGFR1* (**Fig 2e, g**). Again, supporting our hypothesis, we found active transcription across the breakpoint (**Fig. 2e, bottom**). In these tumor cells, we also observed positive staining for pFGFR1 by immunohistochemistry (**Fig. 2f**), similar to what we had observed in the FGFR-inhibitor sensitive tumor (**Fig. 1 h**). However, using the next possible in-frame start codon for *FGFR1* translation would lead to deletion of the immunoglobulin-like domain I, II and the acid box for the sample S00674 and the complete extracellular domain of FGFR1 for the sample A921 (**Fig. 2g**). In summary, we found two additional primary tumors with tail-to-tail rearrangements within *FGFR1*, suggesting that rearrangements within *FGFR1* are recurrent events in 8p11-p12 amplified lung cancer. Genome and transcriptome sequencing, validated by independent rtPCR or IHC staining, enabled the discovery of rearrangements and expression of a N-terminally truncated version of FGFR1 ectodomain-deficiency (up to exon 8). The transmembrane and kinase domains of FGFR1 were not impaired in these cases, which is accompanied by the preservation of the catalytic activity of FGFR1.

Oncogenic potential of ectodomain-deficient FGFR1 variants

The N-terminal part of the FGFR1 protein (comprising the ectodomain) is responsible for ligand specificity, as well as for receptor autoinhibition [8, 9]. Furthermore, deletion of the whole FGFR1

ectodomain leads to ligand-independent dimerization [10]. We therefore hypothesized that the transcribed ectodomain-lacking versions/variants of FGFR1 might be oncogenic or cause oncogene dependency, and thus, sensitivity to FGFR inhibition. In order to test this hypothesis, we cloned *FGFR1* and deleted the ectodomain-encoding portion of the gene, up to all theoretically possible in-frame ATG start codons (termed ΔEC-FGFR1), described in **figures 1g, h, i and 2b, e, g** (**Fig. 3a**, **Suppl. Fig. 4**). Ectopic expression of ΔEC-FGFR1 led to IL3-independent cell growth and was sufficient to induce oncogenic transformation in murine Ba/F3 cells. Furthermore, it induced robust phosphorylation of ΔEC-FGFR1 (**Fig. 3b**). By contrast, overexpression of two different wild-type (wt) FGFR1 versions (FGFR1 alpha and FGFR1 beta) was not sufficient to induce IL-3 independence or FGFR1 phosphorylation (**Fig. 3b, Suppl. Fig. 5**). In these experiments, we noted that the bands of ΔEC-FGFR1 migrated at a larger than expected protein size (**Suppl. Fig. 5**). However, through employment of one HA-Tag version of ΔEC-FGFR1, mass spectrometry studies, and mRNA transcript analysis, we were able to formally exclude the possibility of endogenous murine FGFR transactivation in these murine cells (**Suppl. Fig. 5, 6**). Thus, while we can't fully explain the size difference in immunoblotting, our analyses indicate constitutive FGFR1 kinase activation by amino-terminal deletion of FGFR1. Furthermore, ΔEC-FGFR1 caused FGFR1 oncogene dependency in Ba/F3 cells, evidenced by their high sensitivity to FGFR inhibition by the FGFR inhibitors BGJ398 and AZD4547 (**Fig. 3c**). By contrast, bromodomain inhibition, which would inhibit the oncogenic driver NSD3[7], had no differential viability effects on Ba/F3 cells expressing ΔEC-FGFR1, compared to control Baf3 cell lines (**Suppl. Fig. 7**). Furthermore, Baf3 cells, transformed by ΔEC-FGFR1, formed tumors in a xenograft tumor model. These tumors were highly sensitive to FGFR inhibition, using the FGFR inhibitor BGJ398 *in vivo* (**Fig. 3d**). On the contrary, FGFR inhibition had no impact on Baf3 cells, transformed by the oncogene EML4-ALK, thus confirming the selectivity of the compound (**Fig. 3e, f**). Consequently, somatic genomic amino-terminal deletion, including the deletion of all extracellular domains like IG I and acid box (self-inhibitory function), IG II and IG III (ligand binding and specify domain), causes a novel oncogenic variant of FGFR1. This oncogenic variant includes the transmembrane and kinase domain of FGFR1, ΔEC-FGFR1, and causes therapeutically tractable FGFR dependency in vitro and in vivo.

Distinct rearrangements within the 8p11-p12 locus associate with sensitivity to FGFR inhibition in cancer models lacking ectodomain-deficient FGFR1.

We next sought to investigate the role of genomic rearrangements and their underlying impact on genome structure, copy number, and the potential impact on FGFR inhibition in an orthogonal approach. Therefore we performed a small molecule inhibitor screen of 118 cancer cell lines against FGFR- and Bromodomain inhibitors (**Fig. 4a**). Confirming previous findings [3, 11], we found that cell lines bearing genomic *FGFR* alterations (including FGFR mutations and amplifications) were frequently sensitive to FGFR inhibition (p=0.005). Furthermore, while 8p amplification was predictive of sensitivity to FGFR inhibition (p=0.02), it failed to predict sensitivity to bromodomain inhibition (p=0.8), recently described to associate with 8p amplification affecting *NSD3* (**Fig. 4a**) [7]. Of note, some 8p-amplified cell lines were particularly sensitive to FGFR inhibition, while others were not (**Suppl. Fig. 8**), suggesting predominant dependency on the kinase activity in these cases. We focused our analysis on 8 *FGFR1*-amplified lung cancer cell lines, treated with the small-molecule inhibitors BGJ398 and AZD4547 (**Suppl. Fig. 9**). Of these 8 cell lines, 6 were resistant (GI_{50}, > 1 µM)) and 2 were sensitive (GI_{50} < 1 µM) to FGFR inhibition. As mentioned above, we did not find tail-to-tail rearrangements within the *FGFR1* gene itself or genomic evidence of ectodomain-deficient versions of the FGFR1 kinase that we had observed in the patients with a response to FGFR inhibition, aswell as in the cohort of untreated squamous cell lung carcinomas (see above).

In order to identify additional genomic mechanisms driving FGFR1-dependent cancers, we therefore plotted and analyzed the average copy number of these two groups (**Fig. 4d, Suppl. Fig. 10, 11**). Similar to the copy number profiles generated from primary, patient-derived, tumors **(Suppl. Fig. 2)**, we observed centered amplification on *FGFR1* and *NSD3* in the sensitive cell lines, while resistant cell lines showed no clear center of the amplicon (**Fig. 4d**, **Suppl. Fig 9, 10, 11**). Despite the lack of intragenic *FGFR1* rearrangements, we identified tail-to-tail rearrangements close to *FGFR1* in both FGFR-inhibitor sensitive cell lines, which drove the observed amplification pattern (**Fig. 4d**). In detail, we found tail-to-tail rearrangements close to *FGFR1* in H1581 cells (chr8: 38.595.657 bp) and in DMS114 cells (chr8: 38.382.689 bp), but not in any of the resistant cell lines (**Fig. 4d**). Furthermore, we detected a deleterious *NSD3* break in the FGFR inhibitor-sensitive cell line H1581. This rearrangement fused exon 15 of *NSD3* to exon 10 of *ANK1*, creating an out-of-frame fusion, which strongly suggests inactivation of *NSD3*. Thus,

the H1581 cell line exhibited two rearrangements, which were also found in patient TUM009, who had responded to FGFR inhibition (**Fig. 1f and 4d, Suppl. Fig. 2**). Together, the findings suggest that tail-to-tail rearrangements upstream of *FGFR1* lead to the observation of *FGFR1* / *NSD3*-centered amplification patterns. This goes along with sensitivity to FGFR inhibition and co-occurring deletion of the oncogenic NSD3 SET domain (**Fig. 4d**) [7]. Hereby, further support for a key role of FGFR1 in driving sensitivity to FGFR inhibition is provided. In summary, our findings show that (in addition to ectodomain-deficient versions of FGFR1, resulting from intragenic rearrangements) tail-to-tail rearrangements close to *FGFR1* may drive FGFR1-centered amplification, favoring *NSD3* deletion and FGFR1 dependency.

We next collected 85 patient-derived xenotransplant (PDX) models and tested them for 8p11-8p12-amplification by fluorescence in situ hybridization [12]. We identified 8 models harbouring 8p amplifications and treated them *in vivo* with the FGFR inhibitor BGJ398 (**Fig. 4c, Suppl. Fig. 12**). Of these PDX models, 5 were resistant and 3 were sensitive to FGFR inhibition (**Fig. 4c**) [13]. We sequenced all 8 PDX models and sought for the presence of intragenic deletions of *FGFR1* that might cause ectodomain-deficient *FGFR1*. We also plotted the average chromosomal gene copy number of the sensitive and resistant groups for this cohort (**Fig. 4d**). Similar to the findings in the cell lines described above, we did not observe intragenic *FGFR1* rearrangements or other signs of ectodomain-deficient *FGFR1*. However, similar to the pattern of amplification observed in patient samples and cancer cell lines, FGFR inhibitor-sensitive PDX specimens had *FGFR1*/ *NSD3*- centered amplification, while FGFR-deficient samples showed no clear center of amplification (**Fig. 4d, Suppl. Fig. 11**). The amplification pattern observed in FGFR inhibitor-sensitive PDX samples was again driven by tail-to-tail rearrangements close to *FGFR1* and rearrangements within *NSD3* (**Fig. 4d, e**). In detail, we found two tail-to-tail rearrangements close to *FGFR1* (PDX003, chr8: 38.371.080 bp; PDX008, chr8: 38.481.135 bp) and disrupting *NSD3* rearrangements in all three sensitive PDX models (PDX003, deletion of *NSD3* exons 16-24; PDX006, deletion of *NSD3* exons 9-24; PDX008, deletion of *NSD3* exons 9-24) (**Fig. 4d**). In particular, the detected rearrangements in PDX003 and PDX008 showed strong similarities with the FGFR inhibitor-responsive tumor from patient TUM009 and the cancer cell line H1581, indicating recurrent rearrangements associated with FGFR1 dependency (**Fig. 4d, Suppl. Fig. 2, 10, 11**). By contrast, in resistant PDX models, 8p amplifications were neither centered on *FGFR1*, nor focal within the 8p11-8p12 locus (**Fig. 4d**). Furthermore, we found a destructive head-to-head

rearrangement in the insensitive/resistant PDX001 model, deleting the whole *NSD3* gene and exons 6-18 of *FGFR1* and thus the kinase domain of FGFR1. This suggests that neither *FGFR1* nor *NSD3* were the target of 8p amplification (**Fig. 4e, Suppl. Fig. 11**).

In a pooled analysis of cell lines, PDX models, and patient-derived specimens, tail-to-tail rearrangements in or close to *FGFR1* occurred in 7 out of 9 (78%) "responders" and in 3 out of 12 (25%) "non-responders" (for 5 cell lines only whole exome sequencing data was available) and were associated with sensitivity to FGFR inhibition ($p = 0.03$, Fisher's exact test). Furthermore, we found destructive *NSD3* rearrangements in 5 out of 9 FGFR inhibitor-sensitive samples (56%, deleting exons 6 to 23; $p = 0.007$, Fisher's exact test), but not even one in 12 insensitive/resistant models or patients. Thus, in 8p-amplified tumors tail-to-tail rearrangements affecting *FGFR1* were associated with sensitivity to FGFR inhibition, while destructive rearrangements within *NSD3* make a functional role of NSD3 in these tumors unlikely. Combined with the tumors with amino-terminal truncation of *FGFR1*, the carcinomas with tail-to-tail rearrangements within FGFR1 or in close proximity to the FGFR1 gene locus, may thus constitute the overall population of FGFR1-dependent squamous cell lung cancer with sensitivity to FGFR inhibition.

Tail-to-tail rearrangements close to the *FGFR1* gene in primary human lung cancer

We wondered whether these distinct amplification patterns could also be observed in primary 8p11-12 amplified lung tumors with unknown response to FGFR inhibition. We therefore performed an in-depth re-evaluation of the genome sequencing data of the 26 aforementioned 8p11-12-amplified primary tumors, in which we had also discovered the amino-terminally truncated versions of FGFR1. For 25 of these 26 samples, we were able to calculate the chromosomal gene copy number [14-16]. As expected, the average copy number of all 25 samples revealed high-amplitude amplification that was centered on *FGFR1* and the adjacent *NSD3* gene (FDR $q = 1.3\text{x } 10^{-38}$) **(grey line in Fig. 4f, Suppl. Fig. 13**). We next screened for tail-to-tail rearrangements occurring close to the transcription start site of *FGFR1*. We therefore chose a 400-kilobase (kb) region upstream of *FGFR1*, based on the cell lines and PDX specimens that were sensitive to FGFR inhibition. Confirming our findings in cell lines and PDX models, we identified 9 out of 25 samples to harbor a tail-to-tail rearrangement within this region (36%), and one to have co-occurring intragenic rearrangements or an amino-terminal deletion of *FGFR1* (**Fig. 2a**). Remarkably, only these 9 samples with tail-to-tail rearrangements drove the observed *FGFR1*/

NSD3-centered amplification pattern (**blue line in Fig. 4f left, Suppl. Fig. 13**). These specimens exhibited *FGFR1*-centered and focal amplification (550 kb), whereas the amplified region was three times larger (1.7 Mb) in tumors without distinct tail-to-tail breaks (**red line in Fig. 4f left**). Of note, in 5 of these 9 specimens with tail-to-tail rearrangements close to *FGFR1* we observed an additional rearrangement within the *NSD3* gene. These breaks were mainly head-to-head breaks (4 out of 5 cases) and led to the deletion of exons encoding the SET domain of *NSD3* (required for oncogenic transformation) and thus to an amplification pattern favoring *FGFR1* exclusively[7] (**Fig. 4f right, Suppl. Fig. 14**). These tumors also exhibited increased expression of the short isoform of *NSD3* lacking the catalytic SET domain (**Suppl. Fig. 15**). Thus, in 9 of 25 8p-amplified lung cancer specimens, we found tail-to-tail rearrangements close to the transcriptional start site of *FGFR1*. This rearrangement, in particular, induced a copy number gain of *FGFR1*. *Furthermore,* 5 of these tumors had an additional rearrangement within *NSD3* that caused focal amplification, centered exclusively on *FGFR1*, similar to the observed amplification pattern in FGFR inhibitor-sensitive cell lines, PDX models and patient-derived specimens (**Fig. 4d, e, f, Suppl. Fig. 2**). In summary, tail-to-tail rearrangements upstream of *FGFR1* and presumably destructive rearrangements in *NSD3* can be frequently observed in lung cancer cell lines, patient-derived xenotransplant and primary squamous cell lung carcinomas. They can be clearly separated from other 8p11-p12 amplified lung tumors and are nearly identical to rearrangements observed in patients with a response to FGFR inhibition. These distinct rearrangements exclusively favor the *FGFR1* gene and can be associated with FGFR inhibitor sensitivity.

A mechanistic explanation for the emergence of 8p amplification and FGFR1 dependency

As described above, we identified 10 genomic rearrangements within *NSD3* and 5 within *FGFR1* (**Fig. 5a, b**). While the observed breaks in *NSD3* were disrupting (mainly deleting) the oncogenic SET domain, breaks in *FGFR1* were mainly tail-to-tail rearrangements (4 out of 5), amplifying the oncogenic kinase domain (**Fig. 5a, b**) [10, 17]. We furthermore found that in all 25 whole genome-sequenced and *FGFR1*-amplified tumors telomeric losses were accompanied by frequent intra-chromosomal head-to-head and tail-to-tail rearrangements (**Fig. 4d, f, 5a, b, c and Suppl. Fig. 16, 17**). Each of these rearrangements arose from two breakpoints, which were mainly located on the same chromosome and in close proximity to each other. The palindromic nature of a head-to-head rearrangement, induced by two breaks (S00674, chr8: 36.417.298 bp and chr8: 36.418.018 bp) was

validated by use of a single primer PCR (794 bp) (**Fig.5d**). Thus, the characteristic features of telomeric losses, clipped read orientation, and copy number alterations with at least eight copy number segments, indicated a *Breakage-Fusion-Bridge* (BFB) amplification mechanism (**Fig. 5c**) [18, 19]. By applying these criteria, we reliably identified BFBs as the underlying cause of 8p amplification in 44% (11 out of 25) of the primary 8p11-12 amplified tumors. In 56% of the samples a BFB mechanism was uncertain (48%) or excluded (8%) [19]. Following the assembly logic of BFB as the underlying mechanism, we were able to correctly reconstruct the 8p amplicon of one tumor, considering only the observed genomic breaks (**Suppl. Fig. 17, 18**). However, although BFB mechanism explains the emergence of 8p amplification in squamous cell lung cancer, it is unlikely to have predictive value for *FGFR1* dependency by itself (**Fig. 5c**). It suggests that 8p11-12 amplification emerges over several cell generations, ending in a dominant clone or a heterogenic tumor cell population. In summary, 8p amplifications in squamous cell lung cancer are caused by *Breakage-Fusion-Bridges* (BFBs) in a large fraction of cases. In a subset of 8p11-12-amplified tumors, intra-chromosomal tail-to-tail rearrangements close to the *FGFR1* transcription start site are associated with rearrangements in *NSD3* and cause *FGFR1*-centered amplification with frequent functional inactivation of *NSD3*.

DISCUSSION

Here, we report two types of genomic alterations that associate with FGFR1 dependency and thus sensitivity to FGFR inhibition: intragenic rearrangements of *FGFR1* leading to ectodomain-deficient variants of FGFR1 on the one hand, and tail-to-tail rearrangements close to the *FGFR1* gene that cause *FGFR1*-centered amplification (and, in several instances, genomic inactivation of *NSD3)* on the other.

Intragenic *FGFR1* rearrangements causing ectodomain-deficient *FGFR1* were detected in 4 *FGFR1*-amplified lung cancer samples (8% of *FGFR1* amplified samples used in this study). The tail-to-tail rearrangements occurred within the *FGFR1* open reading frame and deleted various portions of the gene. To our surprise, these versions of *FGFR1* were still transcribed by making use of a non-canonical in-frame ATG start codon. Furthermore, we discovered that these ectodomain-deficient/lacking versions of FGFR1 were oncogenic *in vitro* and *in vivo* and led to

sensitivity to FGFR inhibition. Mechanistically, ligand-independent dimerization by FGFR1 variants lacking the ectodomain, has been reported previously, which may explain the phenotype observed by us [10]. However, somatic genomic alterations that cause such variants have not been described to date. Of note, *FGFR2* alterations causing deletion of the extracellular domain were reported recently in cholangiocarcinoma [20].

Furthermore, tail-to-tail rearrangements close to *FGFR1* frequently cause *FGFR1*-centered amplification in tumors lacking intragenic *FGFR1* rearrangements and the encoded ectodomain-deficient versions of the kinase. These rearrangements are similarly associated with sensitivity to FGFR inhibition and frequently accompanied by destructive rearrangements of *NSD3*, thus arguing in favor of a functional relevance of *FGFR1*, rather than of *NSD3* in driving the oncogenic state in the affected tumors. We were able to show that this pattern of amplification was always caused by similar distinct rearrangements, and thus we could link these rearrangements to FGFR inhibitor sensitivity.

Finally, we found that 8p amplifications in squamous cell lung cancer are frequently caused by *Breakage Fusion Bridges* [18, 19]. This mechanism of amplification induces head-to-head, followed by tail-to-tail, intra-chromosomal rearrangements within the 8p arm. The consecutive order of rearrangements underlies an evolutionary process of tumor development. It also explains the frequently described and observed heterogenous pattern of amplification of 8p11-p12 and may also explain the generally limited degree of response to FGFR inhibition as well as early tumor progression under therapy.

Together, our findings suggest that intragenic tail-to-tail rearrangements in *FGFR1,* causing ectodomain-deficient versions of the kinase as well as tail-to-tail rearrangements close to *FGFR1,* that drive *FGFR1*-centered amplification, may define squamous cell lung cancers with therapeutically relevant FGFR dependency.

MATERIALS AND METHODS

Human lung tumor specimens: We collected a total of 26 fresh-frozen squamous cell lung cancer tumor (SQLC) samples, which were provided by multiple collaborating institutions as fresh-frozen tissue specimen, frozen sections or as genomic DNA extracted from fresh-frozen material [21].

All tumor samples were pathologically assessed to have a purity of at least 60% and no extensive signs of necrosis. Additionally, these tumor samples were reviewed by at least two independent expert pathologists and the diagnosis of SQLC was histomorphologically confirmed by H&E staining and immunohistochemistry [21]. Matching normal material was provided in the form of EDTA-anticoagulated blood or adjacent non-tumorigenic lung tissue. The matched normal tissue was confirmed to be free of tumor contaminants by pathological assessment. Furthermore, tumor and matching normal material were confirmed to be acquired from the same patient by SNP 6.0 array and sequencing analyses. Human tumour samples were obtained from patients under IRB-approved protocols following written informed consent and Patient material was stored at −80 °C.

Genome and transcriptome sequencing: Next generation sequencing data of whole exomes or whole genomes were analyzed using our in-house developed pipeline which has been used and described in previous large-scale cancer genome sequencing projects [16, 22-24]. Briefly, the data is processed by aligning sequencing reads to a reference genome (NCBI build 37/hg19) using bwa-mem (0.7.13-r1126; https://github.com/lh3/bwa), masking potential PCR-duplicates and regions of overlapping read pairs, and the collection of various count statistics which are used to call mutations and, for whole genomes, genomic rearrangements. Finally, Sclust [24] is used for sample purity estimation and copy number analysis. RNA-seq data was analysed using TRUP [25].

ARCHER sequencing: RNA was extracted from FFPE material using the Maxwell RSC in combination with the Maxwell RSC RNA FFPE Kit (Promega) according to the manufacturer's instruction. Removal of genomic DNA was performed with the TURBO DNA-free Kit (Thermo Fisher Scientific) and both tRNA and RNA were quantified using the Qubit RNA HS Assay Kit (Thermo Fisher). Analysis of RNA integrity was done with the 4200 TapeStation System (Agilent Technologies, Santa Clara, CA, USA).

For fusion detection, the Archer FusionPlex Lung Panel (ArcherDX) was used according to manufacturer’s instructions with 200 ng tRNA input for library preparation. Using the KAPA Library Quantification Kit (Roche, Pleasanton, CA, USA), purified libraries were quantified. Sequencing was performed on the Nextseq System (Illumina, San Diego, CA, USA) and results were analyzed with the Archer Suite Analysis v 6.2.7 (ArcherDX). Since the Archer Suite Analysis

software is not able to label inversions, bam files were extracted and loaded directly in the Integrative Genomics Viewer (IGV) to visualize tail-to-tail breaks.

Computational Analysis: Copy numbers were visualized with GISTIC (Mermel CH et al., *Genome Biol.* 2011), web-based interactive builder from the MD Anderson Cancer Center, and ROBOCOP. To summaries the copy number information in the FGFR1 region across samples, a view window of 2MB around the FGFR1 locus was split into 10'000 bins. The available chromosome segment-based copy number data were then mapped onto these bins for each of the samples. Next, the per-bin means of the copy numbers for all samples within a certain condition were calculated and the rolling mean with a window size of 10 (for whole genome sequencing data) and 500 of these numbers was determined to reduce noise in the CAGE data. The process of determining ROlling Binmeans Of COPy numbers is defined here as the ROBOCOP algorithm. Inference of BFB mechanism was performed using algorithms and criteria from Zakov et al., PNAS 2013. Given a chromosome segmentation and sequence of segment copy number values n at the (amplified) FGFR1 locus, the approach estimates an approximate copy number sequence n' by applying palindromic BFB transformations to an initial wild-type segmentation. A similarity/distance between the sequences is measured as the Poisson likelihood of observing sequence n given the approximation n'. Using the provided Java application, BFB inference was done in two steps: (1) Estimating n' (java bfb.BFB_Algo counts:n mode:substring model:Poisson maxError:0) and (2) converting n' into a sequence of BFB transformations (java bfb.BFB_Algo counts:n' mode:search). Following Zakov et al., only amplicons with a BFB approximation of at least eight segments were considered as high confidence (n=11), while amplicons with less than eight segments and/or no support by fold-back inversions were too ambiguous to reliably infer BFB as the underlying mechanism (n=14).

Cell culture and reagents: Cell lines were obtained from ATCC, DSMZ (German Collection of Microorganisms and Cell Cultures, Germany), and J. Minna, and colleagues and cultured using either RPMI or DMEM High Glucose media, supplemented with 10%-20% fetal calf serum (FCS). Adherent cells were routinely passaged by washing with phosphate buffered saline (PBS) buffer and subsequent incubation in Trypsin/EDTA. Trypsin was inactivated by addition of culture medium and cells were plated or diluted accordingly. Suspension cell lines were passaged by

suitable dilution of the cell suspension. All cells were cultured at 37°C and 5% CO2. *FGFR1* amplification was determined by SNP 6.0 array, and/or whole genome sequencing, and/or whole exome sequencing, and/or CAGE sequencing or downloaded from COSMIC.
Compounds were obtained from Selleck Chemicals, Tocris Bioscience or Sigma Aldrich. They were diluted in DMSO, aliquoted and stored as 10mM stocks at -80°C.

Viability assays: Initial cell line screen was performed on 384-well plates using 500 to 2000 cells (depending on cell line). For validation cell lines were plated as triplicates into sterile 96-well plates at 1500 (adherend cells) and 5000 (suspension cells) cells/well density as described previously[6]. After 24 hours of incubation, compounds were added at increasing dosages, ranging from 30 to 0.00003 µM together with a separate DMSO control. After 96 hours, relative cell viability was determined by comparing the ATP-content of each well - assessed by CellTiter Glo Assay (Promega, US) - to the content of the DMSO control. Finally, half-maximal growth inhibitory concentrations (GI50) were calculated by R programming language.

Patient derived xenografts: In total we performed *FGFR1* fluorescence in situ hybridization (FISH) on 35 patient derived xenograft (PDX) models from EpoBerlin, 39 PDX models from Crown Bioscience Inc., 3 samples of Moro Massimo as previously described [12]. In addition, we established 8 PDX models by implanting subcutaneously 2-3mm^3 tumor pieces from biopsies (transported in DMEM media) together with 5-10µl of Matrigel (Corning) in NSG mice. We identified eight *FGFR1* amplified models and performed CAGE sequencing [6]. CAGE sequenced samples confirming *FGFR1* amplification were daily oral treated with BGJ398 (Tocris Bioscience, 20mg/kg, dissolved in 33% PEG300, 5% Glucose, prepare stock freshly every week and store at 4°C) or vehicle (33% PEG300, 5% Glucose) with at least three animals per group.

Xenograft Model: IL3 independent Baf3 Cells transformed with deltaEC-FGFR1 or EML4-Alk were subcutaneously injected in both flanks of NSG mice. Mice with palpable formed tumours were randomly grouped and either daily treated with BGJ398 (Tocris Bioscience, 20mg/kg, dissolved in 33% PEG300, 5% Glucose, prepare stock freshly every week and store at 4°C) or vehicle (33% PEG300, 5% Glucose)

Patient samples: From the clinical trial Nogova et al. we were able to receive eight formalin fixed biopsies, taken before treatment (NCT01004224) [5]. Three of these patients had a partial response and five demonstrated a stable or progressive disease. Based on the small amount of tissue we were limited to CAGE sequencing [6]. In addition, we received one patient sample with a clear response upon an off-label Pazopanib treatment from Oliver Gautschi. All samples were confirmed to be 8p-amplified by *FGFR1* FISH and CAGE sequencing.

Break validation by PCR: To proof expression over the breakpoint, validate the tail-to-tail rearrangement, as well as the ΔEC-FGFR1 variant, we generated cDNA from 1µg total RNA of the sample S00674 and its corresponding normal (provided from Elisabeth Brambilla, Grenoble). cDNA was generated using the SuperScripIII kit (Invirtrogen) following manufacturer's instructions followed by PCR using the break spanning primers 191_F1_S00674/203_R1_amplify_FGFR1 (Figure 2d) and 193_F2_S00674/ 204_R2_amplify_FGFR1 (Figure 2d, Suppl. Table 1).
A head-to-head rearrangement was validated in this sample by a nested PCR using only one primer per PCR run (based on the palindromic nature of this break). In the first run we used the primer 261_6_R1. In the second run (4µl Template from first PCR) we used the primer 262_6_R2 (Suppl. Table 1). The expected band size was 794 base pairs (Figure 5d, based on the whole genome sequencing data).

ΔEC-FGFR1 cloning:
cDNA of H1581 cells (100ng) was used to amplify *FGFR1* by attB-overhang primers and flip it into pDONR.221 using the BP-clonase (Invitrogen). Bacterial transformation of the competent E. coli strain DH5α (Invitrogen) was carried out according to the manufacturer's instructions. Single clones were sequenced from mini-preparation of plasmid DNA using the NucleoSpin Mini Kit (Machery Nagel). For midi-preparation of plasmid DNA we used the NucleoBond Xtra Midi EF Kit (Machery Nagel). The different *FGFR1* variants were generated by side using Gibson Assembly and the primers 234_F_ΔEC-21-FGFR1, 235_F_ΔEC-30-FGFR1, 236_F_ΔEC-85-FGFR1, 237_F_ΔEC-144-FGFR1, and 211_R_FGFR1_GA (Suppl. Table 1).

Virus production: Virus was produced as described previously [6].

Immunoblotting: Cells were washed with cold PBS and lysed in RIPA Lysis Buffer supplemented with protease (Roche) and phosphatase inhibitor (Calbiochem) cocktails. After 20 minutes of incubation on ice, lysates were centrifuged at 18,000 g for 25 minutes. Protein concentration in supernatants was measured using BCA Protein Assay (ThermoScientific). Equivalent amounts of protein (30–60µg) were denatured and separated on 4–12% SDS-PAGE gels and after blotting on nitrocellulose membranes (Amersham Hybond-C Extra). The following antibodies were used for immunoblotting: β-actin (MP Bioscience), HSP90 (CS4877S, cell signaling) phospho-FGFR (Tyr653, Tyr654, CS3676, cell signaling), HA-TAG (CS3724, cell signaling), phospho-AKT (Ser473, CS4370), AKT (CS2920), phospho-ERK (CS4370), and total ERK (CS4696), total FGFR1 (ab76464, Abcam), and conjugated antibodies to rabbit and mouse (Millipore).

Immunohistochemistry: Tissues were fixed in 4% PBS-buffered formalin and embedded in paraffin (FFPE). Immunohistochemistry was performed as described previously on 3 µm slides with specific antibodies for pFGFR1 (Abnova, Y154).

Study approval: The study was reviewed and approved by the institutional review board of Department I of Internal Medicine, Center for Integrated Oncology Cologne/Bonn, University Hospital Cologne and Ethics Committee of University Cologne (reference number 06-037, 09-172, 10-275, and 13-091; biomarker research 10-242). The animal experiments were reviewed and approved by the local animal ethics committee of North Rhine-Westphalia located in 40208 Düsseldorf (Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen - LANUV NRW). The protocols were performed according to the recommendations of FELASA.

Data availability: Whole-genome and transcriptome sequencing on human specimen have been deposited at the European Genome-phenome Archive under the accession code EGAS00001005059. For additional data request, please contact the corresponding author.

Author contributions:

Conceptualization: FM, RKT

Methodology: MHAvA, LM, CB, RM, JH

Investigation: FM, LN

Visualization: MHAvA, LM, JB, CB, JG

Funding acquisition: RTU, MLAL, JW, HW, RKT

Project administration: LG, GB, SM

Supervision: FM, RKT

Writing – original draft: FM, RKT

Writing – review & editing: JG, OG, EB, MLAL, JDM, MF, HCR, JW, RB, RKT

Acknowledgments and Funding: This work was supported by the German Cancer Aid (Deutsche KrebshilfeI) through grant ID 70112731 to **R.K.T.**, by the German Ministry of Science and Education (BMBF) as part of the e:Med program (InCa, grant ID: 01ZX1901A to **R.K.T.**, **M.P.**,**J.G.**) and by the German Research Foundation Deutsche Forschungsgemeinsaft (DFG, Deutsche Forschungsgemeinsaft) as part of the SFB1399 (grant ID 413326622 to **R.K.T.**, **M.P.**, **R.B.**, **J.W.**, **H.W.**, **M.F.**, **J.G.**). Additional support for this work was provided by a Cancer Research UK Programme Grant (A17341 to **H.W.**), a Wellcome Trust Investigator Award (214342/Z/18/Z to **H.W.**), a Medical Research Council Grant (MR/S00811X/1 to **H.W.**), a collaborative research grant (SFB1403–414786233) funded by the Deutsche Forschungsgesellschaft (DFG) and an Alexander von Humboldt Foundation Professorship awarded to **H.W.**. **M-L.A-L.** is supported by funding from the Viertel Foundation Senior Medical Research Fellowship. **C.W.** is supported by a Deep Machanda Lung Foundation Australia Post-Doctoral Fellowship and a Cure Cancer grant. **J.D.M.** is supported by the Lung Cancer SPORE (P50 CA070907). We also want to thank Dr. Jana Rolf from EPO-Berlin-Buch GmbH, Dr. Sebastian Dempe and Kira Boehmer from Crown Bioscience Inc., Luca Ross and Moro Massimo for providing PDX samples. Finally we thank the CECAD Proteomics Core Facility, specially Dr. Jan-Wilm Lackmann for mass spectrometry analysis and Dr. Jakob Suckale for support with the video.

References and Notes

1. Siegel, R.L., et al., *Cancer Statistics, 2021.* CA Cancer J Clin, 2021. **71**(1): p. 7-33.
2. Cancer Genome Atlas Research, N., *Comprehensive genomic characterization of squamous cell lung cancers.* Nature, 2012. **489**(7417): p. 519-25.
3. Weiss, J., et al., *Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer.* Sci Transl Med, 2010. **2**(62): p. 62ra93.
4. Katoh, M., *Fibroblast growth factor receptors as treatment targets in clinical oncology.* Nat Rev Clin Oncol, 2019. **16**(2): p. 105-122.
5. Nogova, L., et al., *Evaluation of BGJ398, a Fibroblast Growth Factor Receptor 1-3 Kinase Inhibitor, in Patients With Advanced Solid Tumors Harboring Genetic Alterations in Fibroblast Growth Factor Receptors: Results of a Global Phase I, Dose-Escalation and Dose-Expansion Study.* J Clin Oncol, 2017. **35**(2): p. 157-165.
6. Malchers, F., et al., *Cell-autonomous and non-cell-autonomous mechanisms of transformation by amplified FGFR1 in lung cancer.* Cancer Discov, 2014. **4**(2): p. 246-57.
7. Yuan, G., et al., *Elevated NSD3 histone methylation activity drives squamous cell lung cancer.* Nature, 2021. **590**(7846): p. 504-508.
8. Kalinina, J., et al., *The alternatively spliced acid box region plays a key role in FGF receptor autoinhibition.* Structure, 2012. **20**(1): p. 77-88.
9. Beenken, A. and M. Mohammadi, *The FGF family: biology, pathophysiology and therapy.* Nat Rev Drug Discov, 2009. **8**(3): p. 235-53.
10. Sarabipour, S. and K. Hristova, *Mechanism of FGF receptor dimerization and activation.* Nat Commun, 2016. **7**: p. 10262.
11. Dutt, A., et al., *Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer.* PLoS One, 2011. **6**(6): p. e20351.
12. Schildhaus, H.U., et al., *Definition of a fluorescence in-situ hybridization score identifies high- and low-level FGFR1 amplification types in squamous cell lung cancer.* Mod Pathol, 2012. **25**(11): p. 1473-80.
13. Weeden, C.E., et al., *Cisplatin Increases Sensitivity to FGFR Inhibition in Patient-Derived Xenograft Models of Lung Squamous Cell Carcinoma.* Mol Cancer Ther, 2017. **16**(8): p. 1610-1622.
14. Peifer, M., et al., *Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer.* Nat Genet, 2012. **44**(10): p. 1104-10.
15. Nieroda, L., et al., *iRODS metadata management for a cancer genome analysis workflow.* BMC Bioinformatics, 2019. **20**(1): p. 29.
16. George, J., et al., *Comprehensive genomic profiles of small cell lung cancer.* Nature, 2015. **524**(7563): p. 47-53.
17. Ferone, G., et al., *SOX2 Is the Determining Oncogenic Switch in Promoting Lung Squamous Cell Carcinoma from Different Cells of Origin.* Cancer Cell, 2016. **30**(4): p. 519-532.
18. Tanaka, H. and M.C. Yao, *Palindromic gene amplification--an evolutionarily conserved role for DNA inverted repeats in the genome.* Nat Rev Cancer, 2009. **9**(3): p. 216-24.
19. Zakov, S., M. Kinsella, and V. Bafna, *An algorithmic approach for breakage-fusion-bridge detection in tumor genomes.* Proc Natl Acad Sci U S A, 2013. **110**(14): p. 5546-51.

20. Cleary, J.M., et al., *FGFR2 Extracellular Domain In-Frame Deletions Are Therapeutically Targetable Genomic Alterations That Function as Oncogenic Drivers in Cholangiocarcinoma.* Cancer Discov, 2021. **11**(10): p. 2488-2505.
21. Clinical Lung Cancer Genome, P. and M. Network Genomic, *A genomics-based classification of human lung tumors.* Sci Transl Med, 2013. **5**(209): p. 209ra153.
22. Peifer, M., et al., *Telomerase activation by genomic rearrangements in high-risk neuroblastoma.* Nature, 2015. **526**(7575): p. 700-4.
23. Fernandez-Cuesta, L., et al., *Frequent mutations in chromatin-remodelling genes in pulmonary carcinoids.* Nat Commun, 2014. **5**: p. 3518.
24. Cun, Y., et al., *Copy-number analysis and inference of subclonal populations in cancer genomes using Sclust.* Nat Protoc, 2018. **13**(6): p. 1488-1501.
25. Fernandez-Cuesta, L., et al., *Identification of novel fusion genes in lung cancer using breakpoint assembly of transcriptome sequencing data.* Genome Biol, 2015. **16**(1): p. 7.

Figures

Fig. 1. Tail-to-tail rearrangements in patients responding to FGFR inhibition. a) Overview of study cohorts. **b)** Kaplan-Meier curve showing progression-free survival (PFS) of patients with 8p11 amplified squamous cell lung cancer treated with the FGFR inhibitors, BGJ398, or GW786034 (TUM006). *FGFR1* amplification was determined by FISH. * Treatment stopped due to toxicity. **c)** Tumor volume change of patients with *FGFR1*-amplified squamous cell lung cancer treated with BGJ398 (RECIST criteria). Tumor progression (red) or durable response (blue) following FGFR inhibition. TUM003 and TUM007 died during treatment with no sign of response. One patient (TUM006) was treated off label (* no RECIST data available. Tumor shrinkage estimated on computed tomography (CT) scans (Suppl. Fig. 1). **d)** and **e)** CT scans of patient TUM005 and TUM004 without or durable response under BGJ398 treatment. **f)** Copy number of 6 patients with progressive disease or 4 patients with durable response to FGFR inhibition (Nogova, Malchers et al.). Red arrows indicate samples with tail-to-tail rearrangements within *FGFR1* (highlighted by green frame). **g)** Copy number plot magnified to the *FGFR1* locus (615x sequencing coverage). TUM004 with a response to FGFR inhibition (BGJ398). Normal exon structure of *FGFR1* (middle), resulting genomic rearrangement (bottom), and location of detected breaks. Resulting rearrangements are indicated by arrows. **h)** Copy number plot magnified to the *FGFR1* locus (558x sequencing coverage). Tumor sample TUM006 was obtained from a patient responding to off-label treatment with GW786034. Normal exon structure (middle), resulting genomic rearrangement (bottom), and location of detected breaks and resulting rearrangements are indicated by arrows, pFGFR1 IHC image (right) is shown. **i)** Transcript of *FGFR1* wild type (wt, ENST00000397091.9, top) and transcripts of *FGFR1* found in treatment naive patient samples (middle and bottom) with possible ATG-start codons (TAC-motive from right to left, *FGFR1* is located on the negative strand). Light blue indicate UTRs.

Figure 1 Tail-to-tail rearrangements in patients responding to FGFR inhibition

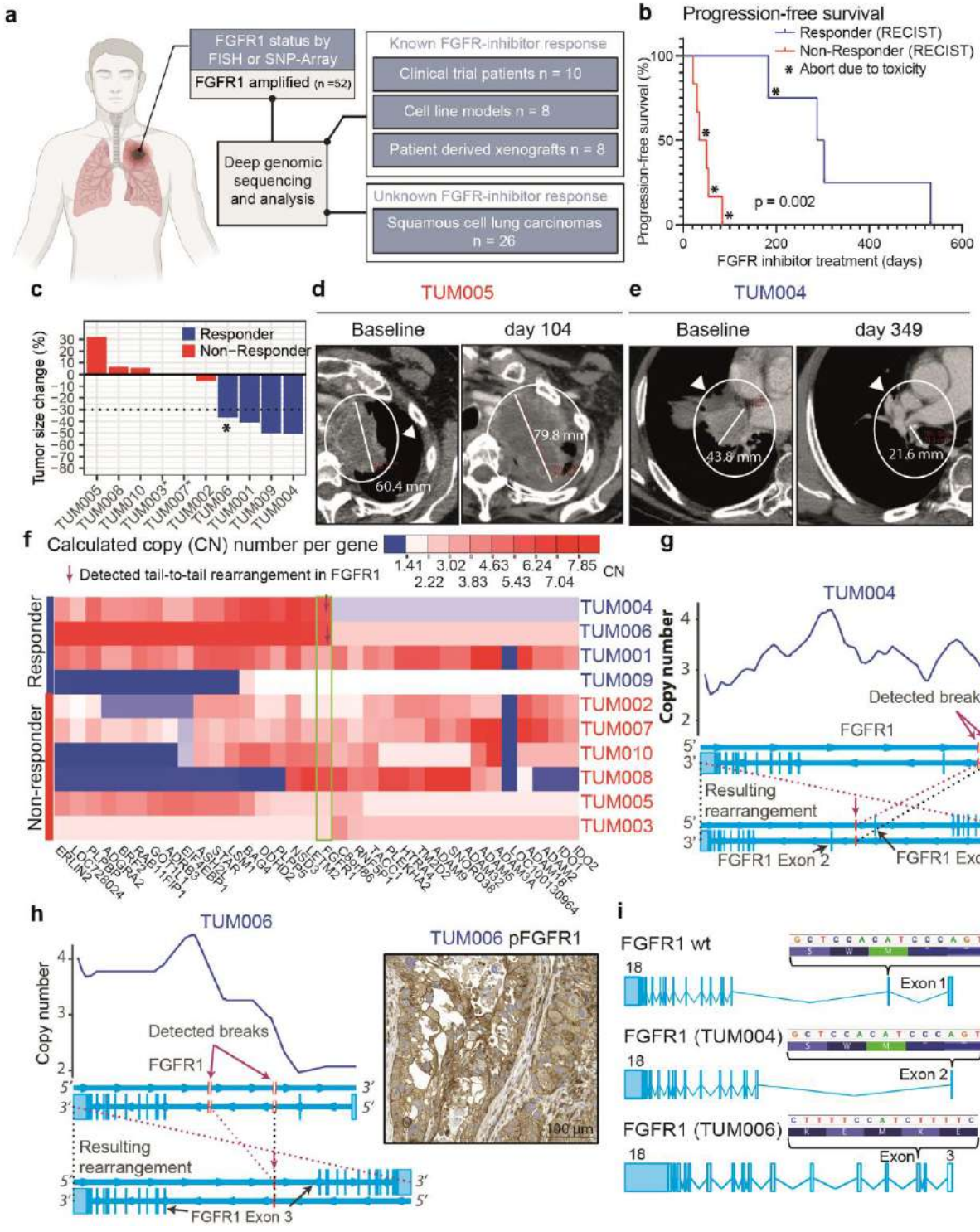

Fig. 2 *FGFR1* tail-to-tail rearrangements in 8p11-8p12 amplified squamous cell lung carcinomas a) Copy number plot (WGS, 30x coverage) of a squamous cell lung carcinoma S00674 (*NSD3* and *FGFR1* are highlighted in orange). The reference genome and the location of genes (wedges) are indicated below (yellow, positive; blue, negative strand, detected breaks are indicated by arrows) **b)** Normal exon structure of *NSD3* and *FGFR1* (top) is indicated. Resulting rearrangement (middle, magenta arrow indicates the tail-to-tail rearrangement, red bars indicate breaks and rearrangement) and breakpoint-spanning reads (from transcriptome sequencing) are shown (bottom). **c)** Expression of NSD3-long and FGFR1-alpha in S00674 determined my transcriptome sequencing is shown. **d)** Electropherogram of a PCR using cDNA generated from tumor and normal (S00674) lung tissue. Two independent primer pairs covering the breakpoint were used (predicted band size: 1, 2.268 bp and 2, 2.407 bp). **e)** Magnified copy number plot showing the genomic *FGFR1* locus (A921, 468x depth, unknown response to FGFR inhibition). Copy number (top), normal exon structure (middle), resulting genomic rearrangement (middle), and break detecting transcriptomic sequencing reads (bottom, magenta arrow indicates the tail-to-tail rearrangement) are indicated. **f)** Microscopic HE (left) and pFGFR1 (right) images of the A921 sample. **g)** Transcripts of *FGFR1* found in patient tumors with unknown FGFR-inhibitor response. Possible ATG-start codons (TAC-motive from right to left, *FGFR1* is located on the negative strand of the reverence genome) and Exons are indicated (light blue are UTRs).

Figure 2 *FGFR1* tail-to-tail rearrangments in 8p11-8p12 amplified squamous cell lung carcinomas

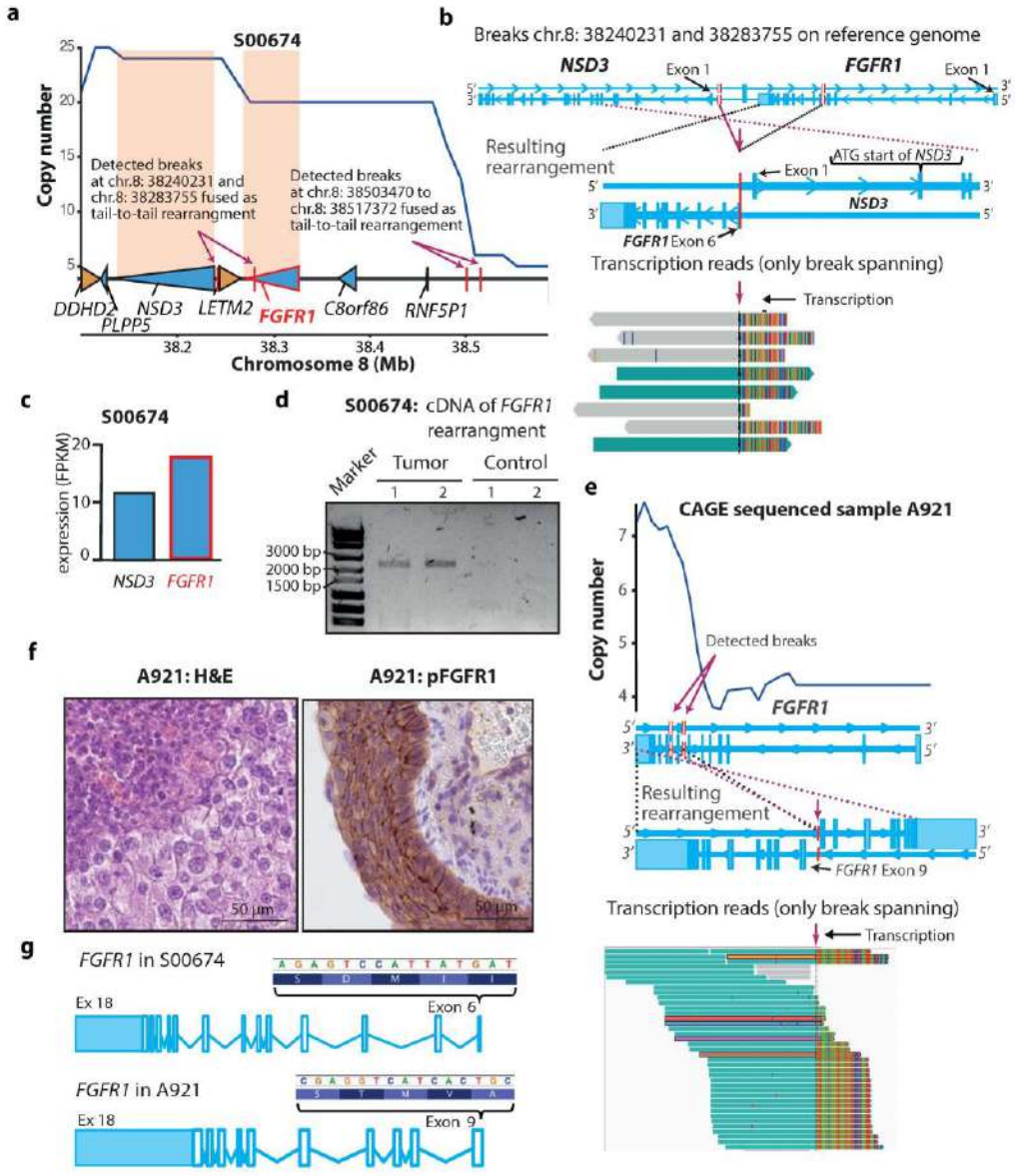

Fig. 3. Oncogenic potential of an ectodomain-deficient version of FGFR1. a) Overview of FGFR1 protein variants using the next possible in frame ATG start codon of transcripts shown in Figure 1i and 2g (AB = acid box, TM = transmembrane domain). **b)** Immunoblots of Ba/F3 cells transduced with retroviruses encoding FGFR1 ΔEC and EML4-ALK (control), as well as parental Baf3 cells or cells transduced with empty Vector (Baf3 e.v.), FGFR1 alpha (Baf3 $FGFR1_{aplpa}$), FGFR1 beta and (Baf3 $FGFR1_{beta}$). Baf3 e.V, FGFR1 alpha and beta are cultured with IL3. **c)** Baf3 e.v., FGFR1beta, and ectodomain lacking *FGFR1* (FGFR1 ΔEC, using an inframe ATG in exon 9) were incubated with increasing concentrations of the FGFR inhibitor BGJ398 (top), or the FGFR inhibitor AZD4547 (bottom) for 96 hours, measuring ATP content to determined viability. Baf3 e.v. and Baf3 FGFR1beta cells was screened in the presence of IL-3, whereas Baf3 FGFR1 ΔEC were screened without IL-3. **d**) Quantification of xenograft tumor models engrafted with Ba/F3 cells expressing FGFR1 ΔEC (blue) or EML4-ALK (red), treated with BGJ398 (20 mg/kg, q.d., red/blue bars) or vehicle (dashed red/blue bars) **e**) Tumor volumes of a xenograft tumor model engrafted with Ba/F3 cells expressing FGFR1 ΔEC, treated with BGJ398 (20 mg/kg, q.d., blue curve), or vehicle (black curve), respectively, upon formation of palpable tumors. Tumor volumes were assessed as indicated and compared by 2-tailed t test. **f**) Representative pictures of the xenograft models are shown before experiment termination.

Figure 3 Oncogenic potential of an ectodomain-deficient version of FGFR1

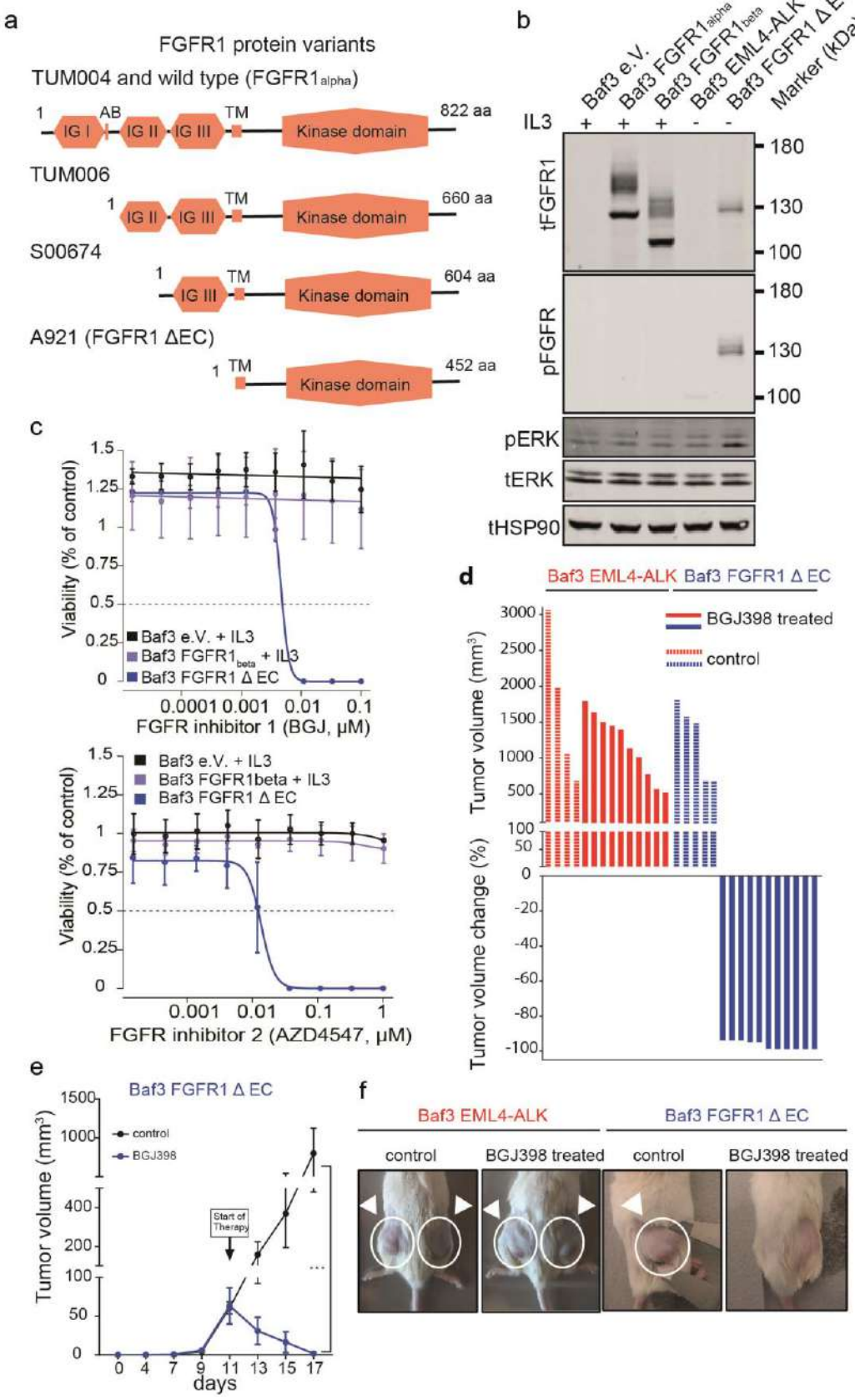

Fig. 4. Rearrangements associated with FGFR inhibitor sensitivity. a) Half-maximal inhibitory concentrations (GI50) of 118 cancer cell lines treated with BGJ398 or JQ1, sorted for the presence or absence of somatic *FGFR* gene family alterations (left) or presence or absence of *FGFR1* amplification (excluding other FGFR alterations, right). **b)** Average GI50 values of 8 cell lines (> 1 µM in red; < 1 µM in blue) treated with the FGFR-inhibitors BGJ398 or AZD4547 **c)** Average tumor growth reduction of 8 patient-derived xenograft tumor models treated with 20 mg/kg BGJ398 or vehicle control (resistant red, tumor reduction < 30 %; sensitive blue, tumor reduction > 30 %). * kindly provided from C E Weeden et al., Clin. Can. Res. 2017 **d)** Average copy number of 6 resistant (red) and 2 sensitive cell lines (blue) to FGFR inhibition (top) and 5 resistant (red) and 3 sensitive (blue) patient-derived xenograft tumor models (bottom). *NSD3* and *FGFR1* are highlighted (orange). Location of genes (wedges) are indicated below (yellow, positive; blue, negative strand). Rearrangements in responding samples are indicated **e)** Illustration of three possible rearrangements and their impact on copy number (for a detailed explanation see suppl. Video 1) **f)** Average copy number of 25 8p-amplified primary squamous cell lung cancer specimens with unknown responsiveness to FGFR inhibition (whole-genome sequenced). Plotted together (grey, n=25) or with (400kb range, blue, n=9) or without observed tail-to-tail rearrangements (>1 mb range, red, n=16) before *FGFR1*. Magnification of the amplification peak is shown (right, blue group, n=9). Only rearrangements observed within the open reading frame of *NSD3* or *FGFR1* are indicated (arrows). Corresponding rearrangements within the same sample are also indicated, if located within the same sample (black lines) and if detected within the magnified area (arrows: head-to-head in green, normal in black, and tail-to-tail in red).

Figure 4 Rearrangements associated with FGFR inibitor sensitivity

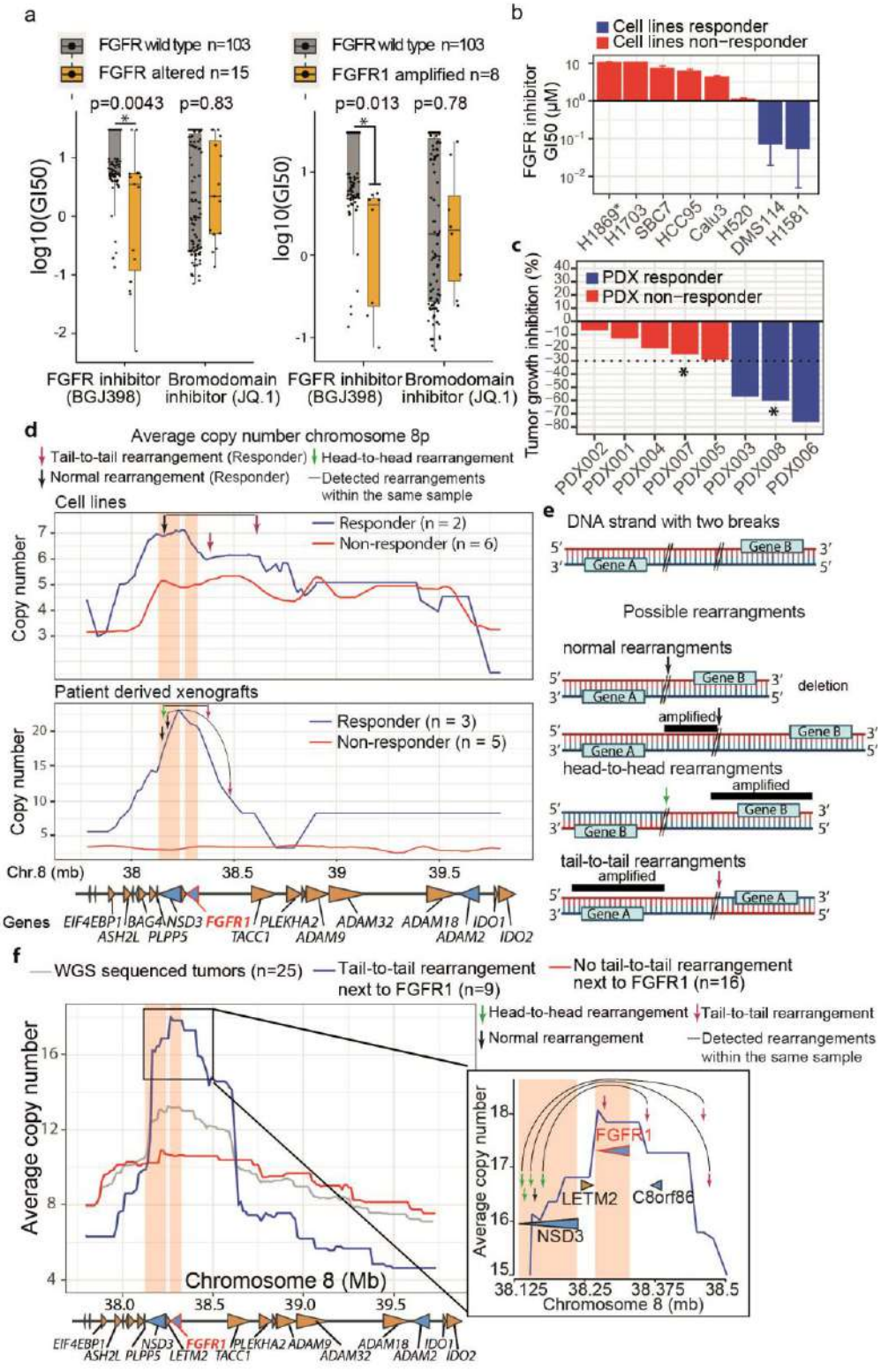

Fig. 5. *FGFR1* dependency and consecutive order of genomic rearrangements. a) Domain architecture of NSD3 long (top, 1-1437 amino acids) and NSD3 short (bottom, 1-645 amino acids) and all detected rearrangements within all study groups (n = 52, arrow color indicate type of rearrangement). **b)** Domain architecture of FGFR1 alpha (1-822 AA, n = 52, arrow color indicates type of rearrangement, TM= transmembrane domain, AB= acid box (self-inhibitory function). **c)** Schematic overview of two breakage-fusion-bridge (BFB) mechanisms forming 8p11-p12 amplifications, differing only in the consecutive order of genomic rearrangements. **d)** Electropherogram of a PCR across a head-to-head rearrangement (S00674, tumor and matched normal DNA) using only one Primer.

Figure 5 FGFR1 dependency and consecutive order of genomic rearrangments