Genetic determinants of *EGFR*-Driven Lung Cancer Growth and Therapeutic Response *In Vivo*

Giorgia Foggetti*[1], Chuan Li*[2], Hongchen Cai*[3], Jessica A. Hellyer[4], Wen-Yang Lin[3], Deborah Ayeni[1], Katherine Hastings[1], Jungmin Choi[5,6], Anna Wurtz[1], Laura Andrejka[3], Dylan Maghini[3], Nicholas Rashleigh[1], Stellar Levy[1], Robert Homer[7,8], Scott Gettinger[9], Maximilian Diehn[4,10,11], Heather A. Wakelee[4], Dmitri A. Petrov[2], Monte M. Winslow[3,12,13,#] and Katerina Politi[1,7,9,#]

[1]Yale Cancer Center, [5]Department of Genetics, [7]Pathology, [9]Medicine (Section of Medical Oncology), Yale School of Medicine, New Haven, Connecticut

[6]Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Korea

[8]VA Connecticut HealthCare System, Pathology and Laboratory Medicine Service, West Haven, Connecticut

[2]Department of Biology, [3]Genetics, [4]Stanford Cancer Institute, Department of Medicine, Division of Oncology, [10]Institute for Stem Cell Biology and Regenerative Medicine, [11]Department of Radiation Oncology, [12]Pathology, [13]Cancer Biology Program, Stanford University School of Medicine, Stanford, California

* Authors contributed equally

To whom correspondence should be addressed:

Katerina Politi, PhD
Departments of Pathology and Internal Medicine (Section of Medical Oncology)
Yale Cancer Center
Yale University School of Medicine
333 Cedar Street, SHM-I 234D
New Haven, CT 06510 USA
Office Tel: 203-737-5921
Lab Tel: 203-737-6215
Fax: 203-785-7531
Email: katerina.politi@yale.edu

Monte Winslow, PhD
Departments of Genetics and Pathology
Stanford University School of Medicine
279 Campus Drive, B256
Stanford, CA 94305 USA
Office Tel: 650-725-8696
Email: mwinslow@stanford.edu

Abstract

Cancer genome sequencing has uncovered substantial complexity in the mutational landscape of tumors. Given this complexity, experimental approaches are necessary to establish the impact of combinations of genetic alterations on tumor biology and to uncover genotype-dependent effects on drug sensitivity. In lung adenocarcinoma, *EGFR* mutations co-occur with many putative tumor suppressor gene alterations, however the extent to which these alterations contribute to tumor growth and their response to therapy *in vivo* has not been explored experimentally. By integrating a novel mouse model of oncogenic *EGFR*-driven *Trp53*-deficient lung adenocarcinoma with multiplexed CRISPR–Cas9-mediated genome editing and tumor barcode sequencing, we quantified the effects of inactivation of ten putative tumor suppressor genes. Inactivation of *Apc*, *Rb1*, or *Rbm10* most strongly promoted tumor growth. Unexpectedly, inactivation of *Lkb1* or *Setd2* – which were the strongest drivers of tumor growth in an oncogenic *Kras*-driven model – reduced *EGFR*-driven tumors growth. These results were consistent with the relative frequency of these tumor suppressor gene alterations in human *EGFR*- and *KRAS*-driven lung adenocarcinomas. Furthermore, *Keap1* inactivation reduced the sensitivity of tumors to osimertinib in the $EGFR^{L858R}$;$p53^{flox/flox}$ model. Importantly, in human *EGFR/TP53* mutant lung adenocarcinomas, mutations in the KEAP1 pathway correlated with decreased time on tyrosine kinase inhibitor treatment. Our study highlights how genetic alterations can have dramatically different biological consequences depending on the oncogenic context and that the fitness landscape can shift upon drug treatment.

59 During tumor evolution, cancer cells accumulate alterations in oncogenes and tumor
60 suppressor genes, which contribute to many of the hallmarks of cancer[1]. Despite their extensive
61 genomic complexity, tumors are frequently classified based on the presence of a single
62 oncogenic driver mutation, while the function of co-incident tumor suppressor gene alterations
63 is largely ignored. There is emerging evidence that the interplay between oncogenic drivers and
64 tumor suppressor genes may influence tumor fitness and impact treatment response[2].
65 However, the combinatorially vast landscape of genomic alterations makes it difficult, except in
66 the most extreme cases, to glean information about the epistatic interactions between tumor
67 suppressor genes and oncogenes[3]. This complexity makes inferring the relationship between
68 genotype and therapy response even more tenuous.
69 Recently, high-throughput, tractable systems that combine autochthonous mouse modeling
70 and genome editing have been developed to directly uncover the functional consequences of
71 genetic alterations during tumorigenesis *in vivo*[4-10]. However, very few studies have
72 investigated the biological consequences of inactivating tumor suppressor genes in the context
73 of different oncogenic drivers *in vivo*, and existing knowledge about the role of specific genetic
74 alterations in tumor suppressor genes has been primarily inferred from correlative human
75 studies[11-13].
76 In lung adenocarcinoma, *EGFR* and *KRAS* are the two most commonly mutated oncogenic
77 driver genes and occur within a background of diverse putative tumor suppressor gene
78 alterations[2,11,12,14]. Among these, *TP53* is the most commonly mutated tumor suppressor gene
79 in both oncogenic *EGFR*- and *KRAS*-driven lung adenocarcinoma, consistent with the
80 importance of disrupting this pathway during lung cancer development[2,13,15-18]. Interestingly,
81 many other putative tumor suppressor genes are mutated at different frequencies in oncogenic
82 *EGFR*- and *KRAS*-driven lung adenocarcinomas[2,11,15]. Whether these differences are due to

different fitness effects that depend on the oncogenic context has never been tested experimentally. Indeed, previous studies on tumor suppressor genes in lung cancer models *in vivo* have been primarily conducted in the context of oncogenic *Kras*-driven tumors, while the functional importance of different tumor suppressor genes in *EGFR*-driven lung tumors remains almost completely unknown (**Supplemental Fig. 1a**).

In addition to driving growth, inactivation of tumor suppressor pathways may affect the therapeutic response to therapies. In advanced *EGFR* mutant lung adenocarcinomas, treatment with EGFR tyrosine kinase inhibitors (TKIs) is the first-line of therapy[19-22]. Response rates to TKIs are high, however, there is large variability in the depth and duration of response between patients and acquired resistance inevitably occurs[14]. Genomic alterations, including those in *RB1* and *TP53*, have been found to correlate with clinical responses to TKIs[13,17,23]. However, given the complexity and diversity of genomic alterations in these tumors, the functional contribution of individual genes to drug resistance remains poorly understood.

To quantify the importance of a panel of ten diverse putative tumor suppressor genes on oncogenic *EGFR*-driven lung tumor growth *in vivo,* we coupled multiplexed CRISPR–Cas9-mediated somatic genome editing and tumor barcoding sequencing (Tuba-seq) with a novel genetically engineered mouse model of *EGFR*^{L858R}-driven *Trp53*-deficient lung cancer. Through the comparison of tumor suppressor effects in oncogenic *EGFR*- and *Kras*-driven lung cancer models, we uncovered prevalent epistasis between tumor suppressor genes and the oncogenic drivers, which explains the different mutational spectra of these tumor suppressor genes in oncogenic *EGFR*- and *KRAS*-driven human lung adenocarcinomas. Moreover, we established a direct causal link between tumor suppressor genotypes and differential responses to osimertinib treatment in *EGFR*-driven lung adenocarcinoma that were supported by correlative human mutational datasets.

107 **Results**

108 **Development of a lentiviral-Cre based mouse model of oncogenic *EGFR*-driven lung**

109 **adenocarcinoma**

110 The use of lentiviral vectors to initiate tumors in genetically engineered mouse models

111 of human cancer enables control of tumor multiplicity, tumor barcoding to map clonality, and

112 the delivery of lentivirus-encoded cDNAs, shRNAs, and sgRNAs to modify neoplastic cells[6,8,24,25].

113 The simplicity of viral-Cre initiated models of oncogenic *Kras*-driven lung cancer has enabled

114 the analysis of genes that co-operate to drive tumor growth within these autochthonous mouse

115 models[26]. To permit the generation of virally-initiated oncogenic *EGFR*-driven *Trp53*-deficient

116 lung tumors, we bred mice with an rtTA/tetracycline-inducible transgene encoding the common

117 lung adenocarcinoma-associated $EGFR^{L858R}$ mutant (*TetO-EGFR*^{L858R}), a Cre-regulated *rtTA*

118 transgene (*Rosa26*$^{CAGs\text{-}LSL\text{-}rtTA3\text{-}IRES\text{-}mKate}$, abbreviated *R26*^{RIK}), and homozygous *Trp53* floxed alleles

119 (*p53*$^{flox/flox}$)[27-29]. In these *TetO-EGFR*^{L858R};*Rosa26*^{RIK};*p53*$^{flox/flox}$ (*EGFR;p53*) mice, lentiviral-Cre

120 transduction of lung epithelial cells leads to the expression of rtTA and mKate as well as the

121 inactivation of *Trp53*. Co-incident doxycycline treatment induces the expression of oncogenic

122 EGFR (**Fig. 1a**).

123 We initiated tumors with a lentiviral-Cre vector in *EGFR;p53* mice and used magnetic

124 resonance imaging (MRI) to monitor tumor development (**Fig. 1b**). Tumors were first detectable

125 in *EGFR;p53* mice 8 weeks after tumor initiation. Histological analysis of lungs 11 weeks after

126 tumor initiation confirmed the development of multifocal lung adenomas and adenocarcinomas

127 (**Fig. 1b**, bottom panels). Tumors stained positively for $EGFR^{L858R}$ and mKate, as well as

128 surfactant protein C (SP-C) and the lung lineage-defining transcription factor NKX2.1/TTF-1 (**Fig.**

129 **1c**, top panels and **Supplemental Fig. 1b**)[30-32]. Importantly, in this model, tumors were more
130 focal than the diffuse tumors that rapidly develop in the previous *CCSP-rtTA;TetO-EGFR*^{L858R}
131 model[27] likely due to tumor initiation from fewer cells in the virally initiated model (**Fig. 1b**, top
132 right panel). Lentivirus-induced tumors in *EGFR;p53* mice predominantly had a micropapillary
133 growth pattern and were also less differentiated that those typically observed in the *CCSP-*
134 *rtTA;TetO-EGFR*^{L858R} model (**Fig. 1b**, bottom panels)[27]. Thus, this new lentiviral-Cre-initiated
135 model recapitulates the genetic and histopathological features of human oncogenic *EGFR-*
136 driven *TP53*-deficient tumors.
137
138 **Multiplexed quantification of tumor suppressor gene function in *EGFR*-driven lung tumors**
139 To enable somatic genome editing in the *EGFR;p53* model, we further incorporated a
140 conditional *Cas9* allele (*Rosa26*$^{LSL\text{-}Cas9\text{-}GFP}$)[4] to generate *TetO-EGFR*^{L858R}*;R26*$^{RIK/LSL\text{-}Cas9\text{-}GFP}$*;p53*$^{flox/flox}$
141 (*EGFR;p53;Cas9*) mice (**Fig. 1a**). Lentiviral-Cre delivery to *EGFR;p53;Cas9* mice initiated
142 multifocal lung adenomas and adenocarcinomas that expressed EGFR^{L858R}, mKate, Cas9, and
143 GFP (**Figs. 1c, d**). Tumors in *EGFR;p53;Cas9* mice were histologically similar to those in the
144 *EGFR;p53* mice (**Figs. 1b, d**).
145 We used an improved version of our Tuba-seq approach to quantify tumor suppressor
146 gene function in oncogenic *EGFR*-driven lung tumors (**Methods**). Genomic integration of
147 barcoded lentiviral vectors uniquely tags each transduced cell and all of the neoplastic cells
148 within the resulting clonal tumors[8]. Each barcode encodes an 8-nucleotide sgID region specific
149 to the sgRNA followed by a random 15-nucleotide barcode; thus, high-throughput sequencing
150 of this sgID-BC region from bulk tumor-bearing lung can be used to quantify the number of cells

151 in each tumor of each genotype (**Methods**)[8]. In Tuba-seq, the absolute number of neoplastic
152 cells in each tumor is calculated by normalizing the number of reads of each unique barcode
153 (sgID-BCs) to the number of reads from benchmark control cells added to each sample.
154 Therefore, using Tuba-Seq we can quantify the size of tumors with different tumor suppressor
155 gene alterations.

156 To assess the function of ten diverse putative tumor suppressor genes, which are
157 frequently altered in human lung adenocarcinoma, we initiated tumors in *EGFR;p53* and
158 *EGFR;p53;Cas9* mice with a pool of barcoded Lenti-sgRNA/*Cre* vectors (Lenti-sgTS^{Pool}/*Cre*; **Fig.**
159 **2a**). In addition to Lenti-sgRNA/*Cre* vectors targeting each putative tumor suppressor gene, this
160 pool contains negative control vectors, including four Lenti-sg*Inert/Cre* vectors and a vector
161 with an sgRNA targeting *Trp53* (which is already inactivated by Cre-mediated recombination in
162 *EGFR;p53;Cas9* mice, **Fig. 1a**)[8]. Lenti-sgTS^{Pool}/*Cre*-initiated tumors were first detectable by MRI 4
163 weeks after tumor initiation in *EGFR;p53;Cas9* mice. At 11 weeks after tumor initiation, when
164 tumors were readily detectable in all mice, tumor-bearing lungs were collected for Tuba-seq
165 analysis and histology (**Fig. 2b** and **Supplemental Fig. 1e**).

166

167 **Putative tumor suppressor genes have distinct effects on *EGFR*-driven lung tumor growth**

168 We used Tuba-seq to quantify the number of neoplastic cells in clonal tumors of each
169 genotype in the *EGFR;p53* model and used two summary statistics to describe the tumor size
170 distribution (different percentiles within the tumor size distributions and the log-normal (LN)
171 mean) (**Methods**; **Figs. 2c-e** and **Supplemental Figs. 2a, b**). For negative controls, tumors with
172 each sgRNA in *EGFR;p53* mice, which lack the *Cas9* allele, had very similar size distributions (**Fig.**

173 **2d**). Furthermore, in *EGFR;p53;Cas9* mice, tumors initiated with each Lenti-sg*Inert/Cre* vector or
174 Lenti-sg*p53/Cre* had very similar tumor size profiles (**Fig. 2c**).
175 Inactivation of *Rb1*, *Apc*, and *Rbm10* most dramatically enhanced the growth of oncogenic
176 *EGFR*-driven *Trp53*-deficient lung tumors (**Figs. 2c, e**). While the importance of *RB1* inactivation
177 in *EGFR*-driven lung adenocarcinomas has begun to be investigated[23,33], very few studies have
178 addressed the functional importance of the APC pathway in *EGFR*-driven lung
179 adenocarcinoma[34]. Interestingly RBM10, an RNA binding protein and splicing regulator that is
180 poorly studied in cancer in general, has not previously been implicated as a critical regulator of
181 *EGFR*-driven lung cancer growth[12,35-37].
182 Surprisingly, inactivation of either *Lkb1* or *Setd2* – which are strong tumor suppressor
183 genes in analogous oncogenic *Kras*-driven tumor models – dramatically reduced tumor growth
184 relative to sg*Inert* tumors[8,38,39]. This effect was consistent across multiple percentiles within the
185 tumor size distribution and as assessed by the tumor size LN mean (**Figs. 2c, e**). These results
186 suggest that in specific contexts, inactivation of genes presumed to be tumor suppressors can
187 have deleterious effects on cancer growth. Other tumor suppressor genes, *Atm*, *Arid1a*, *Cdkn2a*
188 and *Keap1*, did not significantly alter tumor growth in the context of this experiment (**Figs. 2c,**
189 **e**).
190 To assess tumor suppressor gene function at a later time-point of tumor growth, we
191 initiated tumors in *EGFR;p53;Cas9* mice with 8-fold less Lenti-sgTS^{Pool}*/Cre* and performed Tuba-
192 seq after 19 weeks of tumor growth. At this time-point, the histology of the tumors was similar
193 to those observed after 11 weeks of tumor initiation (**Supplemental Figs. 1c, d**). Tuba-seq
194 analysis confirmed the tumor-suppressive function of *Rbm10*, *Apc,* and *Rb1* (**Supplemental Figs.**

195 **2a, b**). Since we used an 8-fold lower viral titer for this experiment, there were proportionally
196 fewer tumors, which limited the resolution of Tuba-seq analysis (**Supplemental Fig. 1f**). Thus,
197 while inactivation of the other genes had no significant effect on tumor growth at this time-
198 point, we cannot exclude that these genes may influence tumor growth. Interestingly, despite
199 the decreased statistical power, inactivation of *Cdkn2a* or *Arid1a* had a positive effect on tumor
200 growth at this 19-week time-point (but not at the 11-week time-point) suggesting a potential
201 role of these tumor suppressor genes at this specific phase of tumorigenesis in the model
202 (**Supplemental Figs. 2a, b**).
203
204 **Validation of *Apc* and *Rbm10*-mediated tumor suppression**
205 We performed further experiments to confirm the role of two less-well studied tumor
206 suppressors, Apc and Rbm10 on the progression of *EGFR*-driven tumors. We initiated lung
207 tumors in *EGFR;p53* and *EGFR;p53;Cas9* mice with Lenti-sg*Apc/Cre*, Lenti-sg*Neo2/Cre* (sg*Inert*),
208 and two Lenti-sg*Rbm10/Cre* vectors each with a unique sgRNA targeting *Rbm10* (*N* = 3
209 *EGFR;p53* mice/group and *N* = 5 *EGFR;p53;Cas9* mice/group; **Fig. 3a**). Inactivation of either *Apc*
210 or *Rbm10* in *EGFR;p53;Cas9* mice increased tumor size and total tumor burden relative to
211 controls as quantified by both MRI and histology (**Figs. 3b-f** and **Supplemental Fig. 3b**). In
212 particular, *Apc* inactivation gave rise to tumors that were strikingly larger than control tumors
213 in *EGFR;p53* mice (**Figs. 3b-d** and **Supplemental Fig. 3b**). Lenti-sg*Apc*/Cre-initiated tumors in
214 *EGFR;p53;Cas9* mice had more cancer cells with nuclear localization of β-catenin (consistent
215 with *Apc* inactivation) and Lenti-sg*Rbm10/Cre*-initiated tumors in *EGFR;p53;Cas9* mice generally
216 lacked Rbm10 protein (**Figs. 3d-f** and **Supplemental Figs. 3c, d**). Tumors with either *Apc* or

Rbm10 inactivation were histologically similar to tumors in *EGFR;p53* mice (**Supplemental Fig. 3a**). These results further confirm the importance of these tumor suppressor pathways in constraining *EGFR*-driven tumor growth *in vivo.* Collectively, our findings underscore the value of coupling Tuba-seq and CRISPR–*Cas9*-mediated somatic genome editing with our virally-induced mouse model to dissect gene function in oncogenic *EGFR*-driven lung cancer.

Oncogenic drivers shape the fitness landscape of tumor suppression

The extent to which different oncogenic drivers affect the landscape of tumor suppression is almost entirely unknown. We approached this question by comparing the fitness landscape of tumor suppression within the contexts of oncogenic *EGFR*- and *Kras*-driven lung tumors. We repeated an experiment previously performed by our group in which we inactivated the same panel of tumor suppressor genes in *Kras*$^{LSL\text{-}G12D/+}$;*p53*$^{flox/flox}$;*R26*$^{LSL\text{-}Tomato}$;*H11*$^{LSL\text{-}Cas9}$ (*Kras;p53;Cas9*) mice and used library preparation methods and the analytical pipeline identical to those used for the *EGFR;p53* and *EGFR;p53;Cas9* mice (**Supplemental Fig. 4a**)[29,40,41]. Inactivation of *Lkb1*, *Setd2,* and *Rb1* were particularly strong drivers of oncogenic *Kras*-driven *Trp53*-deficient tumor growth, while inactivation of *Rbm10*, *Apc*, *Cdkn2a* and *Arid1a* also modestly increased tumor growth (**Figs. 4a, b**). These results are largely consistent with our previous Tuba-seq analysis of *Kras;p53;Cas9* mice, as well as other studies on these genes in oncogenic *Kras*-driven lung cancer mouse models[9,39,42].

Inactivation of several of the tumor suppressor genes (e.g. *Rb1*) had similar effects on *EGFR*- and *Kras*-driven tumors suggesting that these putative tumor suppressor genes limit lung adenocarcinoma growth regardless of the oncogenic context (**Fig.4c** and **Supplemental Fig. 4b**).

239 However, inactivation of either *Lkb1* or *Setd2* greatly increased the growth of oncogenic *Kras*-
240 driven lung tumors but decreased the growth of oncogenic *EGFR*-driven lung tumors (**Figs. 2c,**
241 **e**, **Figs. 4a-c** and **Supplemental Figs. 4b**). Thus, the consequences of tumor suppressor gene
242 inactivation in specific contexts are not limited to the magnitude of tumor-suppressive effects
243 but can also be manifested as opposite effects (known as the sign epistasis) even when the
244 driving oncogenic alterations (in this case *EGFR* and *KRAS*) are traditionally thought to be within
245 a linear pathway.
246
247 **Profound epistasis between tumor suppressor genes and oncogenic drivers drives mutational**
248 **patterns in human lung adenocarcinoma**
249 To compare our functional data in mice with the spectrum of tumor suppressor gene
250 mutations found in human lung adenocarcinomas, we queried data from the AACR Project
251 GENIE database[43]. We calculated the frequency of tumor suppressor gene mutations that occur
252 co-incident with oncogenic *EGFR* (L858R, Exon 19 deletions, L861Q, G719X) or oncogenic *KRAS*
253 (at codons 12, 13 or 61) mutations and *TP53* mutations. This analysis revealed different
254 frequencies of mutations in several tumor suppressor genes in *EGFR/TP53* and *KRAS/TP53*
255 mutant human lung tumors. In line with our data in mice, *RB1*, *RBM10,* and *APC* are frequently
256 altered tumor suppressor genes in *EGFR*/*TP53* mutant lung adenocarcinomas, whereas
257 *LKB1/STK11* and *SETD2,* are among the most frequently mutated tumor suppressor genes in
258 *KRAS/TP53* mutant lung adenocarcinomas (**Fig. 4d**)[44,45]. Further supporting a difference in the
259 function of *LKB1* and *SETD2* in *EGFR/TP53* versus *KRAS/TP53* mutant lung adenocarcinomas,
260 mutations in these genes occurred at significantly higher frequencies in *KRAS/TP53* mutant
261 tumors (**Fig. 4e**). This asymmetry in the mutation frequency of *LKB1* or *SETD2* within oncogenic

262 *EGFR*- or *KRAS*-driven lung tumors is also significant when we extend our analysis to include all
263 tumors regardless of *TP53* mutation status (**Supplemental Fig. 4c**). Collectively, the mouse and
264 human data indicate that the mutational patterns can reflect the biological consequences of
265 profound epistasis between tumor suppressor genes and oncogenic drivers. This highlights the
266 power of our approach in interrogating the functional relevance of genomic combinations on
267 tumorigenesis.
268
269 ***Keap1* inactivation limits the response of tumors to osimertinib**
270 Genetically engineered mouse models have provided insight into the biology of *EGFR*-
271 driven lung tumors and proven valuable in studying mechanisms of resistance to EGFR TKIs,
272 especially on-target (*EGFR*) mechanisms of resistance[27,46-48]. The TKI osimertinib was recently
273 approved for the first-line treatment of *EGFR*-driven lung adenocarcinomas. However,
274 pathways involved in modulating the depth of response and mechanisms of resistance to
275 osimertinib are still under investigation[49,50]. To investigate how tumor suppressors influence
276 the therapeutic response of lung tumors to EGFR inhibition, we treated *EGFR;p53;Cas9* mice
277 with Lenti-sgTS^{Pool}/*Cre*-initiated tumors with osimertinib for two weeks starting at 9 weeks after
278 tumor initiation (**Fig. 5a**). Osimertinib treatment greatly reduced the overall tumor burden
279 relative to vehicle-treated *EGFR;p53;Cas9* mice (**Supplemental Figs. 5a-d**). Residual neoplastic
280 cells were sparse, as determined by staining for $EGFR^{L858R}$ and those cells were not proliferating
281 (**Supplemental Figs. 5e-f**). The overall tumor response was similar when the 2-week treatment
282 was started 17 weeks after tumor initiation (**Supplemental Figs. 5g-l**).
283 To quantify the impact of inactivating each tumor suppressor gene on the response to
284 osimertinib *in vivo*, we performed Tuba-seq on the lungs from osimertinib-treated
285 *EGFR;p53;Cas9* mice 11 and 19 weeks after tumor initiation and compared the results to Tuba-

286 seq results from vehicle-treated controls. Consistent with the imaging data and histological
287 analysis, osimertinib treatment greatly reduced tumor burden as assessed by Tuba-seq (**Fig. 5c**
288 and **Supplemental Figs. 5 and 6c-g**; **Methods**).

289 After two weeks of osimertinib treatment, inactivation of *Apc*, *Rb1* or *Rbm10* was still
290 associated with larger tumors while the size distribution of tumors with inactivation of *Cdkn2a*,
291 *Arid1a*, or *Atm* remained similar to that of tumors with Inert sgRNAs (compare **Fig. 5b** and **Fig.**
292 **2e, Supplemental Fig. 6a** and **Fig. 2c**, **Supplemental Fig. 6b** and **Supplemental Fig. 2a**). The
293 striking exception was tumors with sg*Keap1*. In vehicle-treated mice, the size distribution of
294 sg*Keap1* tumors was almost identical to that of tumors with Inert sgRNAs, however in
295 osimertinib-treated mice sg*Keap1* tumors were significantly larger than the tumors with inert
296 sgRNAs. This suggests that inactivation of *Keap1* limits responses to osimertinib (**Fig. 5b**).
297 Resistance conferred by *Keap1* inactivation was also observed at 19 weeks (**Supplemental Figs.**
298 **6a, b**).

299 We applied an analytical approach that we previously developed and validated to
300 quantify the genotype-specific responses (**Fig. 5c** and **Supplemental Figs. 6c-g**; **Methods**)[51]. By
301 comparing the LN mean of the observed tumor size distributions in osimertinib-treated mice
302 with the expected tumor size distribution based on the overall drug effects, we can estimate
303 genotype specific drug responses (*ScoreRLM*; **Fig. 5c**). After 11 weeks of tumor initiation, sg*Rb1*
304 tumors were 25% smaller than expected (P = 0.04). Conversely, tumors with *Keap1* inactivation
305 were 48% larger than expected (P = 0.07; **Fig. 5c**). The effect of *Keap1* inactivation was even
306 greater at 19 weeks of tumor growth, where tumors were 274% larger than expected (P = 0.13,
307 **Supplemental Fig. 6c**). Other statistical measures of genotype-specific responses, including
308 relative tumor number (*ScoreRTN*) and relative geometric mean (*ScoreRGM*) did not
309 significantly differ between the treated and untreated groups (**Supplemental Fig. 6d**;

310 **Methods**). Our analytical methods allow us to uncover when effects are more pronounced on
311 larger tumors (*ScoreRTN* and *ScoreRGM* have much lower sensitivity when the effects are
312 greater on larger tumors). Thus, our data are consistent with the resistance conferred by *Keap1*
313 inactivation being more pronounced in larger tumors (**Supplemental Figs. 6e-g**; **Methods**).
314

315 ***Keap1*-deficient tumors have reduced sensitivity to osimertinib which correlates with clinical**
316 **outcomes**

317 To further investigate these findings, we initiated tumors with Lenti-sg*Keap1/Cre* in
318 *EGFR;p53* and *EGFR;p53;Cas9* mice followed by treatment with osimertinib or vehicle (**Figs. 5d,**
319 **e**). Osimertinib-treatment reduced the size and number of tumors in *EGFR;p53* mice compared
320 to vehicle-treated *EGFR;p53* mice (**Fig 5d** and **Supplemental Figs. 7a, b**). Conversely, osimertinib
321 treatment of Lenti-sg*Keap1/Cre*-initiated tumors in *EGFR;p53;Cas9* mice did not decrease tumor
322 size or number compared to vehicle-treated *EGFR;p53;Cas9* mice (**Figs. 5d, e** and **Supplemental**
323 **Figs. 7a, b**). Consistent with the inefficiency of CRISPR–Cas9-mediated genome editing in
324 somatic cells, some tumors initiated with Lenti-sg*Keap1/Cre* in *EGFR;p53;Cas9* mice retained
325 expression of Keap1 (**Fig. 5e** top panels). However, the tumors that remained in osimertinib-
326 treated *EGFR;p53;Cas9* mice all had medium to low expression of Keap1 (**Fig. 5e**). Together,
327 these data indicate that while *Keap1* inactivation is not positively selected for during oncogenic
328 *EGFR*-driven tumor growth, osimertinib treatment selects for cancer cells expressing low levels
329 of *Keap1*, thus mitigating the therapeutic response to the drug.

330 To correlate our findings with clinical data, we analyzed the effects of KEAP1 pathway
331 alterations on outcomes with EGFR inhibition in *EGFR/TP53* mutant lung adenocarcinomas.
332 Oncogenic *EGFR*-driven tumors with KEAP1 pathway inactivation have been suggested to be
333 less responsive to TKIs, and we confirmed this association in *EGFR/TP53* tumors (**Figs. 6a, b**)[52].

334 Mutations in the *KEAP1/NFE2L2/CUL3* pathway were associated with a significantly shorter
335 time on EGFR TKI therapy compared with matched patients with *KEAP1/NFE2L2/CUL3* wildtype
336 tumors (5.8 versus 14.3 months, $P = 0.01$; Log-rank test, **Fig. 6a**; **Supplemental Table 1**). Among
337 several other tumor suppressor genotypes, KEAP1 pathway alterations led to the most dramatic
338 reduced response to TKI treatment that remained a significant driver of limited sensitivity after
339 correction for multiple hypothesis testing (**Fig. 6b**). Furthermore, in multivariate analysis, only
340 *KEAP1/NFE2L2/CUL3* mutations significantly reduced time on treatment with EGFR TKI (**Fig. 6b**).
341 We also analyzed a dataset of oncogenic *EGFR* lung adenocarcinoma patient samples
342 collected through the Yale Lung Rebiopsy Program (YLR) prior to first-line TKI treatment
343 (erlotinib, gefitinib, or afatinib) and after the development of resistance to TKIs (**Fig. 6c**;
344 **Supplemental Table 2**). Among 18 patients with *EGFR* and *TP53* mutant tumors, two had *KEAP1*
345 alterations at relapse (TKI-R samples). One case had an acquired missense mutation (Y206N)
346 that lies in a domain of KEAP1 involved in forming the complex with Cullin3 to mediate
347 ubiquitination and degradation of NRF2 (encoded by *NFE2L2*)[53]. The other tumor, which was
348 analyzed only at resistance had heterozygous loss of *KEAP1* (**Fig. 6c**). Moreover, we also
349 observed two cases of *NFE2L2* copy number gain (one detected prior to treatment and
350 maintained at resistance, the other detected only at resistance), and one case of *CUL3*
351 heterozygous loss at resistance. Thus, our *in vivo* functional results are consistent with human
352 data and support a role for inactivation of the KEAP1 pathway in reducing EGFR TKI sensitivity.
353

354 **Discussion**

355 The genomic fitness landscape of lung adenocarcinoma is combinatorially complex, with
356 a large number of genomic alterations in oncogenic drivers and tumor suppressor genes
357 occurring in various combinations[2,3]. While critical for precision medicine approaches,

358 dissecting the functional contribution of these co-occurring genetic alterations to tumor growth
359 and response to therapy is extremely challenging. By using somatic CRISPR–Cas9-mediated
360 gene editing and Tuba-seq[8,9], we evaluated the fitness effects of inactivation of ten tumor
361 suppressor genes commonly altered in lung adenocarcinoma in the contexts of oncogenic *EGFR*
362 and *Kras*-driven tumors. We uncovered distinct roles for specific tumor suppressor genes on
363 tumor growth in the different oncogenic contexts and on sensitivity to the TKI osimertinib *in*
364 *vivo*.

365 *EGFR* and *KRAS* are the most frequently mutated oncogenic drivers in lung
366 adenocarcinoma, collectively occurring in ~40-50% of all patients[2]. Oncogenic alterations in
367 these two genes are mutually exclusive, consistent with them being in the same canonical
368 receptor tyrosine kinase signaling EGFR-RAS-MAPK pathway[54,55]. For this reason, one might
369 anticipate that inactivating tumor suppressor pathways in *EGFR*-driven and *KRAS*-driven lung
370 cancers would have similar biological consequences. It is tempting to make a simplifying
371 assumption that tumor suppressor genes have a constant marginal effect independent of other
372 genomic alterations, or in other words, that the fitness landscape of tumor suppression is
373 smooth and that epistatic interactions are infrequent and/or small in magnitude. Here, we
374 tested this assumption by comparing the fitness landscape of tumor suppression in *EGFR*-driven
375 *Trp53*-deficient and in *Kras*-driven *Trp53*-deficient lung adenocarcinoma and found pervasive
376 epistasis between tumor suppressor genes and oncogenes. Inactivation of *Rb1*, *Rbm10* and *Apc*
377 had a similar effect on *EGFR* and *KRAS*-driven lung tumors (**Figs. 2c, e**, **Figs. 4a-c** and
378 **Supplemental Figs. 4b**). However, while *Lkb1* and *Setd2* were amongst the most potent tumor
379 suppressor genes in *Kras/p53* tumors, inactivating either of these genes reduced the growth of
380 *EGFR/p53* tumors (**Figs. 2c, e** and **Figs. 4a, b**). This observation is consistent with human data,
381 where alterations in either *LKB1* or *SETD2*, are much less common in tumors with oncogenic

382 *EGFR* than in tumors with oncogenic *KRAS* (**Figs. 4d, e** and **Supplemental Fig. 4c**). Thus, we
383 reveal a highly context-dependent fitness landscape of tumor suppression that depends on the
384 nature of the oncogenic driver. Investigation of the mechanisms that underlie the sign epistasis
385 of *LKB1* and *SETD2* may uncover new biological insights and vulnerabilities of *EGFR* mutant
386 tumors. More broadly, sign epistasis within seemingly similar cancer contexts, could help
387 identify genetic interactions for further functional investigation and should be considered when
388 interpreting cancer genomic data (**Figs. 4d, e** and **Supplemental Fig. 4c**).
389 One major advantage of autochthonous genetically engineered mouse models of human
390 cancer is that they can be used to study the impact of inactivation of putative tumor suppressor
391 genes on therapy responses *in vivo*[10,51,56]. We found that inactivation of *Keap1* decreases
392 sensitivity to osimertinib *in vivo* (**Figs. 5b, c**). This finding was supported in a clinical cohort of
393 *EGFR/TP53* mutated lung adenocarcinomas, in which patients with tumors harboring mutations
394 in *KEAP1/NFE2L2/CUL*3 had a significantly shorter time to treatment discontinuation with EGFR
395 TKIs compared with matched controls (**Figs. 6a, b** and **Supplemental Figs. 8a-d**)[17,52]. These
396 results are consistent with this tumor suppressor gene pathway limiting the response to
397 therapy and explains the presence of alterations in *KEAP1* pathway genes in TKI-resistant
398 human tumor specimens. A reduced response to TKI therapy mediated by *KEAP1* inactivation
399 may be associated with the accumulation and transactivation of oxidative stress-related genes
400 by NRF2[57-59]. Alternatively, KEAP1/NRF2-dependent metabolic reprogramming could be
401 involved in mediating drug resistance in lung cancer[60-62]. However, direct genetic alterations in
402 this pathway occur in less than 10% of TKI-resistant *EGFR* mutant lung adenocarcinomas (**Fig.**
403 **6c**)[17]. It is possible that non-genomic alterations that increase NRF2-dependent gene expression
404 or other mechanisms that decrease oxidative stress also occur in TKI-resistant tumors[63-66].
405 Collectively, these results raise the possibility that targeting NRF2 may reduce or delay the

onset of resistance in *EGFR*-driven lung adenocarcinomas. More broadly, our study demonstrates that our approach to can be used to identify clinically relevant pathways that modulate response to therapy *in vivo*. By uncovering the driving forces of the heterogeneity of responses to therapy observed in patients, these types of studies could help define high-risk versus low-risk patient populations and guide therapeutic interventions.

This study provides insight into the complex interplay between tumor suppressor genes and other co-occurring mutations in *EGFR*-driven lung adenocarcinoma tumorigenesis and thus has significant clinical implications. By evaluating interactions between co-occurring alterations in these models, we have avoided confounding factors pervasive in human genomic data (*i.e.,* tumor mutation load, mutation frequency, passenger mutations) and provide clear quantitative data on mutual exclusivity and synergistic effects of genetic alterations. Precise mapping of the fitness consequences of combinations of genetic alterations during tumor evolution will help uncover the biological and clinical relevance of specific alterations during carcinogenesis and identify pathways that can be exploited as therapeutic targets to prevent or overcome resistance to TKIs.

Acknowledgements

G.F was supported by an American Italian Cancer Foundation (AICF) postdoctoral fellowship. C.L. is the Connie and Bob Lurie Fellow of the Damon Runyon Cancer Research Foundation (DRG-2331). H.C. was supported by a Tobacco-Related Disease Research Program Postdoctoral Fellowship (28FT-0019). D.A. was supported by NIH F31 CA203488. K.H. was supported by NIH F32 CA210516. This work was supported by NIH R01 CA231253 (to M.M.W and D.P), NIH R01 CA234349 (to M.M.W and D.P.), NIH R01 CA120247 (to K.P.) and NIH P50 CA196530 (to K.P.). The authors would like to acknowledge the American Association for Cancer

Research and its financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of study authors. We thank members of the Politi, Winslow and Petrov Laboratories for helpful feedback and support in particular, Jacqueline Starrett and Emily Forrest for editing, Fernando de Miguel for creating resources to help with the WES analysis. We also thank Henning Stehr for providing the sequencing data from the STAMP cohort and the Bosenberg Laboratory for sharing their Leica M205 FA stereomicroscope.

Author contributions

G.F. designed and performed *in vivo* experiments, analyzed the data and wrote the manuscript. C.L. performed sequencing analysis, developed and performed statistical analyses, and wrote the manuscript. H.C. designed sgRNAs, generated Lenti-sgRNA/*Cre* vectors, tested sgRNA cutting efficiency, produced lentivirus, and wrote the manuscript. J.A.H. evaluated and analyzed the Stanford Solid Tumor Actionable Mutation Panel (STAMP) cohort of lung cancer patients at Stanford University School of Medicine and wrote the manuscript. W.L. performed *in vivo* experiments, D.A. and K.H. developed the models used for the study. J.C. generated the pipeline for the WES data of the Yale Rebiopsy program (YLR) cohort at the Yale Cancer Center. A.W. coordinates the YLR program. L.A. prepared the libraries for Tuba-seq. D.M. developed the pipeline to analyze the GENIE data. N.R. and S.L. maintained the Yale mouse colony and imaged mice included in this study. R.H. performed the histological analysis of all the mouse lung tumor samples. S.G. conceptualized and leads the Yale Rebiopsy Program at the Yale Cancer Center with K.P.. M.D. conceptualized and analyzed the STAMP cases. H.W. supervised the analysis of the STAMP cohort. D.A.P. conceptualized, supervised the project and wrote the manuscript.

M.M.W. and K.P. designed the experiments, conceptualized, supervised the project and wrote the manuscript. All the authors have revised and approved the submitted version.

Methods

Mice and tumor initiation

TetO-EGFR^{L858R}; *p53*$^{flox/flox}$; *Rosa26*$^{CAGs\text{-}LSL\text{-}rtTA3\text{-}IRES\text{-}mKate}$, *Rosa26*$^{CAGs\text{-}LSL\text{-}Cas9\text{-}GFP}$, *Kras*$^{LSL\text{-}G12D}$, *Rosa26*$^{LSL\text{-}tdTomato}$, and *H11*$^{LSL\text{-}Cas9}$ mice have been described[4,25,27-29,40,41,67,68]. *EGFR*;*p53* and *EGFR*;*p53;Cas9* were on a mixed BL6/129/FVB background and *Kras*;*p53;Cas9* mice were on a mixed BL6/129 background. Approximately equal numbers of males and females were used for each experiment and the number of mice used for each experiment is listed in each figure legend. Lung tumors were initiated by intratracheal administration of Lentiviral-*Cre* vectors as previously described[26]. Tumor burden was assessed by magnetic resonance imaging, fluorescence microscopy, lung weight, and histology, as indicated. Doxycycline was administered by feeding mice with doxycycline-impregnated food pellets (625 ppm; Harlan-Teklad). Osimertinib was resuspended in 0.5% (w/v) methylcellulose (vehicle) and was administered orally (*per os*, 25 mg/kg 5 days a week). All animals were kept in pathogen-free housing under guidelines approved by either the Yale University Institutional Animal Care or the Stanford University Institutional Animal Care and Use Committee guidelines.

Production, purification, and titering of lentivirus

475 The barcoded vectors in the Lenti-sg*TS*^{Pool}*/Cre* have been previously described
476 (**Supplemental Table 3**)[8]. The second Lenti-sg*Rbm10/Cre* vector used in the validation
477 experiments was generated by site-directed mutagenesis (**Supplemental Table 4**). Briefly,
478 Lentiviral-U6-sgRNA*/Cre* vectors contain an 8-nucleotide defined sequence (sgID) that identifies
479 the sgRNA followed by a 15-nucleotide random barcode (BC) to uniquely tag each tumor[8]. To
480 avoid barcode-sgRNA uncoupling driven by lentiviral template switching during reverse
481 transcription of the pseudo-diploid viral genome, each barcoded Lenti-sgRNA*/Cre* vector was
482 generated separately[69,70]. We cultured HEK293T cells in Dulbecco's Modified Eagle Medium
483 with 10% Fetal Bovine Serum and transfected them with individual barcoded Lenti-sgRNA*/Cre*
484 plasmids (*sgLkb1, sgp53, sgApc, sgAtm, sgArid1a, sgCdkn2a, sgKeap1, sgNeo1, sgNeo2, sgNeo3,*
485 *sgNT1, sgRb1, sgRbm10, sgRbm10#2* unbarcoded*, sgSetd2,* or *sgSmad4*) along with pCMV-VSV-
486 G (Addgene #8454) envelope plasmid and pCMV-dR8.2 dvpr (Addgene #8455) packaging
487 plasmid using polyethylenimine. We treated the cells with 20 mM sodium butyrate 8 hours
488 after transfection, changed the culture medium 24 hours after transfection, and collected
489 supernatants 36 and 48 hours after transfection. Subsequently, we removed the cell debris with
490 a 0.45 µm syringe filter unit (Millipore SLHP033RB), concentrated each lentiviral vector by
491 ultracentrifugation (25,000 g for 1.5 hours at 4°C), resuspended the virus in PBS, and stored the
492 virus at -80°C. To determine the titer of each vector, we transduced *Rosa26*$^{LSL\text{-}YFP}$ mouse
493 embryonic fibroblasts (a gift from Dr. Alejandro Sweet-Cordero/UCSF), determined the
494 percentage of YFP-positive cells by flow cytometry, and normalized the titer to a lentiviral
495 preparation of known titer. Lentiviral vectors were thawed and pooled immediately prior to
496 delivery to mice.
497
498 **Lentiviral titers and time of analysis**

Anticipated growth rates were determined by monitoring tumor development through magnetic resonance imaging in pilot experiments and the analysis time-points were selected to ensure that tumors were detectable by MRI such that their response to treatment could be evaluated. Viral titers used in the experiments were chosen to balance the total tumor burden across mice at the time of analysis after tumor initiation. For analysis of tumor growth 11 weeks after tumor initiation, the Lenti-sg*TS*^{Poo}*I/Cre* titer administered to *EGFR*;*p53* mice was 2x10^{6} infectious units (ifu)/mouse, while for *EGFR*;*p53;Cas9* mice we used 1x10^{6} ifu/mouse. We reasoned that using a higher viral titer in the control *EGFR;p53* mice would increase our confidence that any differences observed between *EGFR;p53* and *EGFR;p53;Cas9* mice were due to inactivation of tumor suppressor genes in the latter model. For the 19-week time-point in *EGFR*;*p53;Cas9* mice we initiated tumors with 1x10^{5} ifu/mouse. Two weeks before collection, mice were treated with either vehicle or osimertinib. For the validation experiments in which we used a single vector to initiate tumors (Lenti-sg*Apc/Cre*, Lenti-sg*Rbm10/Cre,* Lenti-sg*Rbm10*#2*/Cre* or Lenti-sg*Neo2/Cre* (sg*Inert*)) we used 1x10^{5} ifu/mouse and analyzed the mice after 14 weeks of tumor growth (**Supplemental Table 4**). For the validation with Lenti-sg*Keap1/Cre* virus (1x10^{5} ifu/mouse), both *EGFR;p53* and *EGFR;p53;Cas9* mice were treated after 15 weeks of tumor initiation and lungs were collected after two weeks of treatment with either vehicle or osimertinib.

Magnetic resonance imaging

All procedures were performed in accordance with protocols approved by the Yale University IACUC and in agreement with the NIH Guide for the Care and Use of Laboratory Animals. Respiratory gated, gradient-echo MR images of mice were collected with a 4T (31-cm bore) small-animal Bruker horizontal-bore spectrometer (Bruker AVANCE). All data were

523 collected as previously described[46]. Tumor volume was quantified by calculating the area of
524 visible lung opacities present in each image sequence per mouse using BioImage Suite 3.01[71].
525

526 **Isolation of genomic DNA from mouse lungs and preparation of sgID-BC libraries**

527 Genomic DNA was isolated from bulk tumor-bearing lung tissue from each mouse as
528 previously described[8]. Briefly, three benchmark control cell lines (~$5x10^5$ cells per cell line) were
529 added to each sample ("spiked-in") prior to lysis to enable the calculation of the absolute
530 number of neoplastic cells in each tumor from the number of sgID-BC reads. Benchmark cell
531 lines carry unique sgID-BCs. Following homogenization and overnight protease K digestion,
532 genomic DNA was extracted from the lung lysates using standard phenol-chloroform and
533 ethanol precipitation.

534 sgID-BC sequencing libraries were prepared by PCR amplifying the sgID-BC region from
535 total genomic DNA. To enable the identification and subsequent computational elimination of
536 index hopped reads after high-throughput sequencing, the sgID-BC region of the integrated
537 Lenti-sg*RNA-BC/Cre* vectors was PCR amplified using unique dual indexing primer pairs. To
538 increase the sequence diversity at each position and reduce the amount of PhiX required to
539 achieve high sequencing quality, we added 6-9 Ns before the sequence-specific primers[51].

540 We used a single-step PCR amplification of sgID-BC regions, which we have found to be
541 a highly reproducible and quantitative method to determine the number of neoplastic cells in
542 each tumor. For each mouse, we performed eight 100 µl PCR reactions per sample (4 µg DNA
543 per reaction, 32 µg per mouse) using Q5 High-Fidelity 2x Master Mix (New England Biolabs,
544 M0494X). The PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter,
545 A63881) using a double size selection protocol. The concentration and quality of the purified

546 libraries were determined using the Agilent High Sensitivity DNA kit (Agilent Technologies,
547 5067-4626) on the Agilent 2100 Bioanalyzer (Agilent Technologies, G2939BA). The libraries
548 were pooled based on lung weight (to have sequencing depth more evenly distributed across
549 samples), cleaned up and size-selected using AMPure XP beads, and sequenced on the
550 Illumina®HiSeq 2500 platform to generate paired-end 150 bp reads (Admera Health).
551

552 **Quantification of tumor sizes from sgID-BC sequencing data**

553 We used stringent filtering to identify the sgID-BC region that minimizes PCR and
554 sequencing error, as previously described[51]. Specifically, we required no mismatch in the
555 barcode region between the forward and reverse reads and removed all spurious tumors with
556 the barcodes within two nucleotides from that of another larger tumor. The absolute number
557 of cells in each tumor was calculated by scaling its sgID-BC read number with the mean read
558 number of three spiked-in cell lines with a known absolute cell number of $5X10^5$.
559

560 **Summary statistics for tumor size distributions**

561 As sequencing depth and PCR efficiency vary across libraries, we focused on tumors that
562 we can repeatedly identify with high confidence, which are tumors over 500 cells as quantified
563 by comparing technical replicates. We used multiple summary statistics to describe the
564 truncated distribution of tumor sizes for all tumors larger than 500 cells. Percentiles and LN
565 means were calculated as two summary statistics. Percentiles are a nonparametric summary of
566 the distribution by taking the 50th, 60th, 70th, 70th, 90th, and 95th percentile of the distribution.
567 The LN mean calculates the maximum likelihood estimator of the mean tumor sizes assuming a
568 log-normal distribution of tumor sizes. For both of these metrics, we normalized to the

corresponding value of the average of Inert tumors to represent the relative growth advantage by carrying the gene.

Quantification of treatment responses of Inert tumors to osimertinib

We quantified the treatment effect of osimertinib by comparing the tumor size distributions of the vehicle- and osimertinib-treated groups. We used two ways to quantify treatment effect. The first way is to calculate the total number of neoplastic cells of tumors carrying the inert control sgRNAs in each mouse, and take the fold change in the average of the total number of neoplastic cells as an approximation of the drug effect.

$$Drug\ effect = \frac{Average\ neoplastic\ cell\ number\ in\ osimertinib - treated\ mice}{Average\ neoplastic\ cell\ number\ in\ vehicle - treated\ mice}$$

This calculation of tumor burden is very intuitive and relevant in the clinical setting, but this measure is quite variable due to the large variations in the sizes of the largest tumors. The second way assumes that each neoplastic cell, regardless of the size of the tumor harboring it, has an equal probability of being killed by osimertinib treatment (K). To make sure that we are evaluating tumors that are large enough to be repeatedly identified, we focus on tumors that are over 1,000 cells in the vehicle-treated mice. We estimated the tumor size reduction after treatment with osimertinib for the tumors with the inert control sgRNAs by matching the distribution of tumor sizes in vehicle- and osimertinib-treated mice. Specifically, we used the binary search algorithm to find the proportion of neoplastic cells remaining after treatment with osimertinib (S) between $K = 1$ (100% cells were killed by osimertinib) and $K = 0$ (0% of cells were killed by osimertinib), such that the median tumor number of tumors with the inert control sgRNAs across the vehicle-treated mice, upon simulated reduction to K, matches the

591 median tumor number of tumors with the inert control sgRNAs larger than 1,000 cells across all
592 osimertinib-treated mice.
593
594 **Estimation of genotype-specific treatment responses**
595 We calculated the expected size distribution of tumors after treatment assuming no
596 genotype-specific treatment responses by reducing all tumors in the vehicle-treated mice by
597 the estimated drug effect (*K*). Then we calculated genotype-specific treatment response for
598 each sg*TS* by comparing the relative LN mean of all tumors in the osimertinib-treated mice and
599 the relative LN mean of all tumors calculated from the expected distribution after treatment.
600 The genotype-specific treatment response is calculated as the log2 ratio of the observed
601 relative LN mean by the expected LN mean, and we named it as *ScoreRLM*. We focus on tumors
602 with the inert control sgRNAs that are over 1,000 cells in untreated mice and take out
603 comparable proportions of tumors with each sgRNA from each vehicle- and osimertinib-treated
604 mice based on the estimated treatment effect and the proportion of tumors carrying each sg*TS*.
605 The *ScoreRLM* is calculated as:

606 $$RLM_{i,j} = Log_2\left(\frac{\text{LN mean}_{\text{i,osimertinib}} / \text{LN mean}_{\text{Inert,osimertinib}}}{\text{LN mean}_{\text{i,vehicle}} / \text{LN mean}_{\text{Inert,vehicle}}}\right)$$

607 $$ScoreRLM_{i,j} = 2^{RLM_{i,j}} - 1$$

608 where $\text{LN mean}_{\text{i,osimertinib}}$ is the LN mean for tumors containing sgID *i* in osimertinib-treated
609 group, $\text{LN mean}_{\text{Inert,osimertinib}}$ is the LN mean for all tumors containing one of the four Inert
610 sgIDs in the osimertinib-treated group. Similarly, $\text{LN mean}_{\text{i,vehicle}}$ is the LN mean for tumors
611 containing sgID *i* in vehicle-treated group and $\text{LN mean}_{\text{Inert,vehicle}}$ is the LN mean for all tumors
612 containing one of the four Inert sgIDs in the vehicle-treated group. When tumors are larger

than expected, the *ScoreRLM* will be positive, indicating resistance conferred by gene inactivation, while when tumors are smaller than expected, the *ScoreRLM* will be negative, indicating sensitivity conferred by tumor suppressor gene inactivation. Although the metric on the log2 scale results in the first formula with good statistical properties ranging from $-\infty$ to $+\infty$ and centered on 0. To make it more directly interpretable by readers, we converted it to the linear scale as shown in the second formula. On the linear scale the metric ranges from -1 to $+\infty$, a value of 0.5 means the tumors are 50% larger than expected and a value of -0.5 means that the tumors are 50% smaller than expected.

Apart from *ScoreRLM* that compares the relative LN mean in the vehicle- and osimertinib-treated group, we can similarly compare the relative tumor number (*ScoreRTN*) and relative geometric mean (*ScoreRGM*) for the observed and expected distribution of tumor sizes following the same logic as shown below:

$$ScoreRGM_{i,j} = Log_2\left(\frac{\text{geometric mean}_{\text{i,osimertinib}} \big/ \text{geometric mean}_{\text{Inert,osimertinib}}}{\text{geometric mean}_{\text{i,vehicle}} \big/ \text{geometric mean}_{\text{Inert,vehicle}}}\right)$$

$$ScoreRGM_{i,j} = 2^{RGM_{i,j}} - 1$$

$$RTN_{i,j} = Log_2\left(\frac{\text{tumor number}_{\text{i,osimertinib}} \big/ \text{tumor number}_{\text{Inert,osimertinib}}}{\text{tumor number}_{\text{i,vehicle}} \big/ \text{tumor number}_{\text{Inert,vehicle}}}\right)$$

$$ScoreRTN_{i,j} = 2^{RTN_{i,j}} - 1$$

where the geometric mean and tumor number were calculated from Inert tumors that are over 1,000 cells in the vehicle-treated mice and the corresponding proportions for other sgIDs in the vehicle-and osimertinib-treated groups considering the proportion of sgIDs and the treatment[51].

The standard deviation of the genotype-specific responses, represented by any of the three metrics, is estimated by bootstrapping mice in both the vehicle- and osimertinib-treated groups and then bootstrapping tumors with the same sgID in each bootstrapped mouse. Such a two-step bootstrap process allows us to control for both variations of tumor sizes across mice and within the same mouse. For each run of bootstrap, we re-estimated the drug effect, estimated the expected tumor size profile between the vehicle- and osimertinib-treated mice. We estimated the standard deviation based on the scores on the log2 scale and then plotted the values on the linear scale, because the latter ranges from -1 to $+\infty$ with no genotype-specific responses located at 1.

Power analysis for three metrics in identifying genotype-specific treatment responses

We calculated the sensitivity and specificity for the three metrics, when 1) the genotype-specific treatment responses to osimertinib only occur in large tumors, while smaller tumors respond similarly to sg*Inert* tumors, 2) the genotype-specific treatment responses are uniform across all tumor sizes such that all sg*TS* tumors have increased or reduced sensitivity to osimertinib relative to sg*Inert* tumors (**Supplemental Figs. 6e, f**). Specifically, we used all ten vehicle-treated mice at 11 weeks for simulation. We first applied a drug effect of K = 75% (75% of cells are expected to be killed by osimertinib assuming no genotype-specific treatment responses) to all tumors. Then we apply the preassigned genotype-specific treatment responses to all tumors to generate the simulated distributed tumor sizes. For the first scenario of size-dependent genotype-specific treatment response, the preassigned genotype-specific responses only occur in tumors with over 10,000 cells, and they are assigned to be four-fold higher than expected, i.e., these tumors do not respond at all to osimertinib considering that 75% of tumor cells are expected to be killed without any genotype-specific treatment responses. For the

658 second scenario of uniform responses, all tumors with the sg*TS* are assigned to be 50% larger
659 than expected, i.e., reduced in size by 62.5% instead of 75% after treatment by osimertinib. The
660 same sample sizes of ten vehicle- and ten osimertinib-treated mice were generated from
661 bootstrapping mice and then bootstrapping tumors from each mouse prior to and after the
662 simulated effects of drug responses and genotype - specific responses, respectively. To
663 calculate the false discovery rate, we also simulated another scenario where no preassigned
664 genotype-specific responses exist and all tumors, regardless of genotypes, were reduced in size
665 uniformly by 75% by osimertinib. The therapeutic sensitivity was calculated as the probability to
666 re-identify the preassigned genotype-specific treatment responses given the cutoff of *P*-values,
667 while specificity was calculated as the probability of falsely identifying genotype-specific
668 treatment responses without any input signal of genotype-specific treatment responses given
669 the cutoff of *P*-values. A total of 100 runs of simulations of the 11 tumor suppressor genes
670 (including Tp53; 1,100 cases) were performed for each scenario. We further plotted the
671 receiver operating characteristic[72] curves for the three metrics by varying the cutoff for the *P*-
672 values.
673
674 **Histology and immunohistochemistry**
675 The auxiliary lobe of the right lung was collected for each experimental mouse from the
676 mice transduced with the Lenti-sgTS^{Pool}/*Cre* pool whereas, both the auxiliary and left lobes were
677 collected for the validation experiments with individual sgRNAs (sg*Apc*, sg*Rbm10*, sg*Rbm10#2*
678 and sg*Neo2*). Right and auxiliary lobes were collected for the experiment with Lenti-
679 sg*Keap1/Cre* virus. Lung lobes were fixed in 4% paraformaldehyde overnight at room
680 temperature, placed in 70% ethanol, and paraffin-embedded and sectioned (Histology @ Yale).
681 Four micrometer sections were used for hematoxylin and eosin (H&E) staining and

682 immunohistochemistry. Tumor sizes were determined by measuring the longest diameter for
683 each tumor in H&E stained sections. Tumor size and tumor area were quantified using ImageJ.
684 The limited tissue collected for histological analysis reduced the number of tumors initiated
685 with Lenti-sg*Rbm10/Cre* that could be measured. *P*-values were calculated from the Mann-
686 Whitney U test. The following antibodies were used for immunohistochemistry: anti-mutant
687 $EGFR^{L858R}$ (1:200, CST #3197), anti-SP-C (1:200, AB40876), anti-TTF-1/Nkx2-1 (1:200, AB76013),
688 anti-phospho-histone H3 (1:200, CST #9701), anti-Ki-67 (1:400, CST #9027), anti-mKate (1:2500,
689 Evrogen #AB233), anti-Cas9 (1:500, Novus Biologicals #7A9-3A3, NBP2-36440), anti-β-catenin
690 (1:500, CST #8814), anti-Rbm10 (1:200, AB224149), and anti-Keap1 (1:500, AB227828). Rbm10
691 expression in tumors was binned as High (over 75% of positive nuclei), Medium (between 25
692 and 75% of positive nuclei) or Low (positive nuclear staining below 25%). Keap1 levels in tumors
693 were binned as High (over 75% of positive cells), Medium (between 25 and 75% of positive
694 cells) or Low (positive staining below 25%). *P*-values were calculated using the Chi-squared test.
695

696 **Analysis of human lung tumor data using GENIE data**

697 The AACR Project GENIE is a registry that contains CLIA-/ISO-certified genomic data
698 collected from the records of more than 9,000 patients who were treated at each of the
699 consortium's participating institutions[43]. Data from GENIE version 4.1-public were accessed
700 through the Synapse Platform. We generated a comprehensive list of all missense, nonsense,
701 and frameshift mutations for all screened genes across all participating centers in the GENIE
702 project, documenting these mutations at the gene and amino acid levels. Based on the sets of
703 genes included in the different screening panels that contribute to Project GENIE, we annotated
704 all lung adenocarcinoma within the database as being wild-type, mutated, or not screened for
705 each gene. From this complete catalog of mutations in each tumor, we determined the rates of

706 co-occurrence of known oncogenic *KRAS* mutations (G12X, G13X, Q61X) or *EGFR* mutations
707 (L858R, Exon 19 deletions, L861Q, G719X) with missense, nonsense, or frameshift mutations in
708 our set of ten tumor suppressor genes. Furthermore, we determined the frequency of co-
709 incident tumor suppressor gene mutations in tumors with oncogenic *KRAS* and inactivating
710 *TP53* mutations (*KRAS/TP53*) and in tumors with oncogenic *EGFR* and inactivating *TP53*
711 mutations (*EGFR/TP53*). A pipeline for this analysis is publicly available at
712 www.github.com/dgmaghini/GENIE, which allows users to input OncoTree codes of interest,
713 generate mutational profiles for the corresponding GENIE tumors, and identify the co-
714 occurrence of up to two background mutations (for example, *EGFR* and *TP53*) with mutations in
715 other genes of interest.
716

717 **Stanford cohort of *EGFR* mutant lung adenocarcinomas**

718 Patients with lung cancer who were evaluated at Stanford Cancer Center and had their
719 tumors analyzed using the Stanford Solid Tumor Actionable Mutation Panel (STAMP)[6] were
720 included in the analysis. This retrospective study was conducted under a molecular analysis
721 protocol approved by the Stanford University Institutional Review Board. All STAMP cases
722 performed between 2015 and 2019 were included; during this time, there were two different
723 assays used, with one covering 198 genes (302 kb) and the other covering 130 genes (232 kb).
724 STAMP was done as the standard of care and thus at the discretion of the treating physician
725 and could have occurred at the time of diagnosis or at the time of progression.

726 From the available STAMP cases, patients were selected if they had stage IV non-small
727 cell lung cancer (NSCLC) and had a pathogenic *EGFR* and *p53* mutation. Patients were excluded
728 if they had incomplete data or were lost to follow up prior to analysis of the primary endpoint,
729 if they elected not to receive treatment or if they received adjuvant tyrosine kinase inhibitor

therapy for early-stage disease. Within this cohort, patients were further selected for the presence of at least one of the tumor suppressor genes investigated in the preclinical setting (*KEAP1, STK11, SETD2, SMAD4, RB1, APC, ARIDIA* and/or *CDKN2A*). To ensure there was enough power to conduct an analysis, genes were included for analysis if there were at least 8 patients with a mutation of that gene who met inclusion criteria. Due to small numbers, patients with *LKB1/STK11* (N = 3), *SETD2* (N = 1) and *SMAD4* (N = 2) mutations were not included. Therefore, only those patients with stage IV, *EGFR/P53* mutant NSCLC who were treated with a tyrosine kinase inhibitor (TKI) and had mutations in either *KEAP1*-pathway component *(KEAP1, NFE2L2, CUL3), RB1, APC, ARIDIA* or *CDKN2A* were included in the analysis.

Demographic data including sex, age at diagnosis, smoking history, and ethnicity were extracted. For each tumor suppressor, patients were matched 1:3 with a wildtype cohort on the basis of sex, smoking history, ethnicity, age and treatment type. Time to Treatment Failure (TTF) was determined by subtracting the date of discontinuation of TKI due to progression, toxicity or death, from the date of initiation of TKI and reported in months. Statistical analysis was performed using Prism 8 and R studio. The Kaplan-Meier method was used to estimate TTF. Comparison of survival curves was made using the Log-rank test. Significance was defined as $P < 0.05$. Hazard ratios (HR) were generated from multivariate regression analysis performed in R and reported with a 95% confidence interval (CI).

Yale cohort of *EGFR* mutant lung adenocarcinomas

Patients with *EGFR*-mutant lung adenocarcinoma were consented and enrolled to a Yale University IRB approved protocol allowing the collection and analysis of clinical data, archival and fresh tissue, blood and the generation of patient-derived xenografts. Patients who received targeted therapy (erlotinib, gefitinib, or afatinib) as first line therapy, either alone or in

754 combination with other therapies such as chemotherapy or cetuximab, were included. For
755 genomic studies, formalin-fixed paraffin embedded tissue was macro-dissected to enrich for
756 tumor material. All of the tumor samples (*N* = 29) analyzed had matched normal tissue and
757 cancer cell purity >20% as assessed by the LOH (Loss of Heterozygosity) events from the whole
758 exome sequencing data.

759 Whole Exome Sequencing

760 DNA was extracted from and analyzed as previously described[73]. Briefly, genomic DNA
761 was captured on the NimbleGen 2.1M human exome array and subjected to 74-bp paired-end
762 reads on the Illumina HiSeq2000 instrument. The mean coverage for normal was 109.1x and the
763 mean coverage for tumor was 189x with 92.78% and 95.56% for the bases covered at least 20
764 independent sequence reads, respectively. Sequence reads were mapped to the human
765 reference genome (GRCh37) using the Burrow-Wheeler Aligner-MEM (BWA-MEM) program.
766 Sequence reads outside the targeted sequences were discarded and the statistics on coverage
767 were collected from the remaining reads using in-house per scripts.

768 Somatic Mutation Calling

769 For all matched tumor-normal pairs, somatic point mutations and indels were called by
770 MuTect2 using Bayesian classifiers. For all somatic mutations called, we extracted base
771 coverage information in all samples and considered the mutations that were supported by at
772 least two independent sequence reads covering non-reference alleles and present in more than
773 5% of all sequencing reads. Identified variants were further filtered based on their presence in
774 repositories of common variations (1000 Genomes, NHLBI exome variant server and 2,577 non-
775 cancer exomes sequenced at Yale) and annotated using ANNOVAR program[74]. All somatic indels
776 were visually inspected to remove the false positive calls.

Somatic Copy Number Variation Analysis

Copy number analysis was performed as previously described[73]. Briefly, copy number variants were identified from the whole exome sequencing data using EXCAVATOR software that normalizes the non-uniform whole exome sequencing data taking GC-content, mappability, and exon-size into account[75]. The Hidden Markov Model was utilized to classify each copy number variant segment into five copy number states (homozygous deletion, heterozygous deletion, normal copy number, homozygous copy gain or multiple copy gain). Tumor purity was estimated from LOH using in-house per scripts.

Data availability

The datasets generated during and/or analyzed in the current study are available in the NCBI Gene Expression Omnibus database (GSE146550). GENIE genomic data analysis is publicly available at www.github.com/dgmaghini/GENIE. All other data supporting the findings are available upon request.

Code availability

The code is available at https://github.com/lichuan199010/Tuba-seq-analysis-and-summary-statistics.

References

1. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
2. Skoulidis, F. & Heymach, J.V. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer* **19**, 495-509 (2019).
3. van de Haar, J. *et al.* Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* **177**, 1375-1383 (2019).
4. Platt, R.J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440-55 (2014).

806 5. Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and
807 metastasis. *Cell* **160**, 1246-60 (2015).
808 6. Chuang, C.H. *et al.* Molecular definition of a metastatic lung cancer state reveals a
809 targetable CD109-Janus kinase-Stat axis. *Nat Med* **23**, 291-300 (2017).
810 7. Winters, I.P. *et al.* Multiplexed in vivo homology-directed repair and tumor barcoding
811 enables parallel quantification of Kras variant oncogenicity. *Nat Commun* **8**, 2053
812 (2017).
813 8. Rogers, Z.N. *et al.* A quantitative and multiplexed approach to uncover the fitness
814 landscape of tumor suppression in vivo. *Nat Methods* **14**, 737-742 (2017).
815 9. Rogers, Z.N. *et al.* Mapping the in vivo fitness landscape of lung adenocarcinoma
816 tumor suppression in mice. *Nat Genet* **50**, 483-486 (2018).
817 10. Winters, I.P., Murray, C.W. & Winslow, M.M. Towards quantitative and multiplexed in
818 vivo functional cancer genomics. *Nat Rev Genet* **19**, 741-755 (2018).
819 11. Campbell, J.D. *et al.* Distinct patterns of somatic genome alterations in lung
820 adenocarcinomas and squamous cell carcinomas. *Nat Genet* **48**, 607-16 (2016).
821 12. Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung
822 adenocarcinoma. *Nature* **511**, 543-50 (2014).
823 13. Blakely, C.M. *et al.* Evolution and clinical impact of co-occurring genetic alterations in
824 advanced-stage EGFR-mutant lung cancers. *Nat Genet* **49**, 1693-1704 (2017).
825 14. Politi, K. & Herbst, R.S. Lung cancer in the era of precision medicine. *Clin Cancer Res*
826 **21**, 2213-20 (2015).
827 15. Jordan, E.J. *et al.* Prospective Comprehensive Molecular Characterization of Lung
828 Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies.
829 *Cancer Discov* **7**, 596-609 (2017).
830 16. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl*
831 *J Med* **376**, 2109-2121 (2017).
832 17. Yu, H.A. *et al.* Concurrent Alterations in EGFR-Mutant Lung Cancers Associated with
833 Resistance to EGFR Kinase Inhibitors and Characterization of MTOR as a Mediator of
834 Resistance. *Clin Cancer Res* **24**, 3108-3118 (2018).
835 18. Scheffler, M. *et al.* K-ras Mutation Subtypes in NSCLC and Associated Co-occuring
836 Mutations in Other Oncogenic Pathways. *J Thorac Oncol* **14**, 606-616 (2019).
837 19. Lynch, T.J. *et al.* Activating mutations in the epidermal growth factor receptor
838 underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**,
839 2129-39 (2004).
840 20. Paez, J.G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to
841 gefitinib therapy. *Science* **304**, 1497-500 (2004).
842 21. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from "never
843 smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc*
844 *Natl Acad Sci U S A* **101**, 13306-11 (2004).
845 22. Soria, J.C. *et al.* Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell
846 Lung Cancer. *N Engl J Med* **378**, 113-125 (2018).
847 23. Offin, M. *et al.* Concurrent RB1 and TP53 Alterations Define a Subset of EGFR-Mutant
848 Lung Cancers at risk for Histologic Transformation and Inferior Clinical Outcomes. *J*
849 *Thorac Oncol* (2019).
850 24. DuPage, M. *et al.* Endogenous T cell responses to antigens expressed in lung
851 adenocarcinomas delay malignant tumor progression. *Cancer Cell* **19**, 72-85 (2011).

852 25. Caswell, D.R. *et al.* Obligate progression precedes lung adenocarcinoma
853 dissemination. *Cancer Discov* **4**, 781-9 (2014).
854 26. DuPage, M., Dooley, A.L. & Jacks, T. Conditional mouse lung cancer models using
855 adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc* **4**, 1064-72 (2009).
856 27. Politi, K. *et al.* Lung adenocarcinomas induced in mice by mutant EGF receptors found
857 in human lung cancers respond to a tyrosine kinase inhibitor or to down-regulation of
858 the receptors. *Genes Dev* **20**, 1496-510 (2006).
859 28. Dow, L.E. *et al.* Conditional reverse tet-transactivator mouse strains for the efficient
860 induction of TRE-regulated transgenes in mice. *PLoS One* **9**, e95236 (2014).
861 29. Jonkers, J. *et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a
862 conditional mouse model for breast cancer. *Nat Genet* **29**, 418-25 (2001).
863 30. Puglisi, F. *et al.* Prognostic value of thyroid transcription factor-1 in primary, resected,
864 non-small cell lung carcinoma. *Mod Pathol* **12**, 318-24 (1999).
865 31. Tan, D. *et al.* Thyroid transcription factor-1 expression prevalence and its clinical
866 implications in non-small cell lung cancer: a high-throughput tissue microarray and
867 immunohistochemistry study. *Hum Pathol* **34**, 597-604 (2003).
868 32. Tanaka, H. *et al.* Lineage-specific dependency of lung adenocarcinomas on the lung
869 development regulator TTF-1. *Cancer Res* **67**, 6007-11 (2007).
870 33. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective
871 clinical sequencing of 10,000 patients. *Nat Med* **23**, 703-713 (2017).
872 34. Nakayama, S. *et al.* beta-catenin contributes to lung tumor development induced by
873 EGFR mutations. *Cancer Res* **74**, 5891-902 (2014).
874 35. Bechara, E.G., Sebestyen, E., Bernardis, I., Eyras, E. & Valcarcel, J. RBM5, 6, and 10
875 differentially regulate NUMB alternative splicing to control cancer cell proliferation.
876 *Mol Cell* **52**, 720-33 (2013).
877 36. Hernandez, J. *et al.* Tumor suppressor properties of the splicing regulatory factor
878 RBM10. *RNA Biol* **13**, 466-72 (2016).
879 37. Zhao, J. *et al.* Functional analysis reveals that RBM10 mutations contribute to lung
880 adenocarcinoma pathogenesis by deregulating splicing. *Sci Rep* **7**, 40488 (2017).
881 38. Walter, D.M. *et al.* Systematic In Vivo Inactivation of Chromatin-Regulating Enzymes
882 Identifies Setd2 as a Potent Tumor Suppressor in Lung Adenocarcinoma. *Cancer Res*
883 **77**, 1719-1729 (2017).
884 39. Ji, H. *et al.* LKB1 modulates lung cancer differentiation and metastasis. *Nature* **448**,
885 807-10 (2007).
886 40. Jackson, E.L. *et al.* Analysis of lung tumor initiation and progression using conditional
887 expression of oncogenic K-ras. *Genes Dev* **15**, 3243-8 (2001).
888 41. Chiou, S.H. *et al.* A conditional system to specifically link disruption of protein-coding
889 function with reporter expression in mice. *Cell Rep* **7**, 2078-86 (2014).
890 42. Walter, D.M. *et al.* RB constrains lineage fidelity and multiple stages of tumour
891 progression and metastasis. *Nature* **569**, 423-427 (2019).
892 43. Consortium, A.P.G. AACR Project GENIE: Powering Precision Medicine through an
893 International Consortium. *Cancer Discov* **7**, 818-831 (2017).
894 44. Skoulidis, F. *et al.* Co-occurring genomic alterations define major subsets of KRAS-
895 mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic
896 vulnerabilities. *Cancer Discov* **5**, 860-77 (2015).

897 45. Arbour, K.C. *et al.* Effects of Co-occurring Genomic Alterations on Outcomes in
898 Patients with KRAS-Mutant Non-Small Cell Lung Cancer. *Clin Cancer Res* **24**, 334-340
899 (2018).
900 46. Pirazzoli, V. *et al.* Afatinib plus Cetuximab Delays Resistance Compared to Single-Agent
901 Erlotinib or Afatinib in Mouse Models of TKI-Naive EGFR L858R-Induced Lung
902 Adenocarcinoma. *Clin Cancer Res* **22**, 426-35 (2016).
903 47. Pirazzoli, V. *et al.* Acquired resistance of EGFR-mutant lung adenocarcinomas to
904 afatinib plus cetuximab is associated with activation of mTORC1. *Cell Rep* **7**, 999-1008
905 (2014).
906 48. Starrett, J.H.G., A.; Cuomo, M.E.; Poels, K.; van Alderwerelt van Rosenburgh I.;
907 Nagelberg, A.; Farnsworth, D.; Price, K.S.; Khan, H.; Asthtekar, K.D.; Gaefele, M.; Ayeni,
908 D.; Stewart, T.F.; Kuhlmann, A.; Keack, S.; Unni, A.M.; Homer, R.; Lockwood, W.W.;
909 Michor, F.; Goldberg, S.B.; Lemmon, M.A.; Smith, P.D.; Cross, D.A.; Politi, K. Drug
910 sensitivity and allele-specificity of first-line osimertinib resistance *EGFR* mutations.
911 *Under Revision* (2020).
912 49. Ramalingam, S.S. *et al.* Osimertinib As First-Line Treatment of EGFR Mutation-Positive
913 Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol* **36**, 841-849 (2018).
914 50. Ramalingam, S.S. *et al.* Overall Survival with Osimertinib in Untreated, EGFR-Mutated
915 Advanced NSCLC. *N Engl J Med* (2019).
916 51. Li, C.L., W.; Rizvi, H.; Cai, H; McFarland, C.D.; Rogers, Z.N.; Yousefi, M.; Winters, I.P.;
917 Rudin, C.M.; Petro, D.A.; Winslow, M.M. Quantitative *in vivo* analyses reveal a complex
918 pharmacogenomic landscape in lung adenocarcinoma. *BioRxiv [**Preprint**] January 29,*
919 *2020. Available from: https://doi.org/10.1101/2020.01.28.923912* (2020).
920 52. Hellyer, J.A. *et al.* Impact of KEAP1/NFE2L2/CUL3 mutations on duration of response
921 to EGFR tyrosine kinase inhibitors in EGFR mutated non-small cell lung cancer. *Lung*
922 *Cancer* **134**, 42-45 (2019).
923 53. Cleasby, A. *et al.* Structure of the BTB domain of Keap1 and its interaction with the
924 triterpenoid antagonist CDDO. *PLoS One* **9**, e98896 (2014).
925 54. Gainor, J.F. *et al.* ALK rearrangements are mutually exclusive with mutations in EGFR
926 or KRAS: an analysis of 1,683 patients with non-small cell lung cancer. *Clin Cancer Res*
927 **19**, 4273-81 (2013).
928 55. Unni, A.M., Lockwood, W.W., Zejnullahu, K., Lee-Lin, S.Q. & Varmus, H. Evidence that
929 synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR
930 mutations in lung adenocarcinoma. *Elife* **4**, e06907 (2015).
931 56. Kersten, K., de Visser, K.E., van Miltenburg, M.H. & Jonkers, J. Genetically engineered
932 mouse models in oncology research and cancer medicine. *EMBO Mol Med* **9**, 137-153
933 (2017).
934 57. Yamadori, T. *et al.* Molecular mechanisms for the regulation of Nrf2-mediated cell
935 proliferation in non-small-cell lung cancers. *Oncogene* **31**, 4768-77 (2012).
936 58. Krall, E.B. *et al.* KEAP1 loss modulates sensitivity to kinase targeted therapy in lung
937 cancer. *Elife* **6**(2017).
938 59. Park, S.H. *et al.* Resistance to gefitinib and cross-resistance to irreversible EGFR-TKIs
939 mediated by disruption of the Keap1-Nrf2 pathway in human lung cancer cells. *FASEB*
940 *J*, fj201800011R (2018).
941 60. DeNicola, G.M. *et al.* NRF2 regulates serine biosynthesis in non-small cell lung cancer.
942 *Nat Genet* **47**, 1475-81 (2015).

61. Romero, R. *et al.* Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. *Nat Med* **23**, 1362-1368 (2017).
62. Galan-Cobo, A. *et al.* LKB1 and KEAP1/NRF2 Pathways Cooperatively Promote Metabolic Reprogramming with Enhanced Glutamine Dependence in KRAS-Mutant Lung Adenocarcinoma. *Cancer Res* **79**, 3251-3267 (2019).
63. Wang, J. *et al.* Nestin regulates cellular redox homeostasis in lung cancer through the Keap1-Nrf2 feedback loop. *Nat Commun* **10**, 5043 (2019).
64. Jain, A. *et al.* p62/SQSTM1 is a target gene for transcription factor NRF2 and creates a positive feedback loop by inducing antioxidant response element-driven gene transcription. *J Biol Chem* **285**, 22576-91 (2010).
65. Wang, Q. *et al.* CDK20 interacts with KEAP1 to activate NRF2 and promotes radiochemoresistance in lung cancer cells. *Oncogene* **36**, 5321-5330 (2017).
66. Marcar, L. *et al.* Acquired Resistance of EGFR-Mutated Lung Cancer to Tyrosine Kinase Inhibitor Treatment Promotes PARP Inhibitor Sensitivity. *Cell Rep* **27**, 3422-3432 e4 (2019).
67. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133-40 (2010).
68. Nakada, D., Saunders, T.L. & Morrison, S.J. Lkb1 regulates cell cycle and energy metabolism in haematopoietic stem cells. *Nature* **468**, 653-8 (2010).
69. Hill, A.J. *et al.* On the design of CRISPR-based single-cell molecular screens. *Nat Methods* **15**, 271-274 (2018).
70. Annunziato, S. *et al.* Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland. *Genes Dev* **30**, 1470-80 (2016).
71. Papademetris, X. *et al.* BioImage Suite: An integrated medical image analysis suite: An update. *Insight J* **2006**, 209 (2006).
72. Skoulidis, F. *et al.* STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. *Cancer Discov* **8**, 822-835 (2018).
73. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* **106**, 19096-101 (2009).
74. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
75. Xu, D., Olman, V., Wang, L. & Xu, Y. EXCAVATOR: a computer program for efficiently mining gene expression data. *Nucleic Acids Res* **31**, 5582-9 (2003).

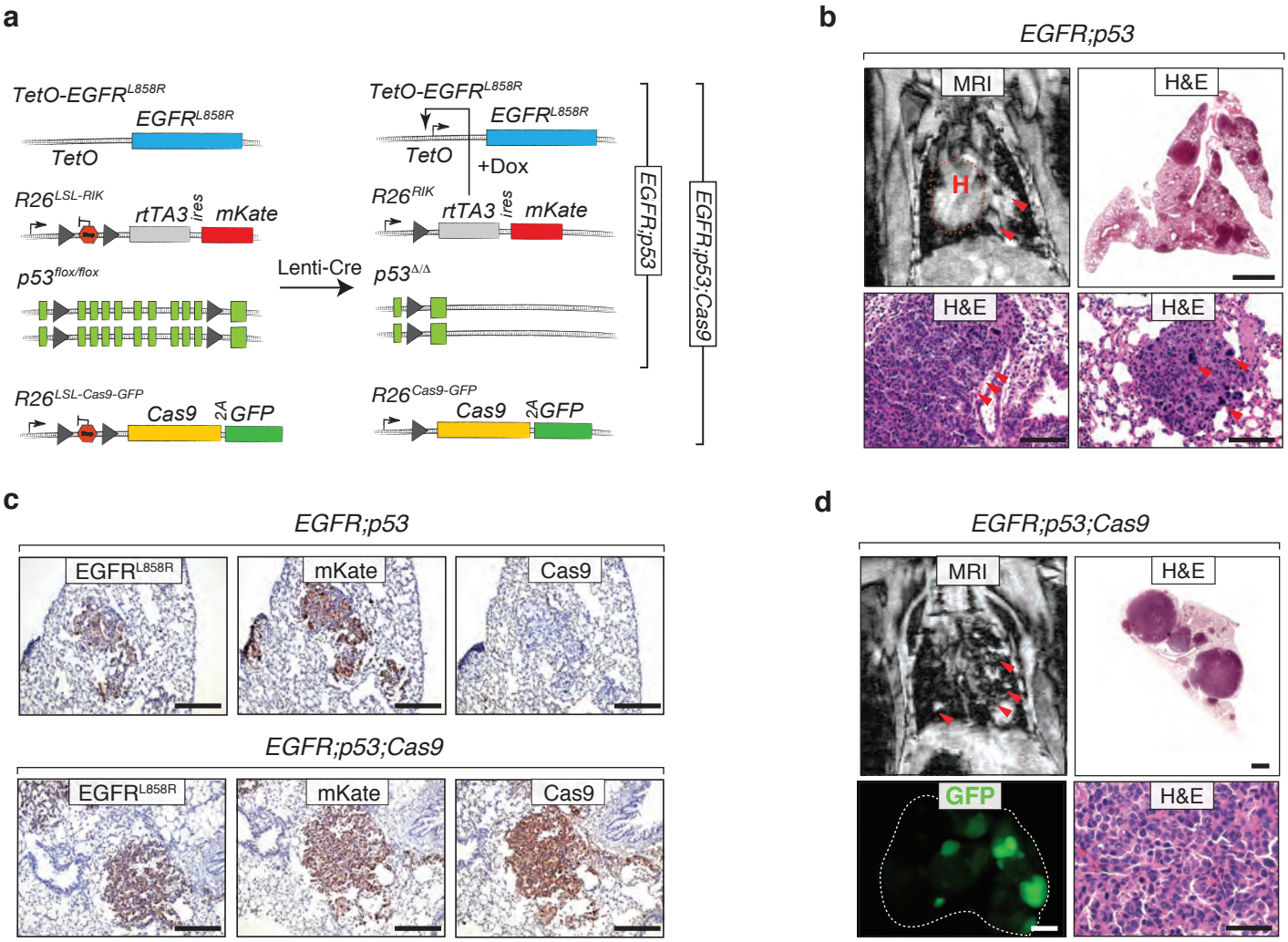
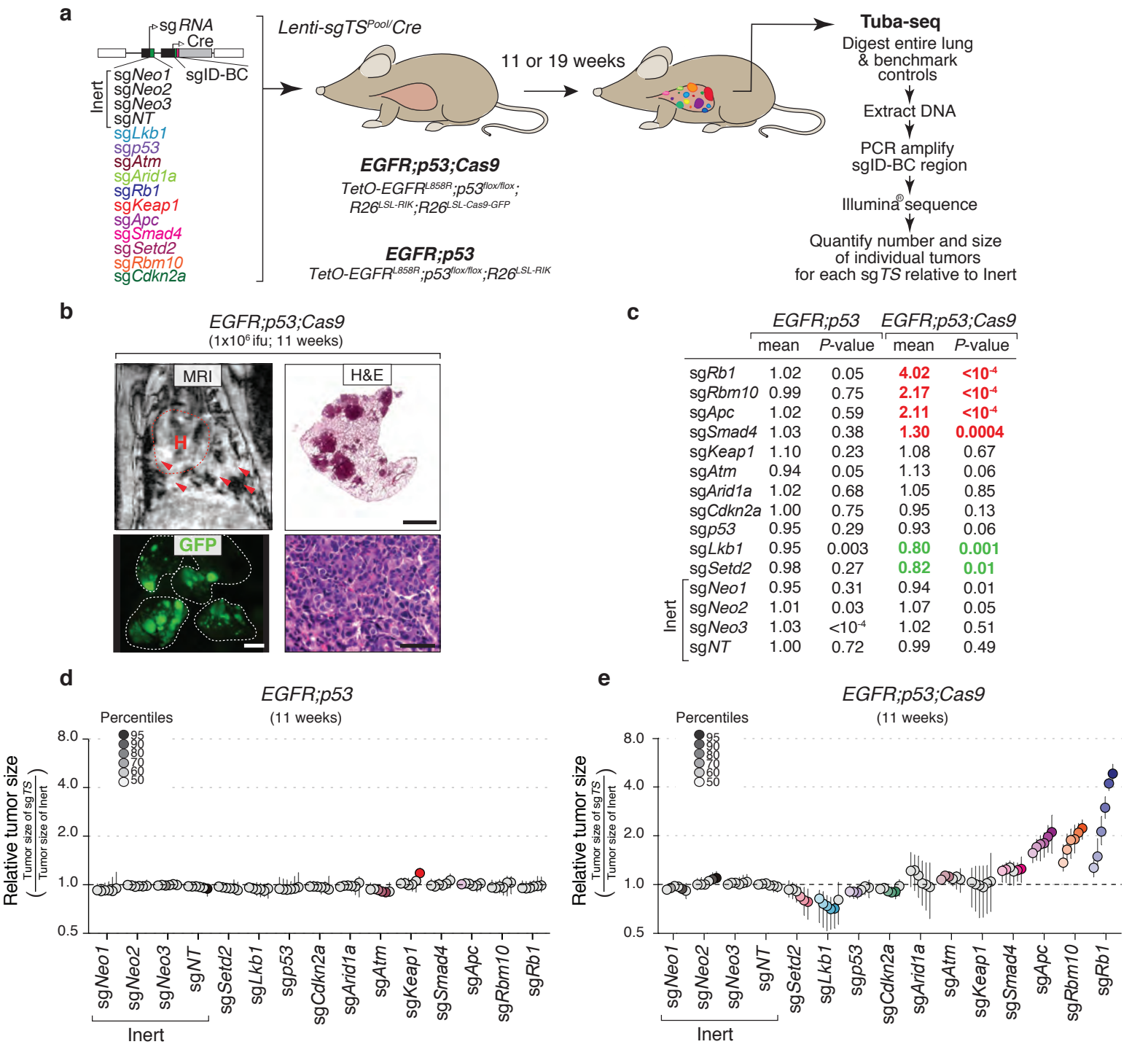


Fig. 1 | Lentiviral Cre-mediated lung tumor initiation in *EGFR;p53* and *EGFR;p53;Cas9* mice. a, Schematic of the *TetO-EGFR*^{L858R}, *R26*$^{LSL\text{-}RIK}$, and *p53*^{flox} alleles in *EGFR;p53* mice, prior to and following Cre-mediated recombination. Conditional expression of oncogenic EGFR^{L858R} is under the control of a tetracycline response element (*TetO*), which is induced by rtTA in the presence of doxycycline (dox). Lentiviral Cre inactivates *Trp53* and enables the expression of the reverse tetracycline-regulated transactivator (*rtTA3*) and mKate expression (from the *R26*$^{LSL\text{-}RIK}$ allele). Cre also allows expression of Cas9 and GFP (from the *R26*$^{LSL\text{-}Cas9\text{-}GFP}$ allele) in the *EGFR;p53;Cas9* model. **b**, MRI showing tumor development 11 weeks after tumor initiation in *EGFR;p53* mice (top left panel) with 2x10^6 ifu Lenti-Cre. The dashed red line surrounds the heart (H) and red arrows indicate areas of tumor. Hematoxylin and eosin (H&E) staining shows lung adenocarcinoma development. Red arrows indicate micropapillary structures and cancer cells with a high nuclear/cytoplasm ratio. Scale bars = 1.2 mm and 100 µm in the top and bottom panels, respectively. Images are from one representative mouse (N = 5). **c**, Immunostaining for EGFR^{L858R} and mKate in tumors in *EGFR;p53* and *EGFR;p53;Cas9* mice, and Cas9 in tumors in *EGFR;p53;Cas9* mice. Scale bars = 200 µm. **d**, MRI and H&E showing tumor development 16 weeks after tumor initiation in E*GFR;p53;Cas9* mice with 1x10^5 ifu Lenti-Cre. Scale bars = 1.2 mm and 100 µm on top and bottom-right panels, respectively. Tumors are positive for GFP in *EGFR;p53;Cas9* mice and tumor-bearing lungs are indicated by the white dashed line. Scale bar = 2.5 mm.



		EGFR;p53		EGFR;p53;Cas9	
		mean	*P*-value	mean	*P*-value
	sg*Rb1*	1.02	0.05	**4.02**	**<10^-4**
	sg*Rbm10*	0.99	0.75	**2.17**	**<10^-4**
	sg*Apc*	1.02	0.59	**2.11**	**<10^-4**
	sg*Smad4*	1.03	0.38	**1.30**	**0.0004**
	sg*Keap1*	1.10	0.23	1.08	0.67
	sg*Atm*	0.94	0.05	1.13	0.06
	sg*Arid1a*	1.02	0.68	1.05	0.85
	sg*Cdkn2a*	1.00	0.75	0.95	0.13
	sg*p53*	0.95	0.29	0.93	0.06
	sg*Lkb1*	0.95	0.003	**0.80**	**0.001**
	sg*Setd2*	0.98	0.27	**0.82**	**0.01**
Inert	sg*Neo1*	0.95	0.31	0.94	0.01
Inert	sg*Neo2*	1.01	0.03	1.07	0.05
Inert	sg*Neo3*	1.03	<10^-4	1.02	0.51
Inert	sg*NT*	1.00	0.72	0.99	0.49

Fig. 2 | Multiplexed somatic CRISPR/Cas9-mediated genome editing uncovers tumor suppressor gene function in *EGFR*-driven lung tumors. a, Experimental strategy. Tumors were allowed to develop for either 11 or 19 weeks in *EGFR;p53;Cas9* and *EGFR;p53* mice after intra-tracheal administration of Lenti-sg*TS*^{Pool}/*Cre*. Whole lungs were collected for tumor barcode deep sequencing and histology. The number of neoplastic cells in each tumor (tumor size) was calculated from barcode sequencing of bulk tumor-bearing lungs. Barcode read number was normalized to benchmark control cells that have known barcodes and were added at a known number to each sample (Methods). **b**, MRI, H&E, and GFP images showing tumor development in *EGFR;p53;Cas9* mice 11 weeks after tumor initiation with 1x10^5 ifu of Lenti-sg*TS*^{Pool}/*Cre*. Scale bars = 1.2 mm and 100 µm in top and bottom panels, respectively. Images are from one representative mouse (N = 10). **c**, Relative log-normal (LN) mean size of tumors with each sgRNA in *EGFR;p53* and *EGFR;p53;Cas9* mice 11 weeks after tumor initiation (normalized to the tumors with inert sgRNAs). *P*-values were calculated from bootstrapping. *P*-values < 0.05 and their corresponding means are highlighted in *EGFR;p53;Cas9* mice for sgRNAs that positively (red) and negatively (green) affect tumor growth. **d**, Relative sizes of tumors at the indicated percentiles of each genotype in *EGFR;p53* mice 11 weeks after tumor initiation with the Lenti-sg*TS*^{Pool}/*Cre*. These mice lack the *R26*$^{LSL\text{-}Cas9\text{-}GFP}$ allele; therefore, all sgRNAs are functionally inert. **e**, Relative sizes of tumors of each genotype in *EGFR;p53;Cas9* mice 11 weeks after tumor initiation with the Lenti-sg*TS*^{Pool}/*Cre*. The relative sizes of tumors at the indicated percentiles were calculated from the tumor size distribution of all tumors from ten mice. 95% confidence intervals are shown. Percentiles that are significantly different from the tumors with inert sgRNAs, are in color.

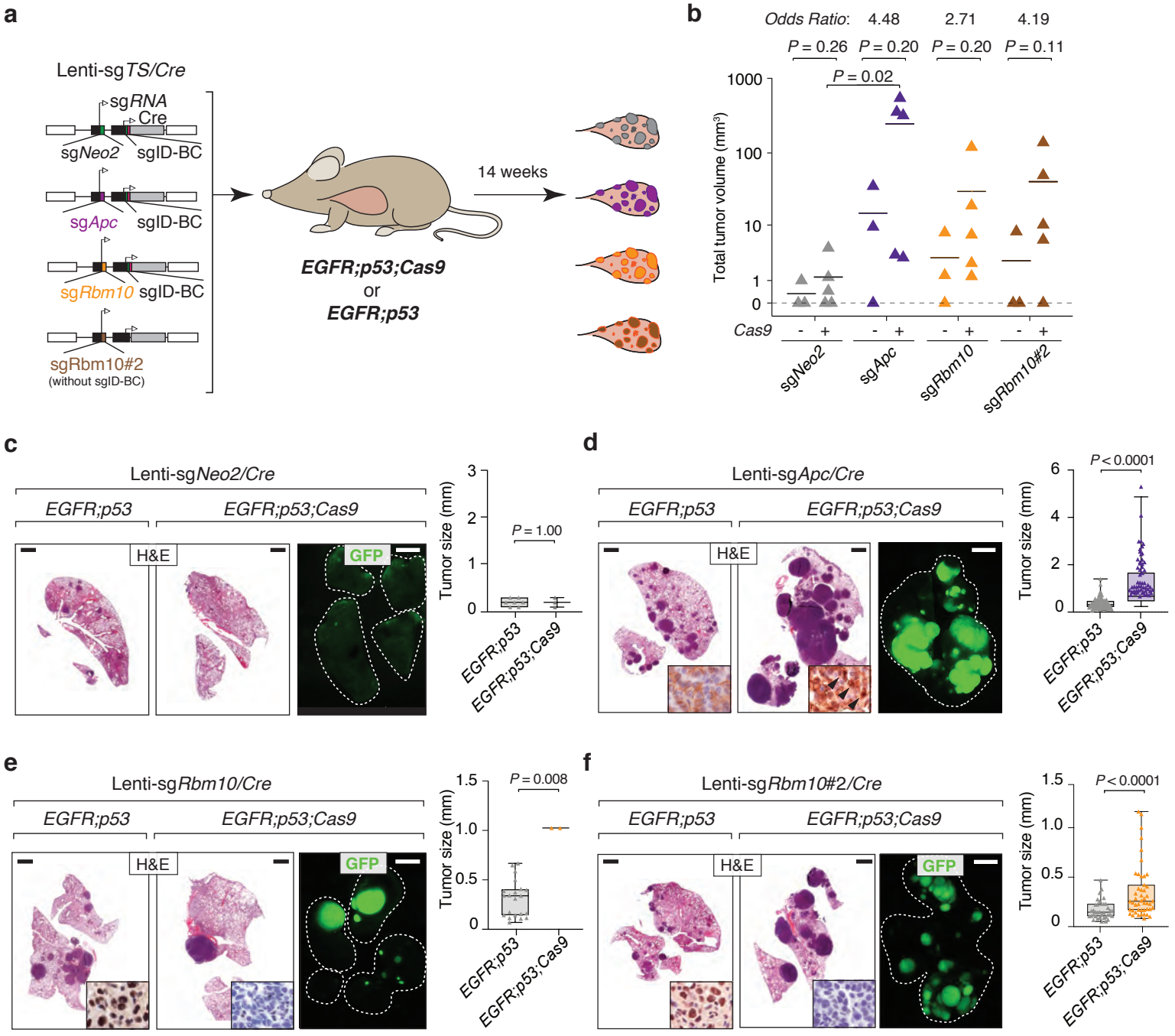
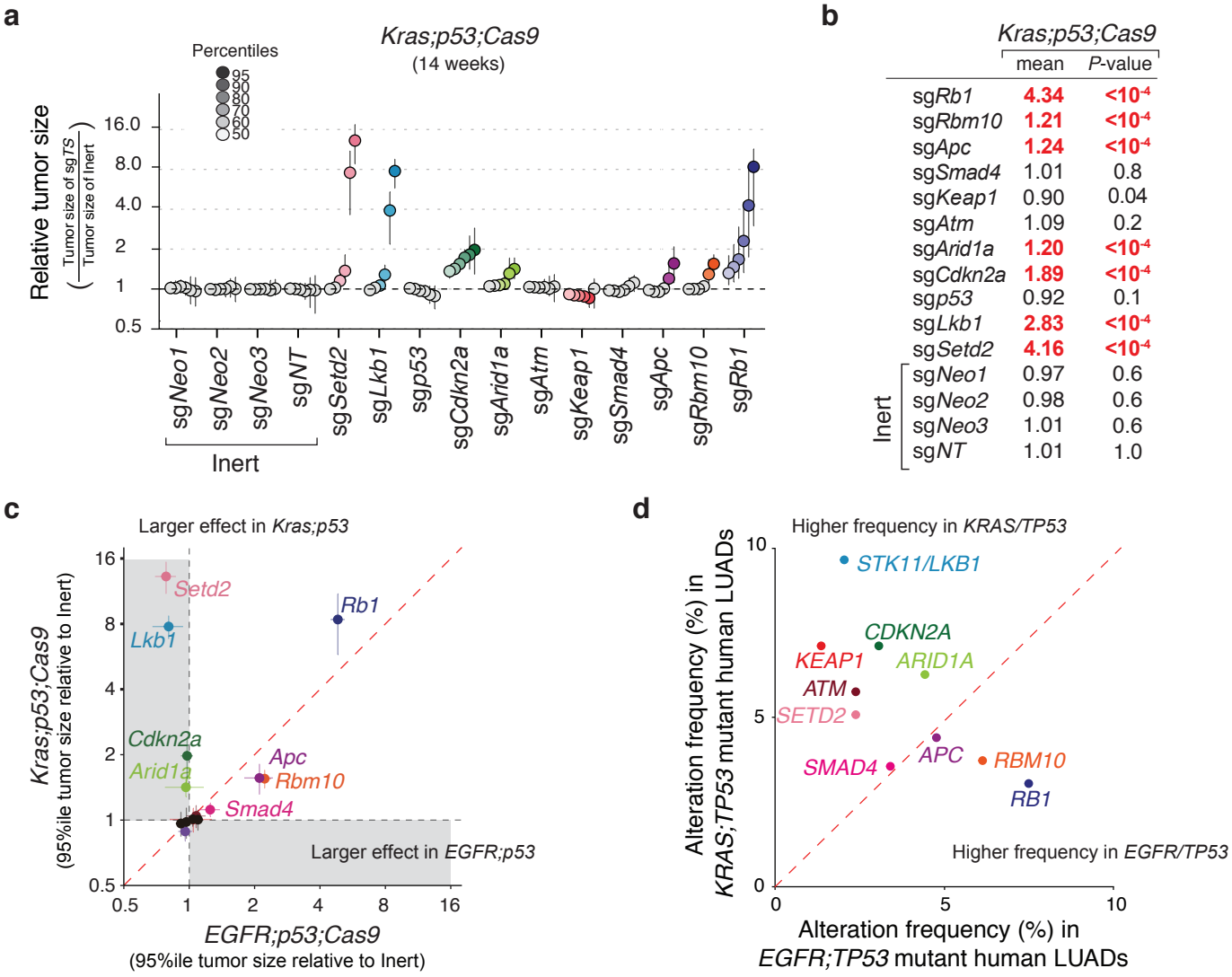


Fig. 3 | *Apc* and *Rbm10* inactivation enhances the growth of oncogenic *EGFR*-driven lung tumors in vivo. a, Experimental strategy for tumor suppressor gene validation. Tumors were initiated in *EGFR;p53* and *EGFR;p53;Cas9* mice with Lenti-sg*TS/Cre* vectors carrying sg*Apc*, sg*Rbm10*, sg*Rbm10#2*, or sg*Neo2* as a control (*EGFR;p53* $N = 3$ mice/group, *EGFR;p53;Cas9* $N = 5$ mice/group, $1x10^5$ ifu/mouse). **b**, Total tumor burden 14 weeks after tumor initiation as quantified by MRI in *EGFR;p53* and *EGFR;p53;Cas9* mice in which tumors were initiated with the indicated vectors. *P*-values were calculated using a one-tailed Mann-Whitney U test. The odds ratio is a measure of the magnitude of increase in tumor volume conferred by inactivation of the gene of interest. **c-f**, H&E and GFP quantification show that *Apc* (**d**) and *Rbm10* (**e, f**) inactivation enhance tumor growth (**c**). Tumor size was calculated by measuring the longest diameter of each tumor. Scale bars = 1.2 mm and 2.5 mm for histology and GFP images, respectively. The left and auxiliary lobes were analyzed using histology. *P*-values were calculated using the Mann-Whitney U test. The panels on the bottom right of the H&E images (**d-f**) show nuclear accumulation of β-catenin in tumors in *EGFR;p53;Cas9* mice following *Apc* inactivation (**d**) whereas, Rbm10 protein expression is greatly reduced or absent after *Rbm10* targeting in *EGFR;p53;Cas9* mice (**e, f**).



e

LKB1/STK11	wt	mut	
$EGFR^{onc}$ $TP53^{mut}$	287	6	2.0%
$KRAS^{onc}$ $TP53^{mut}$	515	57	10.0%

Odds ratio = 5.29
$P = 6\times10^{-6}$

SETD2	wt	mut	
$EGFR^{onc}$ $TP53^{mut}$	212	7	3.2%
$KRAS^{onc}$ $TP53^{mut}$	370	30	7.5%

Odds ratio = 2.46
$P = 0.03$

RBM10	wt	mut	
$EGFR^{onc}$ $TP53^{mut}$	137	18	11.6%
$KRAS^{onc}$ $TP53^{mut}$	206	22	9.6%

Odds ratio = 0.81
$P = 0.61$

APC	wt	mut	
$EGFR^{onc}$ $TP53^{mut}$	279	14	4.8%
$KRAS^{onc}$ $TP53^{mut}$	546	26	4.5%

Odds ratio = 0.95
$P = 0.87$

Fig. 4 | Dramatic differences in the effects of tumor suppressor genes in oncogenic *EGFR*- versus *Kras*-driven lung cancer. a, Tumors were initiated in *Kras;p53;Cas9* mice with Lenti-sg*TS*^{Pool}/*Cre* virus (1x10^6 ifu) and collected for Tuba-seq preparation 14 weeks after tumor initiation ($N = 10$). **b**, Relative size of tumors of each genotype in *Kras*^{G12D};*p53*$^{flox/flox}$;*H11*$^{LSL\text{-}Cas9}$ (*Kras;p53;Cas9*) mice 14 weeks after tumor initiation with Lenti-sg*TS*^{Pool}/*Cre* relative to tumors with inert sgRNAs. The relative size of tumors at the indicated percentiles was calculated from the tumor size distribution of all tumors from six mice. 95% confidence intervals are shown. **c**, The relative effect of inactivating each tumor suppressor gene on tumor size (comparison to tumors with inert sgRNAs) at the indicated percentiles in the *EGFR;p53;Cas9* and *Kras;p53;Cas9* models. Tumor suppressor genes that when inactivated have opposite effects in the two models are within the shaded gray areas of the plot. *Lkb1* and *Setd2* inactivation strongly increased tumor size only in the *Kras;p53;Cas9* model. **d**, The frequency of tumor suppressor gene alterations that co-occur with *EGFR* or *KRAS* and *TP53* mutations in human lung adenocarcinomas (data from AACR Project GENIE). **e**, *LKB1* and *SETD2* alterations co-occur significantly more frequently in *KRAS*^{onc};*TP53*^{mut} than in *EGFR*^{onc};*TP53*^{mut} human lung adenocarcinomas. The frequency of *RBM10* and *APC* alterations were not significantly different between in *KRAS*^{onc};*TP53*^{mut} and *EGFR*^{onc};*TP53*^{mut} tumors. The odds ratios represent the strength of the frequencies associated with tumor suppressor gene alterations that co-occur in *KRAS*^{onc};*TP53*^{mut} cases compared to *EGFR*^{onc};*TP53*^{mut} cases. *P*-values were calculated using a Fisher's exact test.

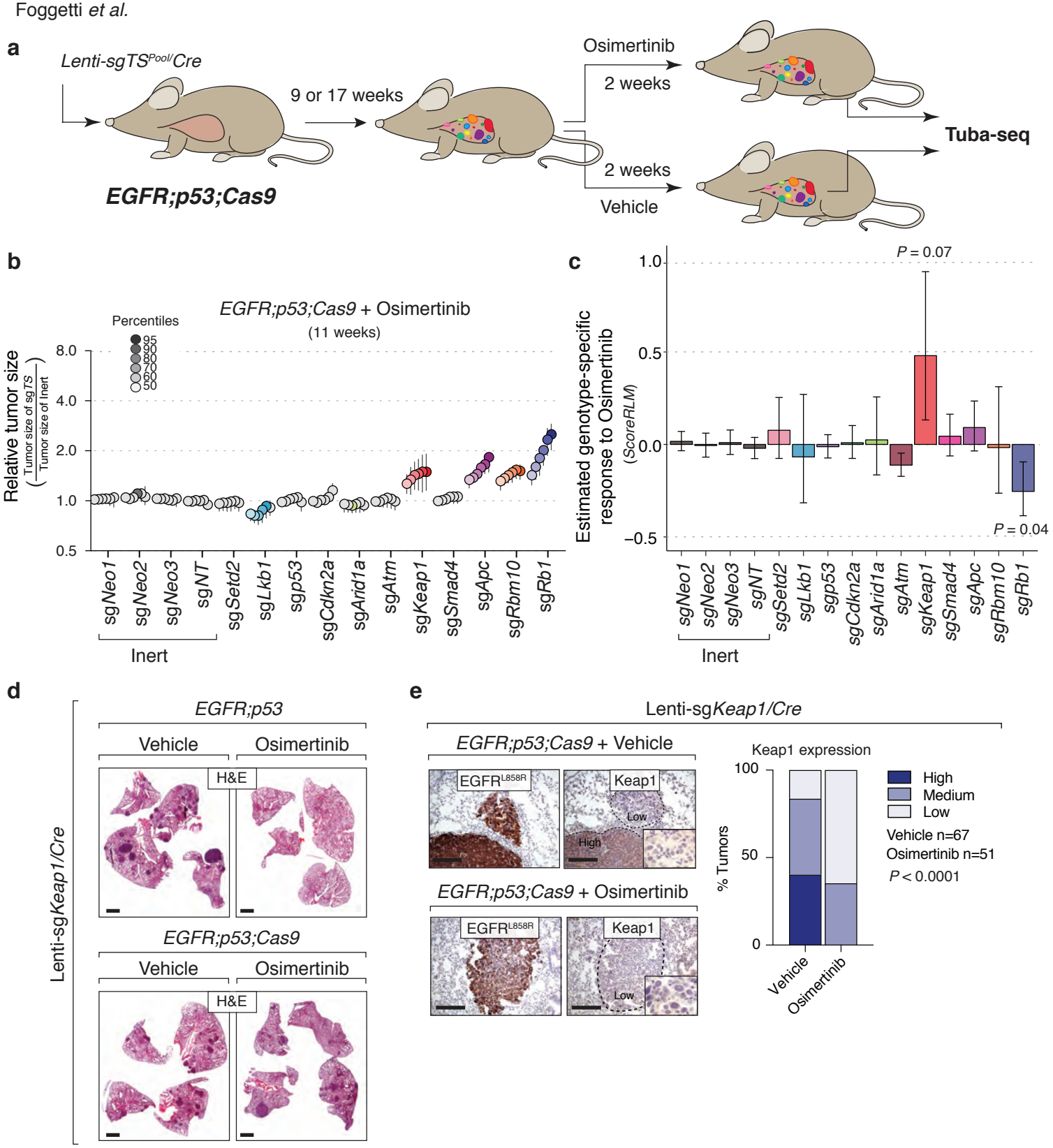


Fig. 5 | Identification of genotype-specific responses to osimertinib treatment. a, Experimental strategy. Tumors were initiated with Lenti-sg*TS*^{Pool}*/Cre* in *EGFR;p53;Cas9* mice for 9 or 17 weeks and were treated for two weeks with either vehicle or osimertinib (25 mg/kg, 5 days/week). **b**, Relative sizes of the tumors at the indicated percentiles for each genotype in *EGFR;p53;Cas9* mice after treatment with osimertinib. **c**, Estimate of the genotype-specific treatment response (*ScoreRLM*) calculated by comparing the LN mean of tumors treated with osimertinib to the LN mean of vehicle-treated tumors in *EGFR; p53;Cas9* mice 11 weeks after tumor initiation (Methods). Error bars indicate the standard deviation. **d**, H&E staining of tumor bearing-lungs in *EGFR;p53* and *EGFR;p53;Cas9* mice with Lenti-sg*Keap1/Cre* initiated tumors. **e**, Immunostaining of Lenti-sgLenti-sg*TS*^{Pool}*/Cre* initiated tumors in *EGFR;p53;Cas9* mice for EGFR^{L858R} and Keap1 after vehicle or osimertinib treatment. Tumors are still detectable after two weeks of treatment with osimertinib and mostly expressing medium/low levels of Keap1 compared to tumors treated with vehicle in *EGFR;p53;Cas9* mice (bar graph). Dashed lines delineate tumors and the level of Keap1 is indicated with a label (high/low). Scale bars = 1.2 mm and 200 μ m. *P*-values were calculated using a Chi-squared test.

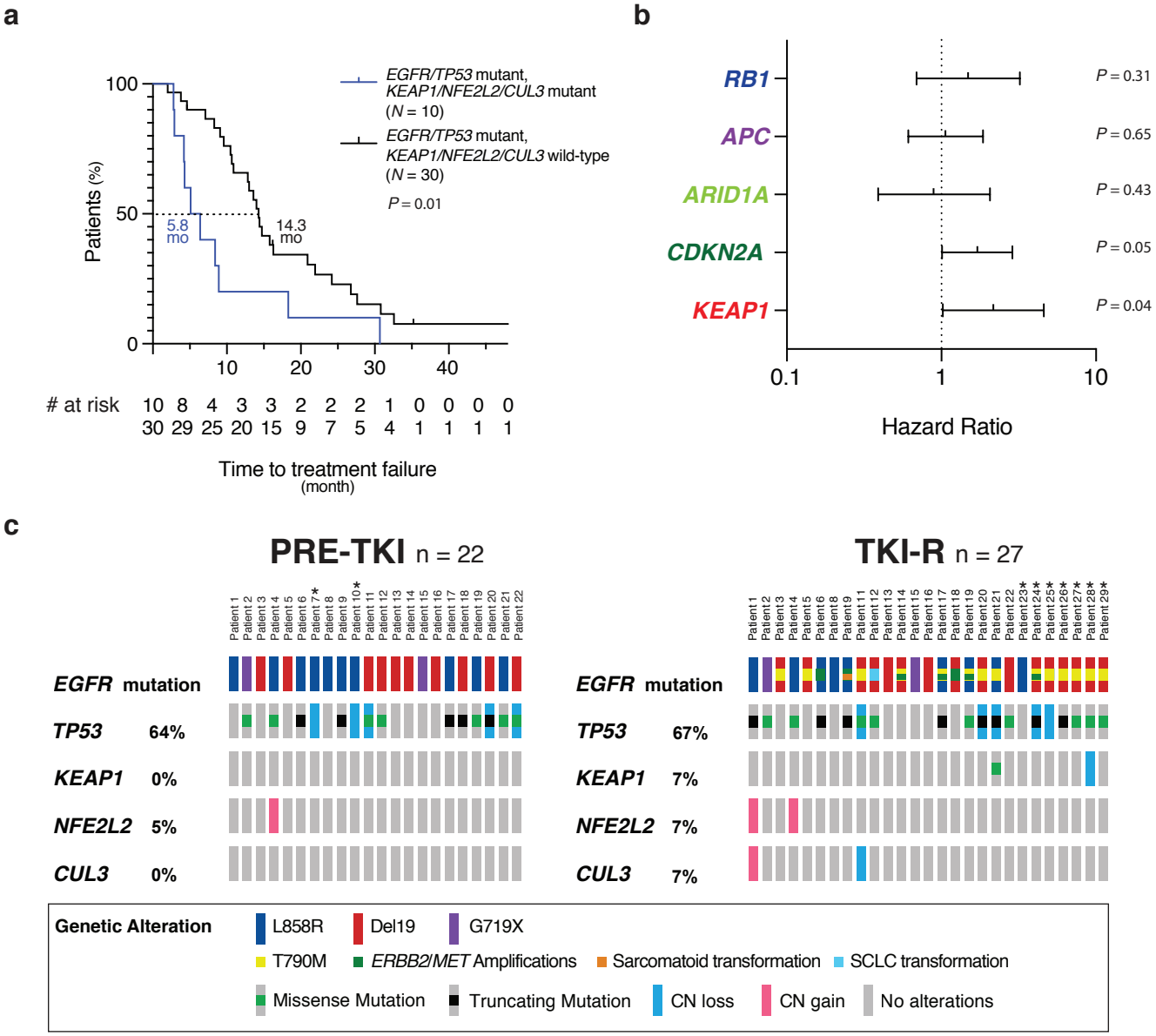


Fig. 6 | *KEAP1* inactivation correlates with reduced therapeutic response to TKIs in human *EGFR*-driven lung adenocarcinomas. a, Kaplan-Meier curve showing time to TKI treatment failure for patients with *EGFR/TP53* lung adenocarcinomas that do or not have mutations in the *KEAP1/NFE2L2/CUL3* axis. Time to treatment failure is the time from the initiation of TKI treatment to the date of discontinuation of TKI due to progression, toxicity or death. **b**, Forest plot of the time on treatment hazard ratios for tumors with *KEAP1*, *RB1*, *APC*, *ARIDA* or *CDKN2A* alterations in *EGFR/TP53* mutant lung adenocarcinomas. Hazard ratios (HR) were calculated from multivariate regression analysis, 95% confidence intervals are shown. **c**, *KEAP1, NFE2L2*, and *CUL3* alteration frequency in oncogenic *EGFR* and *TP53* mutant lung tumors in the Yale Lung Rebiopsy Program dataset (Methods). OncoPrint of oncogenic *EGFR* samples prior to first-line TKI (erlotinib, gefitinib, afatinib) treatment (PRE-TKI, left panel) ($N = 22$), and at acquired resistance to TKIs (TKI-R, right panel) ($N = 27$). In each row, the percentage and the type of alterations in each gene are indicated for each patient. For the TKI-R samples, known resistance mechanisms such as acquisition of T790M in *EGFR*, are reported. Cases having paired pre- and resistant samples are indicated by the patient number. Unpaired samples are indicated with an asterisk (*).