### **Supplementary Methods and Statistical Analysis**

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

###  A diagram describing the analysis steps that are part of the Tuba-seq method in more detail are summarized in **Supplementary Fig. S2A**. We used stringent filtering to identify the sgID-BC region that minimizes PCR and sequencing error, as previously described (56). Specifically, when compared with the previous Tuba-seq method (9), we now 1) prepare each vector separately to eliminate template switching, 2) use Q5 polymerase for PCR reactions to reduce PCR errors, 3) use unique dual indexing, design new primer sequences, and paired-end sequencing to minimize index hopping and sequencing errors for amplicon sequencing, and 4)required the absence of mismatches in the barcode region between the forward and reverse reads and removed all spurious tumors with the barcodes within two nucleotides from that of another larger tumor (56). The absolute number of cells in each tumor was calculated by scaling its sgID-BC read number with the mean read number of three spiked-in cell lines with a known absolute cell number of 5x105.

### **Summary statistics for tumor size distributions**

### As sequencing depth and PCR efficiency vary across libraries, we focused on tumors that we can repeatedly identify with high confidence, which are tumors over 500 cells as quantified by comparing technical replicates. We used multiple summary statistics to describe the truncated distribution of tumor sizes for all tumors larger than 500 cells. Percentiles and LN means were calculated as two summary statistics. Percentiles are a nonparametric summary of the distribution by taking the 50th, 60th, 70th, 70th, 90th, and 95th percentile of the distribution. The LN mean calculates the maximum likelihood estimator of the mean tumor sizes assuming a 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 of inactivating 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 taking the fold change of 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 median tumor number of tumors with the inert control sgRNAs larger than 1,000 cells across all osimertinib-treated mice.

### **Estimation of genotype-specific treatment responses**

### We calculated the expected size distribution of tumors after treatment assuming no genotype-specific treatment responses by reducing all tumors in the vehicle-treated mice by the estimated drug effect (*K*). Then we calculated the genotype-specific treatment response for each sg*TS* by comparing the relative LN mean of all tumors in the osimertinib-treated mice and the relative LN mean of all tumors calculated from the expected distribution after treatment. The genotype-specific treatment response is calculated as the log2 ratio of the observed relative LN mean by the expected LN mean, and we named it as *ScoreRLM*. We focus on tumors with the inert control sgRNAs that are over 1,000 cells in untreated mice and take out comparable proportions of tumors with each sgRNA from each vehicle- and osimertinib-treated mice based on the estimated treatment effect and the proportion of tumors carrying each sg*TS*. The *ScoreRLM* is calculated as:

$$RLM\_{i,j}=Log\_{2}\left(\frac{^{LN mean\_{i, osimertinib}}/\_{LN mean\_{Inert, osimertinib}}}{^{LN mean\_{i, vehicle}}/\_{LN mean\_{Inert, vehicle}}}\right)$$

$$ScoreRLM\_{i,j}=2^{RLM\_{i,j}}-1$$

### where $LN mean\_{i,osimertinib}$ is the LN mean for tumors containing sgID *i* in osimertinib-treated group, $LN mean\_{Inert,osimertinib}$ is the LN mean for all tumors containing one of the four inert sgIDs in the osimertinib-treated group. Similarly, $LN mean\_{i,vehicle}$ is the LN mean for tumors containing sgID *i* in vehicle-treated group and $LN mean\_{Inert,vehicle}$ is the LN mean for all tumors 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 –∞ to +∞ 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 +∞, 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:

$$ScoreRGM\_{i,j}=Log\_{2}\left(\frac{^{geometric mean\_{i, osimertinib}}/\_{geometric mean\_{Inert, osimertinib}}}{^{geometric mean\_{i, vehicle}}/\_{geometric mean\_{Inert, vehicle}}}\right)$$

$$ScoreRGM\_{i,j}=2^{RGM\_{i,j}}-1$$

$$RTN\_{i,j}=Log\_{2}\left(\frac{^{tumor number\_{i, osimertinib}}/\_{tumor number\_{Inert, osimertinib}}}{^{tumor number\_{i, vehicle}}/\_{tumor number\_{Inert, vehicle}}}\right)$$

$$ScoreRTN\_{i,j}=2^{RTN\_{i,j}}-1$$

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(56).

### 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 size 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 +∞ 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 (**Supplementary Figs. S11E, 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 second scenario of uniform responses, all tumors with the sg*TS* are assigned to be 50% larger than expected, i.e., reduced in size by 62.5% instead of 75% after treatment by osimertinib. The same sample sizes of ten vehicle- and ten osimertinib-treated mice were generated from bootstrapping mice and then bootstrapping tumors from each mouse prior to and after the simulated effects of drug responses and genotype - specific responses, respectively. To calculate the false discovery rate, we also simulated another scenario where no preassigned genotype-specific responses exist and all tumors, regardless of genotypes, were reduced in size uniformly by 75% by osimertinib. The therapeutic sensitivity was calculated as the probability of re-identifying the preassigned genotype-specific treatment responses given the cutoff of *P*-values, while specificity was calculated as the probability of falsely identifying genotype-specific treatment responses without any input signal of genotype-specific treatment responses given the cutoff of *P*-values. A total of 100 runs of simulations of the 11 tumor suppressor genes (including Tp53; 1,100 cases) were performed for each scenario. We further plotted the receiver operating characteristic (ROC) curves for the three metrics by varying the cutoff for the *P*-values. Statistical analyses were performed using R3.6.