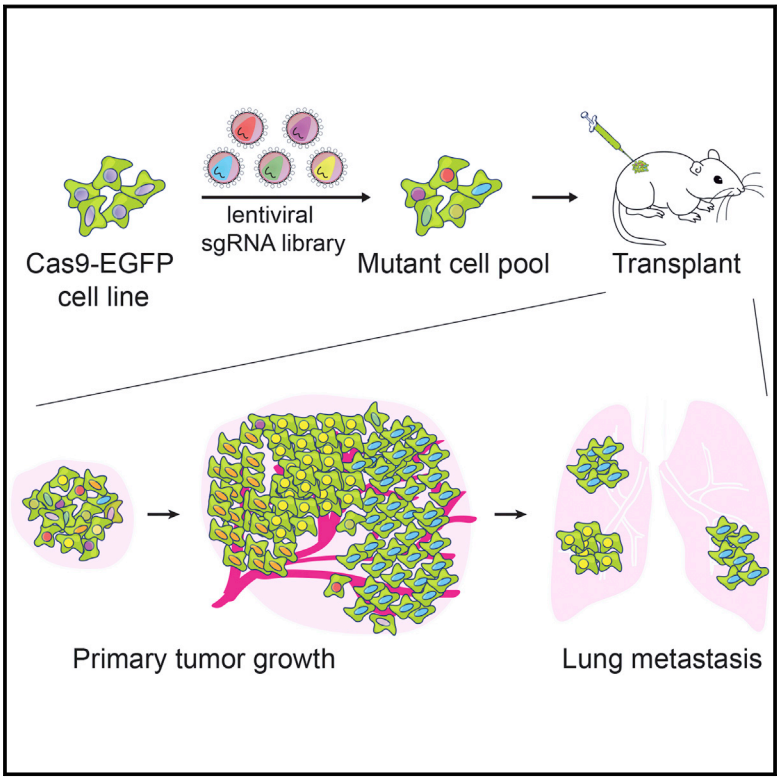
Resource

Cell

Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis

Graphical Abstract



Authors

Sidi Chen, Neville E. Sanjana, ...,
Feng Zhang, Phillip A. Sharp

Correspondence

zhang@broadinstitute.org (F.Z.),
sharppa@mit.edu (P.A.S.)

In Brief

Using an in vivo genome-wide CRISPR/Cas9 screen, loss-of-function mutations that drive tumor growth and metastasis to the lung have been identified, demonstrating Cas9-based screening as a robust method to systematically assay gene phenotypes in cancer evolution.

Highlights

- Genome-wide in vivo CRISPR-Cas9 screen in mice reveals genes regulating lung metastasis
- Screen identifies loss-of-function mutations in known tumor suppressors and novel genes
- Candidate metastasis genes are validated using a pooled competition assay
- Effect of mutations on primary tumor growth positively correlates with metastasis

Chen et al., 2015, Cell *160*, 1246–1260
March 12, 2015
http://dx.doi.org/10.1016/j.cell.2015.02.038

CellPress

Cell

Resource

Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis

Sidi Chen,[1,2,3,10] Neville E. Sanjana,[3,4,5,6,7,10] Kaijie Zheng,[3,4] Ophir Shalem,[3,4] Kyungheon Lee,[8] Xi Shi,[3,4] David A. Scott,[3,4] Jun Song,[8] Jen Q. Pan,[3,4] Ralph Weissleder,[8,9] Hakho Lee,[8] Feng Zhang,[3,4,5,6,7,*] and Phillip A. Sharp[1,2,*]
[1]David H. Koch Institute for Integrative Cancer Research, Cambridge, MA 02139, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]Stanley Center for Psychiatric Research, Cambridge, MA 02142, USA
[5]McGovern Institute for Brain Research
[6]Department of Brain and Cognitive Sciences
[7]Department of Biological Engineering
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]Center for Systems Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[9]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[10]Co-first author
*Correspondence: zhang@broadinstitute.org (F.Z.), sharppa@mit.edu (P.A.S.)
http://dx.doi.org/10.1016/j.cell.2015.02.038

SUMMARY

Genetic screens are powerful tools for identifying genes responsible for diverse phenotypes. Here we describe a genome-wide CRISPR/Cas9-mediated loss-of-function screen in tumor growth and metastasis. We mutagenized a non-metastatic mouse cancer cell line using a genome-scale library with 67,405 single-guide RNAs (sgRNAs). The mutant cell pool rapidly generates metastases when transplanted into immunocompromised mice. Enriched sgRNAs in lung metastases and late-stage primary tumors were found to target a small set of genes, suggesting that specific loss-of-function mutations drive tumor growth and metastasis. Individual sgRNAs and a small pool of 624 sgRNAs targeting the top-scoring genes from the primary screen dramatically accelerate metastasis. In all of these experiments, the effect of mutations on primary tumor growth positively correlates with the development of metastases. Our study demonstrates Cas9-based screening as a robust method to systematically assay gene phenotypes in cancer evolution in vivo.

INTRODUCTION

Cancer genomes have complex landscapes of mutations and diverse types of genetic aberrations (Lawrence et al., 2013; Weinberg, 2007). A major challenge in understanding the cancer genome is to disentangle alterations that are driving the processes of tumor evolution from passenger mutations (Garraway and Lander, 2013). Primary tumor growth and metastasis are distinct yet linked processes in the progression of solid tumors (Nguyen et al., 2009; Valastyan and Weinberg, 2011; Vanharanta and Massagué, 2013). It has been observed in the clinic that the probability of detecting metastases in a patient correlates positively with the size of a primary tumor (Heimann and Hellman, 1998). Several possible explanations have been suggested: metastatic properties may only be acquired in late-stage tumors, larger tumors may seed proportionally more cells into circulation that eventually migrate to other sites, or cells with a strong ability to proliferate may also have enhanced ability to metastasize (Weinberg, 2007). In early studies using random insertional mutagenesis, it was observed that metastatic cell subpopulations overgrow to complete dominance in the primary tumor, suggesting progressive selection at both sites (Korczak et al., 1988; Waghorne et al., 1988).

Genetic screens are powerful tools for assaying phenotypes and identifying causal genes in various hallmarks of cancer progression (Hanahan and Weinberg, 2011). RNAi and overexpression of open reading frames (ORFs) have been utilized for screening cancer genes in several models of oncogenesis in mice (Schramek et al., 2014; Shao et al., 2014; Zender et al., 2008). Recently, the Cas9 nuclease (Barrangou et al., 2007; Bolotin et al., 2005; Chylinski et al., 2013, 2014; Deltcheva et al., 2011; Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012; Sapranauskas et al., 2011) from the microbial type II CRISPR (clustered regularly interspaced short palindromic repeats) system has been harnessed to facilitate loss-of-function mutations in eukaryotic cells (Cong et al., 2013; Mali et al., 2013). When the Cas9 nuclease is targeted to specific locations in the genome, DNA cleavage results in double-stranded breaks (DSBs), which are repaired via non-homologous end-joining (NHEJ) (Rouet et al., 1994). NHEJ repair results in insertion or deletion (indel) mutations that can cause loss of function if the DSB occurs in a coding exon. The Cas9 nuclease can be guided to its DNA target by a single-guide RNA (sgRNA) (Jinek et al., 2012), a synthetic fusion between the CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) (Deltcheva et al., 2011). In cells, Cas9-mediated gene disruption requires the full-length tracrRNA (Cong et al., 2013; Mali et al., 2013), in which secondary structures at the 3′ end of tracrRNA are

critical for Cas9-mediated genome modification (Cong et al., 2013; Hsu et al., 2013).

Screens utilizing Cas9 have identified genes that are essential for cell survival and genes involved in drug resistance in various cell lines (Shalem et al., 2014; Wang et al., 2014; Koike-Yusa et al., 2014; Zhou et al., 2014). In vivo pooled screens are challenging due to many factors, such as the complexity of the library, limitations of virus delivery and/or cell transplantation, uniformity of viral transduction at a low MOI, and the complex dynamics and interactions of cells in animals. In this study, we report a genome-wide Cas9 knockout screen in a mouse model of tumor evolution. This screen provides a systematic phenotypic measurement of loss-of-function mutations in primary tumor growth and metastasis.

RESULTS

CRISPR/Cas9 Library-Mediated Mutagenesis Promotes Metastasis

We derived and cloned a cell line (Chen et al., 2014) from a mouse non-small-cell lung cancer (NSCLC) (Kumar et al., 2009). This cell line possesses an oncogenic *Kras* in conjunction with homozygous *p53* and heterozygous *Dicer1* loss of function (*Kras*$^{G12D/+}$;*p53*$^{-/-}$;*Dicer1*$^{+/-}$, denoted KPD) and is capable of inducing tumors when transplanted into immunocompromised mice (Chen et al., 2014; Kumar et al., 2009). We transduced this cell line with a lentivirus carrying a Cas9 transgene fused to a GFP and generated clonal cell lines (Cas9-GFP KPD) (Experimental Procedures) (Figures S1A and S1B). A clonal Cas9-GFP KPD cell line (clone 5) was selected to provide genetic and cellular homogeneity for subsequent screens.

We utilized a pooled genome-wide mouse sgRNA library (termed mouse genome-scale CRISPR knockout library A, or mGeCKOa) containing 67,405 sgRNAs targeting 20,611 protein-coding genes and 1,175 microRNA precursors in the mouse genome (Sanjana et al., 2014). The library also contains 1,000 control sgRNAs (termed non-targeting sgRNAs) designed to have minimal homology to sequences in the mouse genome (Sanjana et al., 2014; Shalem et al., 2014). We transduced the Cas9-GFP KPD cell line with the mGeCKOa library in three independent infection replicate experiments; for each replicate, the library representation (cells per lentiviral CRISPR construct) was greater than 400× (Figure 1A) (Experimental Procedures).

After in vitro culture for 1 week, we subcutaneously transplanted 3×10^7 cells into the flanks of immunocompromised *Nu/Nu* mice (Figure 1A). We transplanted the cells from each infection replicate into four mice, using one mouse for early tumor sequencing and three mice for sequencing of late-stage primary tumor and metastases (Figure 1A). Both mGeCKOa-transduced and untransduced Cas9-GFP KPD cells formed tumors at the injection site (Figure 1B). Like most subcutaneously transplanted tumors, these tumors were poorly differentiated. The primary tumors induced by mGeCKOa-transduced cells grew slightly faster than tumors from the untransduced cells at an early stage (Figure 1C) (2 weeks post-transplantation) (paired two-tailed t test, $p = 0.05$), but at late stages all tumors were similar in size (paired two-tailed t test, $p = 0.18$ for data at 4 weeks, $p = 0.6$ for data at 6 weeks) (Figure 1C).

At 6 weeks post-transplantation, we imaged the mice using micro-computed tomography (μCT) and found tumors in the lungs of the mice transplanted with mGeCKOa-transduced Cas9-GFP KPD cells (mGeCKOa mice), but not in the mice transplanted with untransduced Cas9-GFP KPD cells (control mice) (Figure 1D, Figure S1C). Mice were sacrificed and examined for metastases in various organs. Under a fluorescent stereoscope at 6× magnification, metastases were visually detected in the lung in 89% (8/9) of the mGeCKOa mice (Figure S1D). The mGeCKOa mice on average had 80% of their lung lobes positive for metastases (Figure 1E). In contrast, none (0/3) of the control mice developed detectable metastases in the lung (Figure 1E). At this time, metastases were not detected in the liver, kidney, or spleen in either group (Figure 1F). These data indicated that mGeCKOa library transduction enhanced the ability of the Cas9-GFP KPD cells to form metastases in the lung.

Dynamic Evolution of sgRNA Library Representation during Tumor Growth and Metastasis

To investigate the sgRNA representation through different stages of tumor evolution and to identify genes where loss of function confers a proliferative or metastatic phenotype, we used deep sequencing to readout the sgRNA representation (see Data S1 in Dataset S1). At 6 weeks post transplantation, we sequenced the late-stage primary tumor and three random lobes from the lung of each of the nine mGeCKOa mice (Figure 1A) (Experimental Procedures). In parallel, we also sequenced the mGeCKOa input plasmid library, the pre-transplantation mGeCKOa-transduced Cas9-GFP KPD cells (cultured in vitro for 7 days after transduction), and early-stage primary tumors (2 weeks post transplantation, one mouse from each infection replicate). In the cell samples, the sgRNA representations showed high concordance between technical replicates (correlation, $\rho = 0.95$ on average, $n = 3$) and biological infection replicates (correlation, $\rho = 0.84$ on average, $n = 3$) (Figures 2A, S2A, S2B, and S2E). The sgRNA representation of cell samples correlates highly with the plasmid representation (correlation, $\rho = 0.93$ on average, $n = 3$) (Figures 2A, S2C, and S2E). Furthermore, different sgRNAs that target the same gene are correlated in terms of rank change (correlation, $\rho = 0.49$ on average, $n = 3$) (Figure S2D). Using gene set enrichment analysis (GSEA), we found that the sgRNAs with significantly decreased abundance in cells compared to plasmid are enriched for genes involved in fundamental cellular processes, such as ribosomal proteins, translation factors, RNA splicing factors, and RNA processing factors, indicating selection against the loss of these genes after 1 week in culture (Figure S2F).

To investigate the sgRNA library dynamics in different sample types (plasmid, pre-transplantation cells, early primary tumor, late primary tumor, and lung metastases), we compared the overall distributions of sgRNAs from all samples sequenced. Cell samples clustered tightly with each other and the plasmid, forming a cell-plasmid clade (Figures 2A and S2E). Early primary tumor samples also clustered with each other and then with the cell-plasmid clade, whereas late tumors and lung metastases clustered together in a distinct group (Figures 2A and S2E). The overlap of detected sgRNAs between different pre-transplantation infection replicates is over 95% (Figure S3A). The detected sgRNAs in the three infection replicates of early tumor

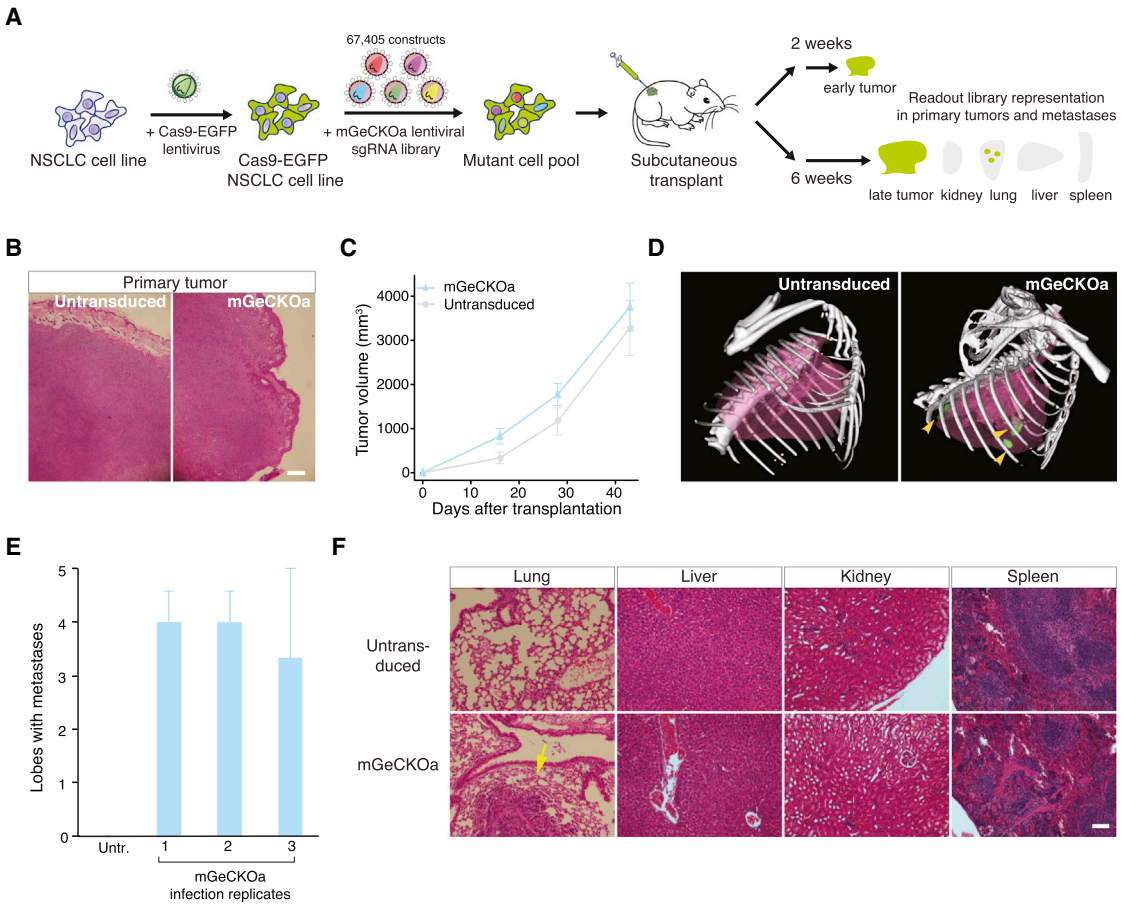


Figure 1. Tumor Growth and Metastasis in Transplanted Cas9-GFP KPD Cells with mGeCKOa Library

(A) Schematic representation of the loss-of-function metastasis screen using the mouse genome-scale CRISPR/Cas9 knockout library (mGeCKOa).

(B) Representative H&E stains of primary tumor from *Nu*/*Nu* mice subcutaneously transplanted with a Cas9-GFP *Kras*$^{G12D/+}$;*p53*$^{-/-}$;*Dicer1*$^{+/-}$ (KPD) NSCLC cell line that was either untransduced or transduced with the mGeCKOa lentiviral library. Scale bar, 200 μm.

(C) Primary tumor growth curve of *Nu*/*Nu* mice transplanted with untransduced cells (n = 3 mice) or mGeCKOa-transduced Cas9-GFP KPD cells (n = 9 mice). Error bars indicate SEM.

(D) MicroCT 3D reconstruction of the lungs of representative mice transplanted with control (untransduced) and mGeCKOa-transduced (mGeCKOa) cell pools. Lung metastases were identified and traced in each 2D section (green).

(E) Percent of lobes with metastases visible after dissection under a fluorescence stereoscope in *Nu*/*Nu* mice transplanted with untransduced Cas9-GFP KPD cells (n = 3 mice) or mGeCKOa-transduced Cas9-GFP KPD cells with three independent infection replicate experiments (1, 2, and 3; n = 3 mice per replicate). Error bars indicate SEM.

(F) Representative H&E stains from various organs of *Nu*/*Nu* mice subcutaneously transplanted with untransduced and mGeCKOa-transduced Cas9-GFP KPD cells. Yellow arrow indicates a lung metastasis. Scale bar, 40 μm.

See also Figure S1.

samples overlap 63%–76% with each other (Figure S3B). Early primary tumors retained less than half (32%–49%) of the sgRNAs found in the transplanted cell populations (Figures 2B, 2C, S3C, and S3D). Compared to the cell populations, sgRNAs whose targets are genes involved in fundamental cellular processes are further depleted in early tumors (Table S1).

Interestingly, only a small fraction of sgRNAs (less than 4% of all sgRNAs, or less than 8% of sgRNAs in the early primary tumor of the corresponding replicate) were detected in the late-stage primary tumor samples (Figures 2B, 2C, S3C, and S3D). The sgRNA diversity (i.e., number of different sgRNAs detected) further decreased in samples from lung metastases (Figures 2B, 2C, S3C, and S3D). The lung samples retained $\leq$0.4% of all sgRNAs in the mGeCKOa library, or $\leq$1.1% of sgRNAs found in the early primary tumor of the corresponding replicate, with a subset of highly enriched sgRNAs (Figures 2B, 2C, S3C, and S3D). The global patterns of sgRNA distributions in different sample types are distinct, as is evident in the strong shifts in the respective cumulative distribution functions (Kolmogorov-Smirnov [KS] test, $p < 10^{-15}$ for all pairwise comparisons) (Figure 2D).

Enriched sgRNAs in Primary Tumors

Late primary tumors retain few sgRNAs (on average 813 ± 264 sgRNAs, n = 9 mice), with even fewer at high frequencies

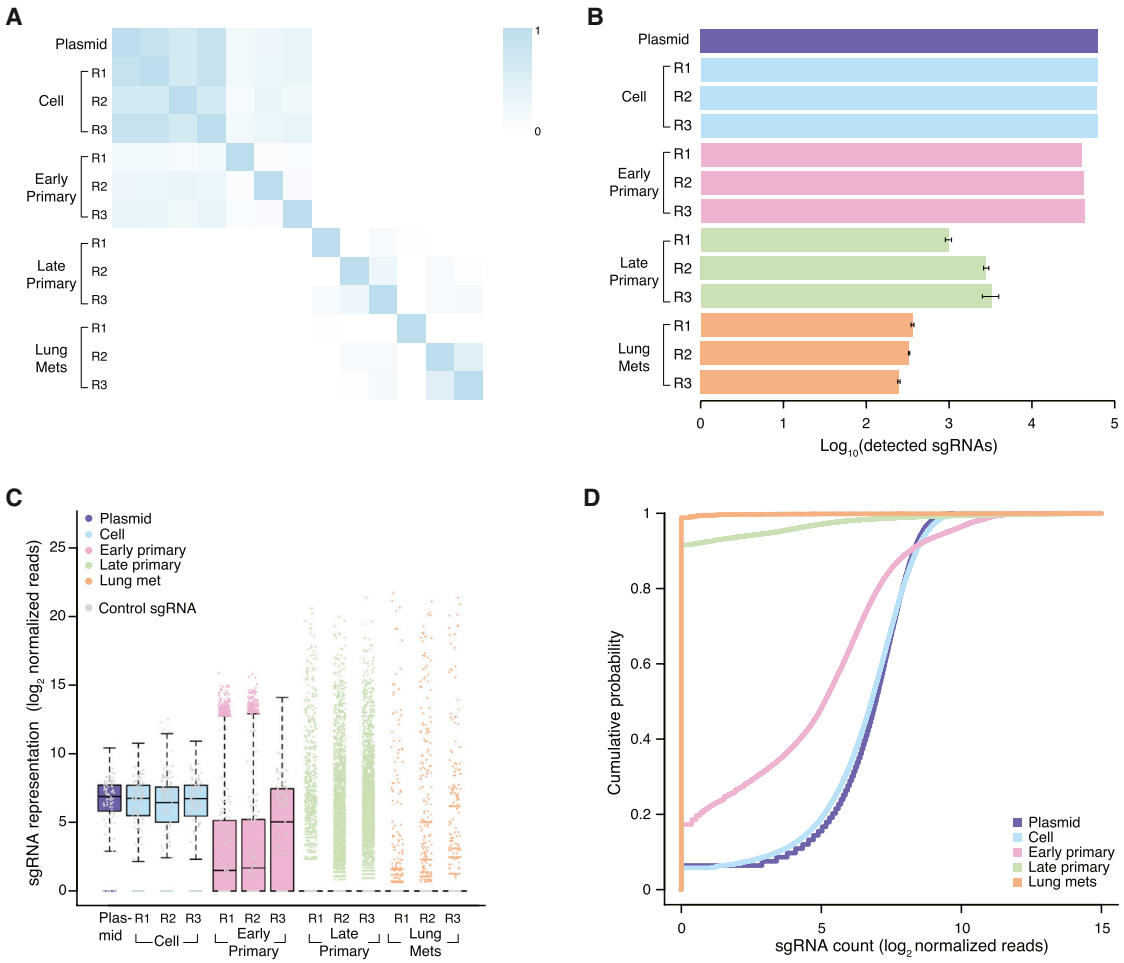


Figure 2. Representation of mGeCKOa Library at Different Stages of Tumor Growth and Metastasis

(A) Pearson correlation coefficient of the normalized sgRNA read counts from the mGeCKOa plasmid library, transduced cells before transplantation (day 7 after spinfection), early primary tumors (~2 weeks after transplantation), late primary tumors (~6 weeks after transplantation), and lung metastases (~6 weeks after transplantation). For each biological sample type, three independent infection replicates (R1, R2, and R3) are shown. n = 1 mouse per infection replicate for early primary tumors; n = 3 mice per infection replicate for late primary tumors and lung samples.

(B) Number of unique sgRNAs in the plasmid, cells before transplantation, early and late primary tumors, and lung metastases as in (A). Error bars for late primary tumors and lung metastases denote SEM for n = 3 mice per infection replicate.

(C) Boxplot of the sgRNA normalized read counts for the mGeCKOa plasmid pool, cells before transplantation, early and late primary tumors, and lung metastases as in (A). Outliers are shown as colored dots for each respective sample. Gray dots overlaid on each boxplot indicate read counts for the 1,000 control (non-targeting) sgRNAs in the mGeCKOa library. Distributions for late primary tumors and lung metastases are averaged across individual mice from the same infection replication.

(D) Cumulative probability distribution of library sgRNAs in the plasmid, cells before transplantation, early and late primary tumors, and lung metastases as in (A). Distributions for each sample type are averaged across individual mice and infection replications.

See also Figures S2 and S3.

(4 ± 1 sgRNAs with >5% of total reads) in each mouse (Figures 2B, 2C, S2C, S2D, 3A, and S4H). We used three methods to identify enriched sgRNAs in late primary tumors: (1) sgRNAs above a certain threshold, (2) top-ranked sgRNAs in the tumor of each mouse, and (3) using false discovery rate (FDR), i.e., sgRNAs enriched compared to the distribution of the 1,000 non-targeting sgRNAs. All three methods generated similar results (Figure S4A). Taking the results from (3) as an example, a total of 935 sgRNAs (targeting 909 genes) are enriched over the non-targeting controls (FDR cutoff = 0.2%) in the late primary tumor of one or more mice (Figures 3B and 3C). These sgRNAs are targeting genes highly enriched in apoptosis pathways (Table S2), with many of them being pro-apoptotic, such as BH3 interacting-domain death agonist (*Bid*), phosphatase and tensin homolog (*Pten*), cyclin-dependent kinase inhibitor 2a (*Cdkn2a*), and O-6-methylguanine-DNA methyltransferase (*Mgmt*), suggesting strong selection for mutations that inactivate apoptosis in primary tumor cells.

We identified 24 candidate genes that were targeted by two or more independent sgRNAs enriched in late primary tumors

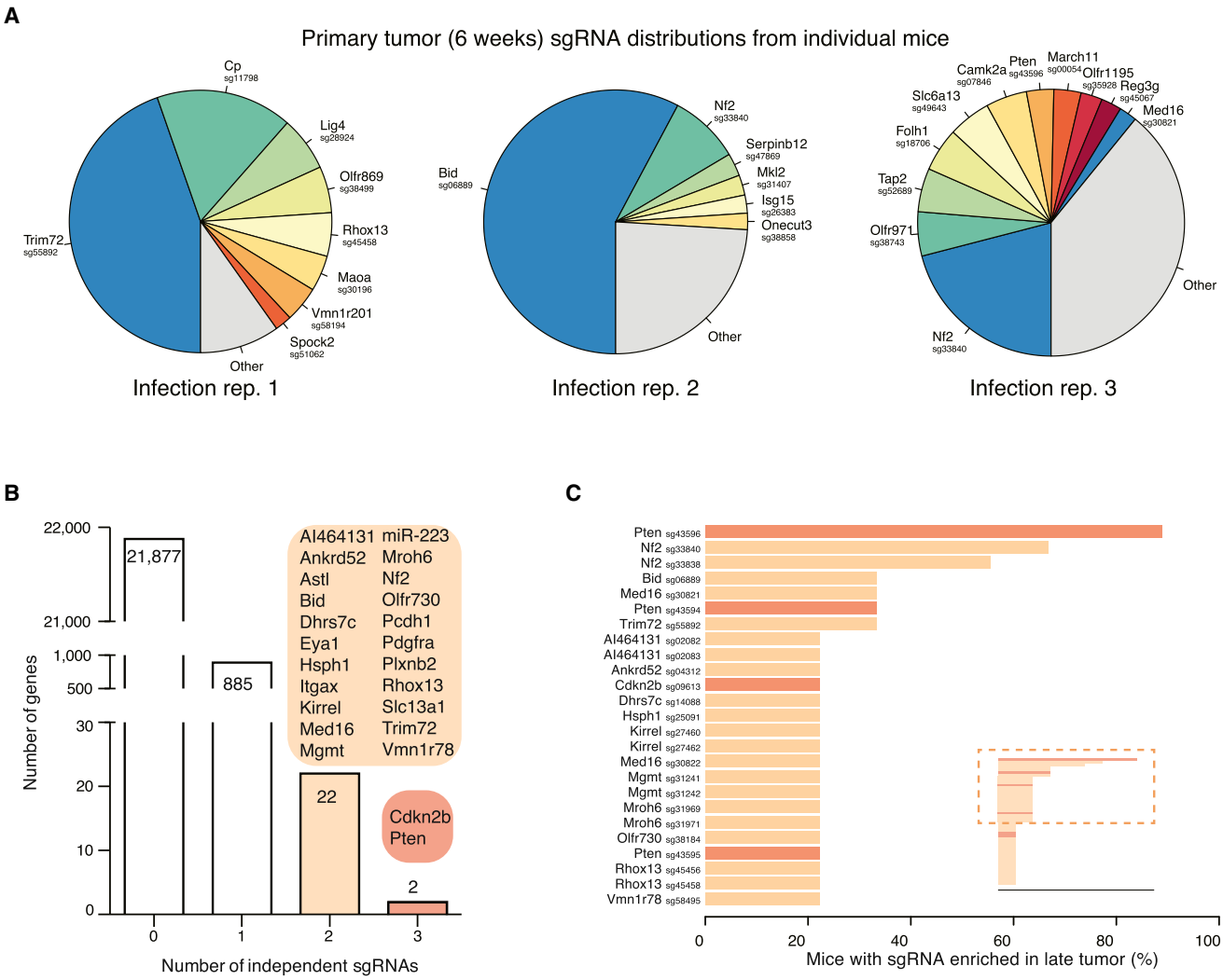


Figure 3. Enriched sgRNAs from the mGeCKOa Screen in Primary Tumors

(A) Pie charts of the most abundant sgRNAs in the primary tumors (at ~6 weeks post-transplantation) of three representative mice (one from each replicate mGeCKOa infection). The area for each sgRNA corresponds to the fraction of total reads from the primary tumor for the sgRNA. All sgRNAs with ≥2% of total reads are plotted individually.

(B) Number of genes with 0, 1, 2, or 3 significantly enriched (FDR < 0.2% for at least one mouse) mGeCKOa sgRNAs targeting that gene. For genes/miRs with 2 or more enriched sgRNAs, genes/miRs are categorized by how many sgRNAs targeting that gene/miR are enriched as indicated in the colored bubbles adjacent to each bar.

(C) Inset: waterfall plot of sgRNAs where multiple sgRNAs targeting the same gene are significantly enriched in primary tumors. Each sgRNA is ranked by the percent of mice in which it is enriched. Only sgRNAs enriched in two or more mice are shown in the main panel. Main panel: enlargement and gene labels for sgRNAs at the top of the list from the inset (boxed region).

See also Figures S3, S4, and S5.

(Figures 3B and 3C). These genes were found to be mutated in patients in many previously reported cancer sequencing studies curated by cBioPortal (Cerami et al., 2012; Gao et al., 2013) (Figure S5A). For example, in somatic mutations identified by The Cancer Genome Atlas (TCGA) for NSCLC, including adenocarcinoma (LUAD) (Cancer Genome Atlas Research Network, 2014) and lung squamous cell carcinoma (LUSC) (Cancer Genome Atlas Research Network, 2012), 36% (107/407) of patients have one or more of these 24 genes mutated (Figures S5B and S5C). Several candidates were well-known tumor suppressors, such as *Pten*, cyclin-dependent kinase inhibitor 2b (*Cdkn2b*), neurofibromin 2 (*Nf2*/*Merlin*), alpha-type platelet-derived growth factor receptor (*Pdgfra*), and integrin alpha X (*Itgax*).

Enriched sgRNAs in Metastases

We also sequenced the sgRNA distributions from three lung lobes for each mouse transplanted with mGeCKOa-transduced Cas9-GFP KPD cells. In each lobe, the sgRNA representation is dominated by one or a few sgRNAs (Figures 4A, S3D, and S4I). In each mouse, the lung sgRNA representation (average of

normalized sgRNA representations from three lobes) is also dominated by a small number of sgRNAs (on average, 3.4 ± 0.4 sgRNAs with >5% of total reads) (Figure 4B), suggesting that metastases were seeded by a small set of cells, which grew to dominance over this timescale. Non-targeting sgRNAs were occasionally detected in the metastases but were never observed at high frequency (<0.1% of total reads in any lobe; Figures 2C, 4A and 4B, and S4I). These observations are consistent with our finding that untransduced tumors are not metastatic (Figure 1E), suggesting that specific sgRNA-mediated mutations led to metastasis.

The sgRNA representations in the lung metastases are similar to those in the late-stage primary tumors in several ways. First, the detected sgRNAs in lung samples overlap significantly with those in late tumor samples (chi-square test, $p < 10^{-15}$) (Figure S3E). Second, the number of sgRNAs detected in lung samples correlates, albeit weakly, with the number of sgRNAs detected in late primary tumor samples ($\rho = 0.42$, F test, $p = 0.097$) (Figure S3F). Third, the abundance (number of reads) of sgRNAs in the lung correlates positively with that in the late primary tumors of the same mouse (correlation, $\rho = 0.18$ on average, F test, $p < 0.01$, $n = 9$) (Figure S3G). Fourth, in most mice (8/9), the lung metastasis enriched sgRNAs also occupy a large fraction of reads in the late primary tumor of the same mouse (Figure 4C, left panel), significantly larger than a random sampling of the same number of sgRNAs from the mGeCKOa library (Figure 4C, right panel). These data indicate that mutants with preferential ability to proliferate in late primary tumors are more likely to dominate the metastases.

The three methods (threshold, rank, or FDR) of finding enriched sgRNAs in the lung metastases yield similar results (Figure S4B). Using the non-targeting sgRNA distribution to set a FDR-based cutoff for enrichment, the enriched sgRNAs in different lobes of the same mouse overlap with each other by 62% ± 5% (chi-square test, $p < 10^{-15}$) (Figure S4C), while different mice show greater variability while still overlapping significantly (29% ± 3%, chi-square test, $p < 10^{-15}$) (Figure S4D). The overlap between sgRNAs in different biological/infection replicate experiments when pooling enriched sgRNAs from all mice in the same replicate is 54% (chi-square test, $p < 10^{-15}$) (Figure S4E), suggesting that pooling sgRNAs from mice in the same experiment facilitates the identification of shared hits. These data suggest that the three independent experiments reproducibly captured a common set of hits and provide a picture for in vivo experimental variation between different lobes, different animals, and different infection replicates.

We found 147 sgRNAs enriched in more than one lobe, and 105 sgRNAs enriched in the lung of more than one mouse (Figures 4D and 4E). These include sgRNAs targeting *Nf2*, *Pten*, tripartite motif-containing protein 72 (*Trim72*), fibrinogen alpha chain (*Fga*), *Bid*, cyclin-dependent kinase inhibitor 2a (*Cdkn2a*), zinc finger FYVE domain-containing 28 (*Zfyve28*), reproductive homeobox 13 (*Rhox13*), and BRISC and BRCA1 A complex member 1 (*Babam1*), as well as microRNA genes *miR-152* and *miR-345*. Intriguingly, a few sgRNAs targeting the Pol II subunits and olfactory receptor are also enriched in the lung, possibly due to off-target effects or unknown roles of these genes. For most sgRNAs detected in lung metastases, the relative abundance in metastases is lower than that in the late primary tumor of the same mouse, with a metastasis-primary ratio (MPR) less than 1 (Figure S4F), likely due to more skewed distributions of sgRNAs in the metastases compared to those in the late primary tumors. A small subset of sgRNAs, however, are more abundant in metastases than in primary tumors (MPR > 1) in multiple mice, e.g., sgRNAs targeting *Nf2*, *Trim72*, prostaglandin E synthase 2 (*Ptges2*), or ubiquitin-conjugating enzyme E2G 2 (*Ube2g2*) (Figure 4F).

For four genes, *Nf2*, *Pten*, *Trim72*, and *Zfyve28*, two independent sgRNAs targeting different regions of the same gene were enriched in lung metastases (Figure 4G). One of the *Zfyve28*-targeting sgRNAs, however, is enriched in only one mouse, whereas *Nf2*, *Pten*, and *Trim72* all have two sgRNAs enriched in multiple mice (Figure 4H). These three genes, several representative genes with one frequently enriched sgRNA (*Cdkn2a*, *Fga*, and *Cryba4*), and the two top-scoring microRNAs (*miR-152* and *miR-345*) were chosen to assay individually for primary tumor growth and metastases formation.

Validation In Vivo Using Individual sgRNAs

For these eight genes (*Nf2*, *Pten*, *Trim72*, *Cdkn2a*, *Fga*, *Cryba4*, *miR-152*, and *miR-345*), we cloned multiple sgRNAs targeting each of them into the lentiGuide-Puro vector and transduced them into the Cas9-GFP KPD cell line (Figure 5A) (Experimental Procedures). As expected, these sgRNAs generated a broad distribution of NHEJ-mediated indels at the target site when examined 3 days post-transduction, with a bias toward deletions (Figure 5B). For protein-coding genes, the majority (>80%) of indels are out of frame, which potentially disrupts the protein functions. For *miR-152* and *miR-345*, the sgRNAs generated mostly deletions (>90% of indels are deletions, average indel size –7 bp) (Figure 5B), overlapping with the loop or mature microRNA sequences in the hairpins, which are structures required for maturation of microRNAs. For proteins where specific antibodies are available (Nf2 and Pten), we found that the majority of the protein products were significantly reduced 1 week after lentiviral sgRNA infection (Figure S6A).

When these single-sgRNA-transduced cells were transplanted into the flanks of immunocompromised mice, they all formed tumors in situ. With two mice injected per sgRNA and three sgRNAs per gene, all genes tested showed increased lung metastasis formation compared to controls (untransduced and non-targeting sgRNAs), with the most significant ones being *Nf2*, *Pten*, and *Cdkn2a* (Fisher's exact test, one-tailed, $p < 10^{-3}$) (Figures 5C and 5D). *Fga* and *Trim72* also have effects on metastasis acceleration (*Fga* $p = 0.001$, *Trim72* $p = 0.046$). *Cryba4* is not statistically different from controls ($p = 0.1$). sgRNAs targeting *miR-345* or *miR-152* significantly increased the rate of metastasis (*miR-345* $p = 0.01$, *miR-152* $p = 0.046$). These data suggest that loss-of-function mutations in any of *Nf2*, *Pten*, *Cdkn2a*, *Trim72*, *Fga*, *miR345*, or *miR-152* are sufficient to accelerate the rate of metastasis formation in this genetic background.

Most genes targeted by single sgRNAs also contributed to accelerated primary tumor growth compared to controls (Figure 5E). *Nf2* and *Pten* loss of function dramatically speed up tumor growth (KS test, $p < 0.001$) (Figure 5E); *Cdkn2a*-, *Trim72*-, and *Fga*-targeting sgRNAs slightly accelerate primary tumor growth (KS test, $p = 0.003$–0.01); *Cryba4* has a marginal effect

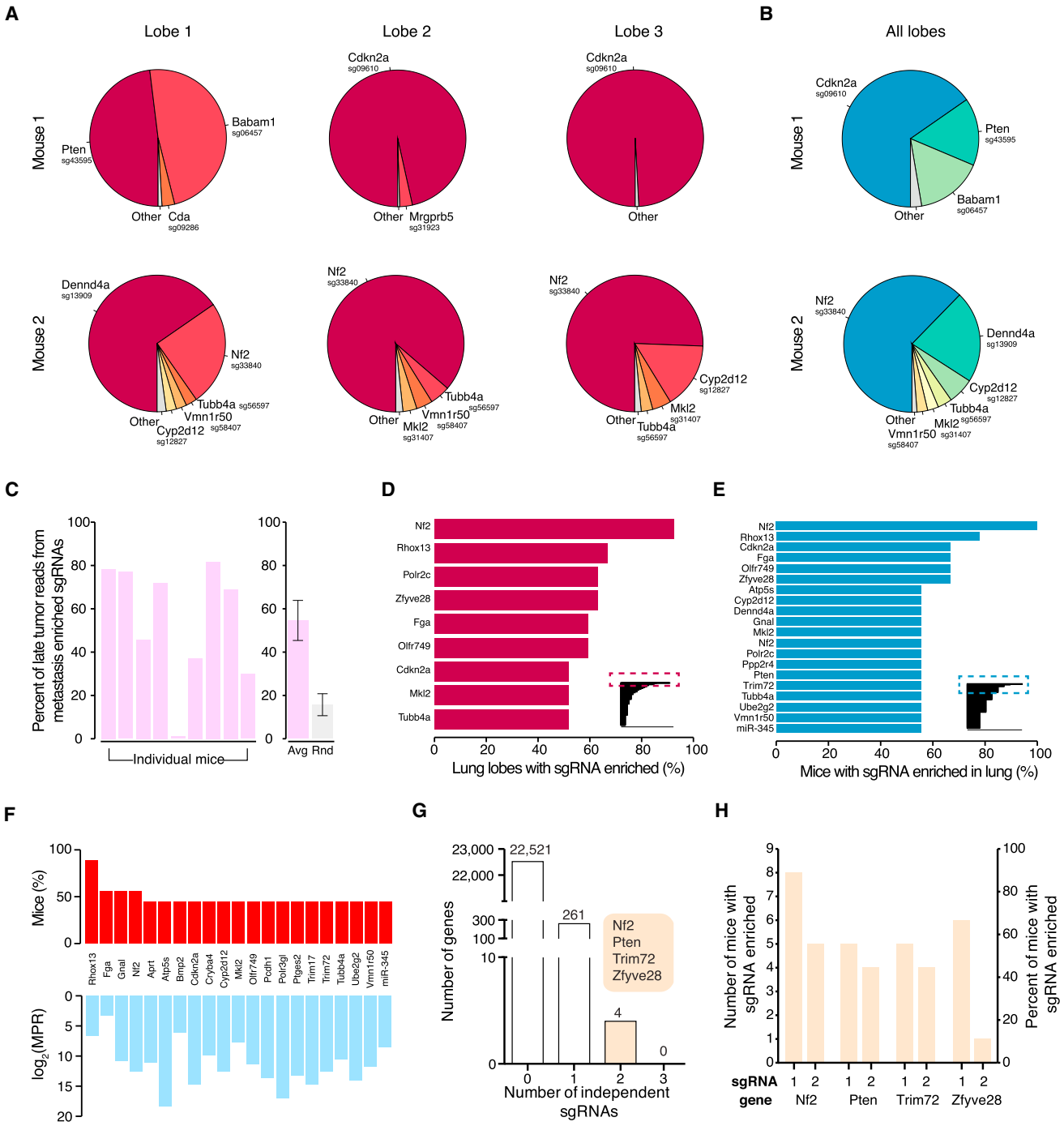


Figure 4. Enriched sgRNAs from the mGeCKOa Screen in Lung Metastases

(A) Pie charts of the most abundant sgRNAs in three individual lobes of the lungs of two representative mice transplanted with mGeCKOa-transduced cells. The area for each sgRNA corresponds to the fraction of total reads from the lobe for the sgRNA. All sgRNAs with ≥2% of total reads are plotted individually.

(B) Pie charts of the most abundant sgRNAs in the lung (averaged across three individual lobes) for the two mice shown in (A). All sgRNAs with ≥2% of average reads are plotted individually.

(C) Left: percentage of late tumor reads for the significantly enriched (FDR < 0.2%) mGeCKOa sgRNAs found in the lung metastases (averaged across three dissected lobes). Right: in purple, the percentage of late tumor reads for the significantly enriched (FDR < 0.2%) mGeCKOa sgRNAs found in the lung metastases (average across all mice, n = 9 mice). In gray, the percentage of late tumor reads for random, size-matched samplings of sgRNAs present in the late tumor (n = 100 samplings). Error bars indicate SD.

(legend continued on next page)

(KS test, $p = 0.08$); and neither *miR-152*- nor *miR-345*-targeting sgRNAs promote primary tumor growth (KS test, $p > 0.1$). Overall, for the targets we examined using individual sgRNAs, the number of lobes with lung metastases strongly correlates with the terminal volume of the late primary tumor (or average primary tumor growth rate) (correlation, $\rho = 0.83$, F test, $p < 0.01$) (Figure 5F), indicating at a single-gene level that mutant cells with a stronger ability to promote primary tumor growth generate metastases faster.

To analyze blood samples for the presence of circulating tumor cells (CTCs), we designed a microfluidic device based on the physical size of the Cas9-GFP KPD cells (Figures S6B and S6C). We performed CTC capture with terminal blood samples from mice injected with Cas9-GFP KPD cells transduced with sgRNAs targeting *Nf2*, *Pten*, *Trim72*, *Cdkn2a*, and *miR-152* and from mice injected with Cas9-GFP KPD control cells (untransduced or non-targeting sgRNA) (Figures S6C and S6D). Mice transplanted with cells transduced with sgRNAs targeting *Nf2*, *Pten*, *Trim72*, or *Cdkn2a* had a higher concentration of CTCs as compared to controls (Figures S6D–S6G), consistent with the higher rate of lung metastasis formation.

Competitive Dynamics of Top Hits Assessed Using an sgRNA Minipool

To better understand the relative metastatic potential of multiple genes from our genome-wide screen, we designed a targeted pooled screen with a smaller library. This small library (termed validation minipool) contains 524 sgRNAs targeting 53 genes that had highly enriched sgRNAs in lung metastases in the genome-wide screen (ten sgRNAs per gene for most genes) plus 100 non-targeting sgRNAs. We also created a size-matched library containing 624 non-targeting sgRNAs (termed control minipool) (Figure 6A). Lentiviruses from these two pools were used to transduce the Cas9-GFP KPD cells, which were cultured in vitro for 1 week and then transplanted into *Nu/Nu* mice (Figure 6A). Both validation minipool- and control minipool-transduced cells induced primary tumor growth at a similar rate (Figure 6B). However, mice transplanted with validation minipool cells had a dramatically elevated rate of lung metastasis formation (Figure 6C).

We sequenced the validation minipool plasmid library and the transduced cells pre-transplantation, as well as the late-stage primary tumors and whole lungs of the mice at 5 weeks post-transplantation (see Data S2 in Dataset S1). The sgRNA representations correlate strongly between technical replicates of the transduced cell pool, late primary tumors, and lung metastases (Figures S7A and S7D). The sgRNA representation in the cell sample strongly correlated with the plasmid (correlation, $\rho = 0.91$) (Figures S7B and S7D). Almost all (99.4%) sgRNAs were recovered in the plasmid and the cell population (Figure S7C). The late primary tumors retained less than half of the sgRNAs, and the metastases in the whole lung retained only a small fraction (2%–7%) of all sgRNAs (Figure S7C). Enriched sgRNAs from lung metastases clustered with each other and with late primary tumors (Figure S7D). Similar to the genome-wide library, in this validation minipool, the plasmid and cell samples had a unimodal distribution of sgRNAs, whereas the late primary tumors and lung metastases contained a bimodal distribution, with the majority of sgRNAs being absent and a small fraction spanning a large range of non-zero read counts (Figure 6D). Intriguingly, two mice retained relatively high sgRNA diversity in late primary tumors (Figure 6D), likely due to dormant or slowly proliferating cells that remained in low numbers during tumor growth. Similar to the genome-wide library, large shifts in the sgRNA distribution exist between different sample types (KS test, $p < 10^{-15}$ for pairwise comparisons between the cell, primary tumor, and lung metastases, $p = 0.02$ between plasmid and cell) (Figure 6E).

In the validation minipool, the sgRNAs detected in the late primary tumors or the lungs of five different mice significantly overlap with each other (Figures S7E and S7F). The late primary tumors and lung metastases are dominated by a few sgRNAs (Figures 7A and S7G–S7I), suggesting that these sgRNAs outcompete others during tumor growth and metastasis. With the validation library, the sgRNA representations are highly correlated between late primary tumors and lung metastases (correlation, $\rho = 0.55$ on average, F test, $p < 0.01$, $n = 5$) (Figure 7B). The late primary tumors and lung metastases have dozens of sgRNAs at moderate to high frequencies (Figures 7B and 7C). Several genes have multiple independent sgRNAs that are enriched in the lung over the primary tumor (MPR > 1), such as *Nf2* (eight sgRNAs), *Pten* (four sgRNAs), *Trim72* (three sgRNAs), *Ube2g2* (three sgRNAs), *Ptges2* (two sgRNAs), and ATP-dependent DNA ligase IV (*Lig4*) (two sgRNAs) (Figures 7C and 7D). Two *Cdkn2a* sgRNAs were present in both late primary tumors and lung metastases in two mice, but with MPR < 1. *Fga*-, *Cryba4*-, *miR-152*-, and *miR-345*-targeting sgRNAs were not found at high frequency in either late primary tumors or lung metastases, suggesting that they are outcompeted by other loss-of-function mutations (such as *Nf2*), which agrees with the relatively reduced metastasis formation of these genes in the

(D) Inset: all sgRNAs found in individual lung lobes, ordered by the percent of lobes in which a particular sgRNA was among the significantly enriched (FDR < 0.2%) sgRNAs for that lobe. Only sgRNAs enriched in two or more lobes are shown. Main panel: enlargement and gene labels for sgRNAs at the top of the list from the inset (boxed region).

(E) Inset: all sgRNAs found in individual mice (averaged across three dissected lobes), ordered by the percent of mice in which a particular sgRNA was among the significantly enriched (FDR < 0.2%) sgRNAs for that mouse. Only sgRNAs enriched in two or more mice are shown. Main panel: enlargement and gene labels for sgRNAs at the top of the list from the inset (boxed region).

(F) Bottom: metastasis primary ratio (MPR) for the sgRNAs in mGeCKOa with enrichment in metastases over late tumors (MPR > 1) observed in at least three mice. The sgRNAs are sorted by the number of mice in which the MPR for the sgRNA is greater than 1. Top: number of mice in which the MPR for this sgRNA is greater than 1. In both panels, individual sgRNAs are labeled by gene target.

(G) Number of genes with 0, 1, 2, or 3 significantly enriched (FDR < 0.2% for at least one mouse) mGeCKOa sgRNAs in the lung metastases. For genes with 2 enriched sgRNAs, gene names are indicated in the colored bubble adjacent to the bar.

(H) Number of mice and percentage of mice in which each sgRNA was enriched in the lung metastases for all genes with multiple enriched sgRNAs.

See also Figures S4 and S5.

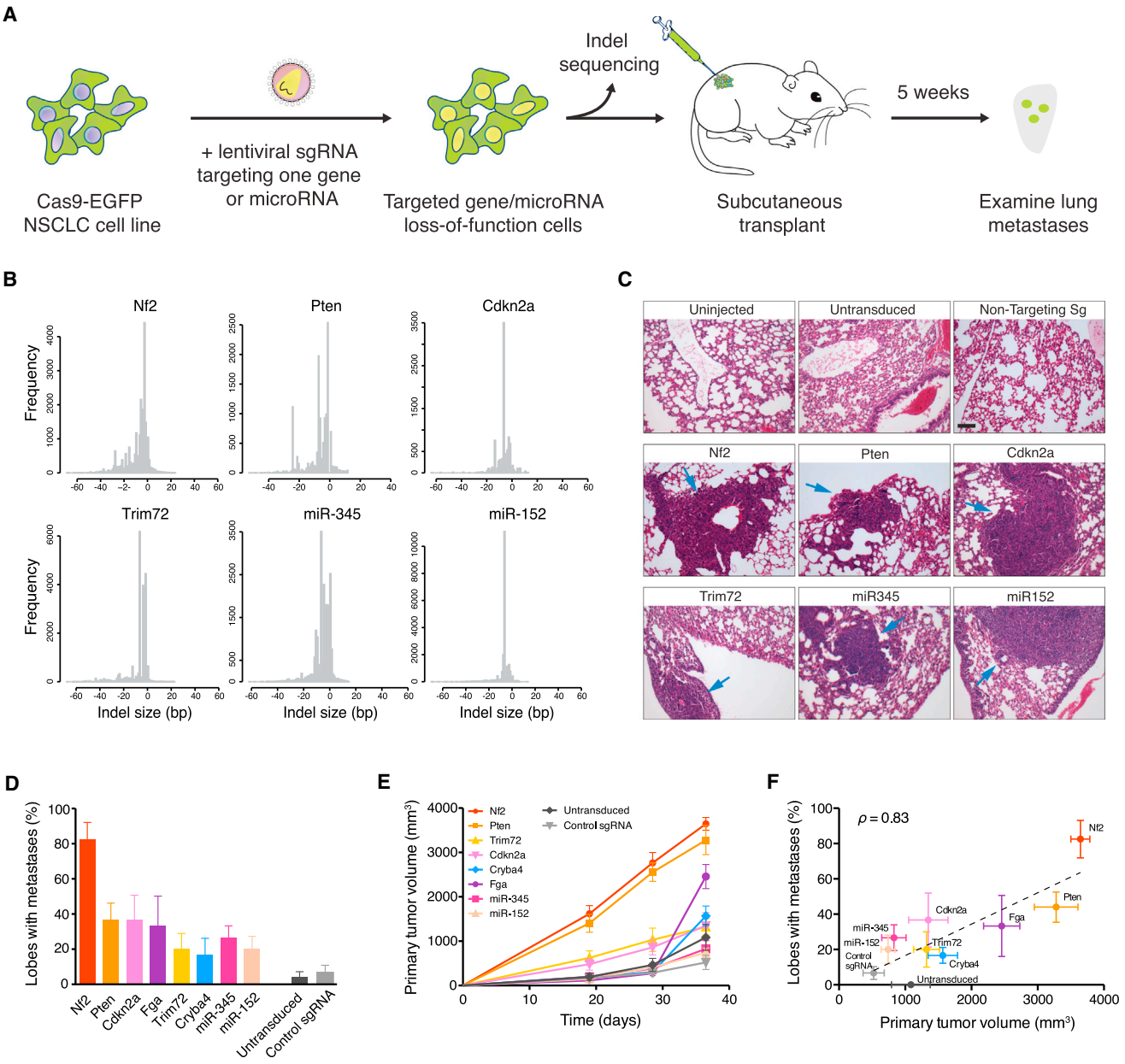


Figure 5. Validation of Target Genes and MicroRNAs from mGeCKOa Screen Using Individual sgRNAs

(A) Schematic representation of lentiviral transduction of Cas9-GFP KPD cells with single sgRNAs designed to target one gene or miR. After puromycin selection, the cell population was transplanted into *Nu*/*Nu* mice and also deep sequenced to examine the distribution of indels at the target site. After 5 weeks, the primary tumor and lungs were examined.

(B) Histograms of indel sizes at the genomic locus targeted by a representative sgRNA for each gene/miR after 3 days of puromycin selection. Indels from sgRNAs targeting the same gene were pooled (6 sgRNAs for each protein-coding gene; 4 sgRNAs for each miR).

(C) Representative H&E staining of lung lobes from uninjected mice (n = 3 mice), mice transplanted with cells transduced with Cas9 only (n = 5), and mice transplanted with cells containing Cas9 and a single sgRNA (n = 6). Single sgRNAs are either control/non-targeting sgRNAs (n = 6 mice for control sgRNAs, 3 distinct control sgRNAs with 2 mice each) or targeting sgRNAs (n = 6 mice for each gene/miR target, 3 sgRNAs per target with 2 mice each). Blue arrows indicate lung metastases. Scale bar, 10 μm.

(D) Percent of lung lobes with metastases after 6 weeks for the mice in (C). Error bars indicate SEM.

(E) Primary tumor growth curve of *Nu*/*Nu* mice transplanted with NSCLC cells transduced with Cas9 only (n = 5) or single sgRNAs (n = 6 mice per gene/miR target, 3 sgRNAs per target with 2 mice each; n = 6 mice for control sgRNAs, 3 control sgRNAs with 2 mice each). Error bars indicate SEM.

(F) Correlation between primary tumor volume and percent of lobes with metastases for each gene in (D) and (E). Error bars indicate SEM.

See also Figure S6.

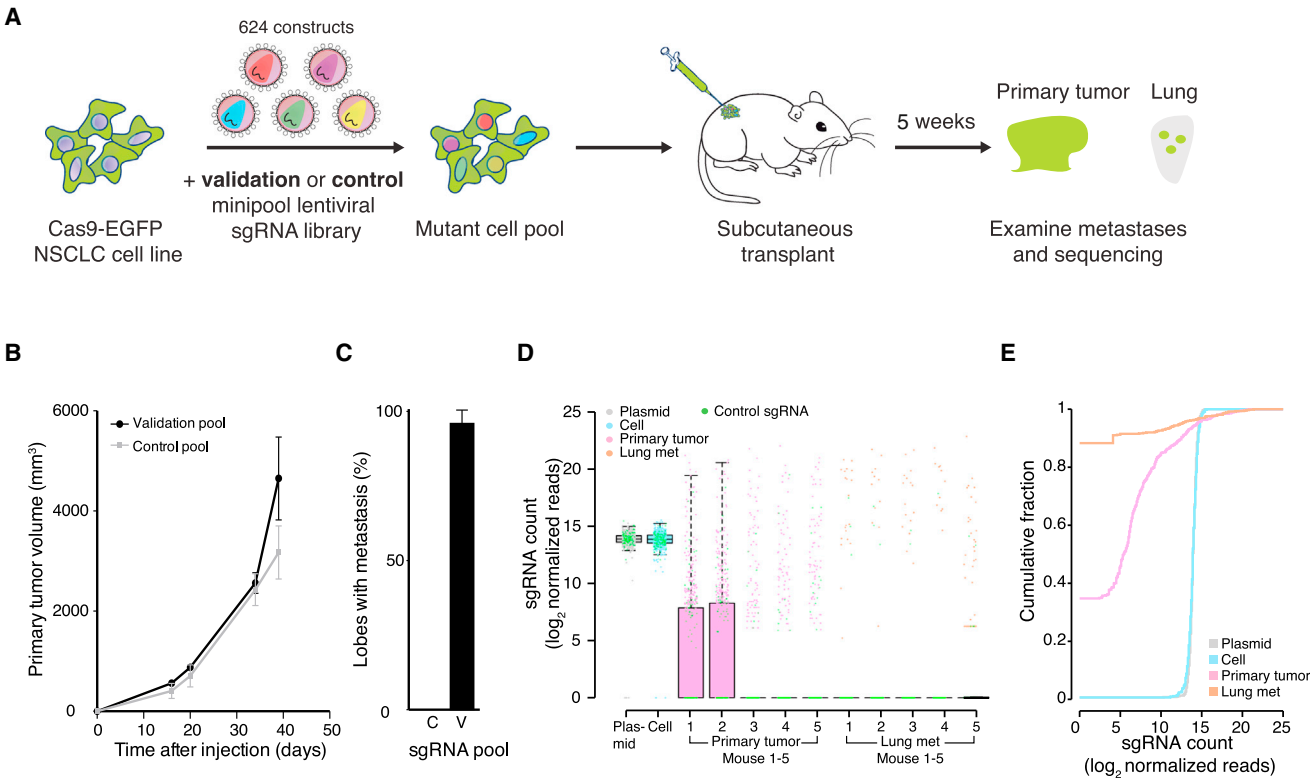


Figure 6. Tumor Evolution and Library Representation in Transplanted Cas9-GFP KPD Cells with Minipool Libraries

(A) Schematic representation of the loss-of-function metastasis minipool screen. Briefly, Cas9-GFP KPD cells were transduced with either validation minipool (524 gene-targeting + 100 non-targeting sgRNAs) or control minipool (624 non-targeting sgRNAs). After puromycin selection, the cell pools were transplanted into *Nu*/*Nu* mice. After 5 weeks, validation minipool sgRNAs were sequenced from primary tumor and lung samples.

(B) Primary tumor growth curve of *Nu*/*Nu* mice transplanted with Cas9 vector + validation minipool cells (n = 5 mice) or Cas9 + control minipool cells (n = 5 mice). Error bars indicate SEM.

(C) Percent of lung lobes with metastases after 6 weeks for the mice in (B). C, control minipool; V, validation minipool. Error bars indicate SEM.

(D) Boxplot of the sgRNA normalized read counts for the plasmid library, cells before transplantation, primary tumors, and lung metastases using the validation minipool.

(E) Cumulative probability distribution of library sgRNAs in the validation plasmid pool, cells before transplantation, primary tumors, and lung metastases. Distributions of primary tumor and lung metastases are averaged across five mice.

See also Figure S7.

individual sgRNA validation. These results further validate several of the top hits from the primary screen, using either sgRNA dominance (e.g., *Nf2*, *Pten*, *Trim72*) or MPR (e.g., *Nf2*, *Trim72*, *Ube2g2*, *Ptges2*). This validation minipool reveals the dynamics of multiple competing mutants chosen from the primary screen hits and indicates that mutants with strong pro-growth effects tend to enhance metastasis (Figure 7E).

TCGA Gene Expression of Screen Hits in Human Lung Cancer

To assess the relevance of our mGeCKOa and validation minipool screen hits (genes targeted by sgRNAs enriched in lung metastases) to pathological metastasis in human cancer, we performed gene expression analysis of the human orthologs of these genes. We compared mRNA levels in metastatic compared to non-metastatic primary tumors in patient samples using TCGA mRNA sequencing data. We found that most (61%–75%) of these genes are downregulated in metastatic tumors in NSCLC patients (Figures S5D and S5E; Table S6). These data suggest that downregulation of these genes is selected for in metastatic tumors from patients.

DISCUSSION

Pooled Mutagenesis in a Metastasis Model

Distal metastases develop as primary tumors shed CTCs into the circulation, from which CTCs travel to the destination site, move out of the blood or lymphatic vessels, and initiate clonal growth (Valastyan and Weinberg, 2011; Vanharanta and Massagué, 2013; Weinberg, 2007). In this study, cancer cells transplanted into the flanks of mice form primary tumors in situ, and cells from this mass undergo the intravasation-circulation-extravasation-clonal growth cascade to form distal metastases (Francia et al., 2011). The initial lung cancer cell line has little capacity to form metastases; in contrast, after being mutagenized with the mGeCKOa genome-scale Cas9 knockout library, the cell

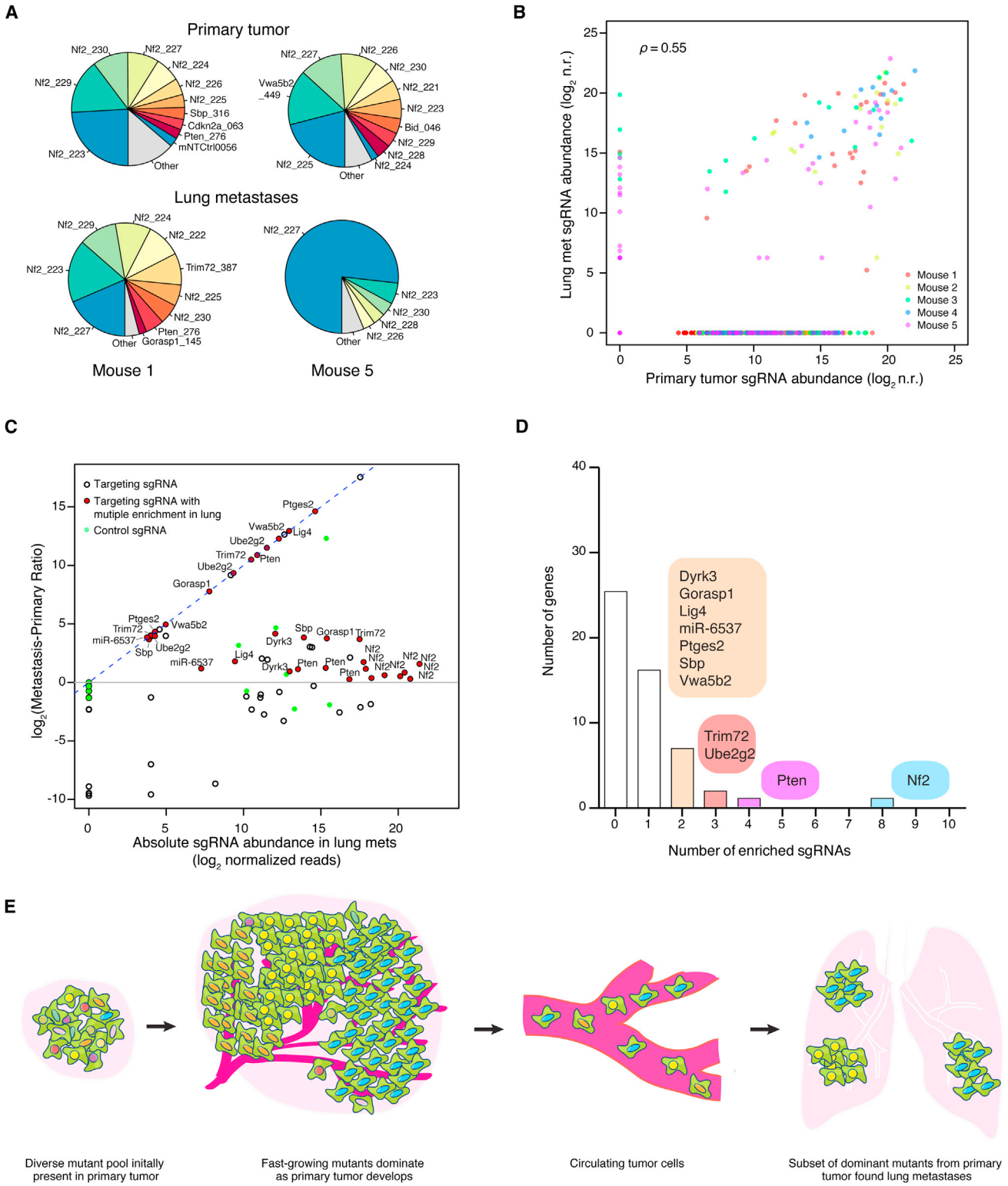


Figure 7. Enriched sgRNAs from the Validation Minipool Screen in Primary Tumors and Lung Metastases

(A) Pie charts of the most abundant sgRNAs in the primary tumor and the whole lung of two representative mice transplanted with validation minipool-transduced Cas9-GFP KPD cells. The area for each sgRNA corresponds to the fraction of total reads from the tissue (primary tumor or lung metastases) for the sgRNA. All sgRNAs with ≥2% of total reads are plotted individually.

(legend continued on next page)

population forms highly metastatic tumors. Thus, these mutations, acting in simple or complex pleiotropic ways, accelerate metastasis. In this model, the effect of mutations on metastasis strongly correlates with their abundance in late-stage primary tumors.

sgRNA Dynamics during Tumor Evolution

The dynamics of the sgRNA population changed dramatically over the course of tumor development and metastasis, reflecting the selection and bottlenecks of cellular evolution in vitro and in vivo. After a week in culture, cells retained most of the sgRNAs present in the plasmid library, with decreases in sgRNAs targeting genes involved in fundamental cellular processes. The distribution of non-targeting control sgRNAs is almost identical to those targeting genes, suggesting that the selective pressure of in vitro culture alone does not radically alter sgRNA representation, similar to previous observations in human melanoma cells (Shalem et al., 2014).

In contrast, less than half of the sgRNAs survive in an early-stage primary tumor. This loss of representation occurs with both gene-targeting sgRNAs and non-targeting control sgRNAs, suggesting that random sampling influences sgRNA dynamics during the transplantation and tumor initiation processes, although we cannot exclude that some of the non-targeting sgRNAs might have detrimental or pro-growth effects. We also detected further dropout of genes involved in fundamental cellular processes in early tumor samples compared to cell samples. Thus, it is likely that the sgRNA dynamics are influenced by a combination of selection and random sampling during transplantation and tumor initiation.

As primary tumors grow, the mutant cells proliferate and compete as a pool. This creates strong selection for sgRNAs targeting anti-apoptotic genes and other tumor suppressors. The majority of the genetic diversity in early tumors is lost during the subsequent 4 weeks of primary tumor growth in mice. Accordingly, sequencing revealed a smaller set of dominant sgRNAs, usually on the order of hundreds to a few thousand per mouse. In addition, almost all of non-targeting sgRNAs are lost during primary tumor growth, which is consistent with selection for cells with special growth and survival properties. This observation is also consistent with earlier transplantation studies by Kerbel and colleagues using small pools of randomly mutagenized cells, which found that the majority of clonal variants detectable by Southern blot disappeared within 6 weeks of primary tumor growth, leaving one dominant clone (Korczak et al., 1988; Waghorne et al., 1988).

Each step toward metastasis has a bottleneck effect. In the lung metastases, we detected very few sgRNAs at high abundance. As with the primary tumor, we found only a few non-targeting sgRNAs at low frequencies in metastases. Their presence could be due to unknown off-target effects of these sgRNAs, random shedding of CTCs in the primary tumor, or clustering together with other strongly selected CTCs during metastasis (Aceto et al., 2014).

Relevance of Screen Hits to Human Cancer

Several of the genes enriched in late-stage primary tumors are associated with cancer, but their functions in tumor growth are poorly understood. For example, *Mgmt*, a gene with two enriched sgRNAs, is required for DNA repair and is thus crucial for genome stability (Tano et al., 1990). Mutation, silencing, or promoter methylation of *MGMT* is associated with primary glioblastomas (Jesien-Lewandowicz et al., 2009). *Med16*, another gene with two enriched sgRNAs, encodes a subunit of the mediator complex of transcription regulation, which has been recently implicated in cancer (Huang et al., 2012; Schiano et al., 2014).

We found that the genes that are significantly enriched in lung metastases largely overlap with those found in abundance in the late primary tumor. Several of these hits were validated in vivo using multiple individual sgRNAs, including *Nf2*, *Pten*, *Cdkn2a*, *Trim72*, *Fga*, *miR-152*, and *miR-345*. *Nf2*, *Pten*, and *Cdkn2a* are well-known tumor suppressor genes. Intriguingly, the *NF2* locus is mutated at only 1% frequency in primary tumors of human NSCLC patients (LUAD and/or LUSC) (Cancer Genome Atlas Research Network, 2012, 2014). *Nf2* mutant mice develop a range of highly metastatic tumors (McClatchey et al., 1998). It is possible that *NF2* mutations influence metastases to a greater degree than primary tumor growth, but this awaits metastasis genomics from patient samples. *Pten* mutations are also associated with advanced stages of tumor progression in a mouse model of lung cancer (McFadden et al., 2014), and *PTEN* was found to be mutated at 8% in adenocarcinoma patients (LUAD). *CDKN2A* has been shown to be often inactivated in lung cancer (Kaczmarczyk et al., 2012; Yokota et al., 2003). *Fga* encodes fibrinogen, an extracellular matrix protein involved in blood clot formation. *Fga* mutations have been found in various cancer types in TCGA (Lawrence et al. 2013), as well as circulating tumor cells (Lohr et al., 2014). *Trim72* is an E3 ubiquitin ligase, and its role in cancer metastasis is largely unknown. Studies have shown that *miR-152* and *miR-345* are associated with cancer and metastasis (Cheng et al., 2014; Tang et al., 2011). *FGF2* and *BAG3*, which promote metastasis, were predicted targets of *miR-152* and *miR-345*; thus, loss of these

(B) Scatterplot of normalized sgRNA read counts in primary tumor and lung metastases for all sgRNAs in the validation minipool for each mouse (different color dots indicate sgRNAs from different mice). $\log_2$ n.r., $\log_2$ normalized reads.
(C) $\log_2$ ratio of sgRNA abundance in the lung metastases over the primary tumor (MPR) plotted against the abundance in the lung metastases (n = 5 mice per sgRNA). Green dots are the 100 control sgRNAs. Dots with black outlines are non-control sgRNAs that target genes or miRs. Red dots indicate non-control sgRNAs for which more than one sgRNA targeting the same gene/miR is enriched in the lung metastases over the primary tumor (i.e., $\log_2$(MPR) > 0) and are labeled with the gene/miR targeted. The lung-primary ratio is calculated for individual mice, and these quantities are averaged across mice.
(D) Number of genes with 0 to 10 significantly enriched validation minipool sgRNAs in lung metastases. For genes/miRs with 2 or more enriched sgRNAs, genes/miRs are categorized by how many sgRNAs targeting that gene/miRs are enriched, as indicated in the colored bubbles adjacent to each bar.
(E) Schematic illustration of tumor growth and metastasis in the library-transduced NSCLC transplant model. The initially diverse set of loss-of-function mutations in the subcutaneously transplanted pool is selected over time for mutations that promote growth of the primary tumor. A subset of these mutants also dominate lung metastases.
See also Figure S7.

microRNAs may lead to acceleration of metastases, likely due to de-repression of these genes (Cheng et al., 2014; Tang et al., 2011).

In our own analysis of TCGA samples from lung cancer patients, we observed downregulation of the human orthologs of the genes identified in the genome-wide and validation minipool screens at the mRNA level in metastatic tumors compared to non-metastatic tumors, suggesting that these genes may also be inactivated during pathological metastasis. Human orthologs of these genes are often found to be mutated in cancers. Moreover, these genes have been implicated in various pathways and biological processes in tumorigenesis and/or metastasis in human cancer (Tables S7A–S7C). However, most cancer sequencing studies involve samples from primary tumors of patients. In the clinic, metastases are rarely sampled. Future patient sequencing directly from metastases may further connect genes identified in the mouse model to those mutated or silenced in clinical metastases.

Future In Vivo Functional Genomic Screens

Our study provides a roadmap for in vivo Cas9 screens, and future studies can take advantage of this model to explore other oncogenotypes, delivery methods, or metastasis target organs. Genome-scale CRISPR screening is feasible using a transplant model with virtually any cell line or genetic background (e.g., mutations in *EGFR*, *KRAS*, *ALK*, etc.), including a large repertoire of human cell lines from diverse cancer types (Barretina et al., 2012). Other cell delivery methods, such as intravenous injection or orthotopic transplantation, may help identify genes regulating extravasation and clonalization. Examining samples from other stages or sites, such as CTCs or metastases to other organs, can provide a more refined picture of tumor evolution.

In addition to these parameters, several aspects of the screen perturbations themselves can also be modified. Targeted drug therapies or immunotherapies can be applied in conjunction with the in vivo screening strategy to identify genes involved in acquired resistance. Other screening technologies, such as Cas9-mediated activation (Gilbert et al., 2014; Konermann et al., 2015), can identify metastasis-regulating factors that act in a gain-of-function manner. Activation screens that identify oncogenes, as well as dropout screens that identify genetic dependencies, may facilitate identification of novel therapeutic targets. Targeted subpool strategies can be used to reduce the library size and facilitate further confirmation of primary screens. In a customized library, genes can be chosen based on genomic analysis, pathways, or clinical relevance for focused screening libraries. Additionally, application of pooled sgRNA libraries using individually barcoded cells will allow quantitative assessment of the robustness and significance of each candidate hit and will enable analysis of the competitive dynamics among different perturbations. With these promising future directions and the results of our study, Cas9-based in vivo screening establishes a new platform for functional genomics discovery.

EXPERIMENTAL PROCEDURES

Generation of Cas9-GFP Expression Vector

A lentiviral vector, lenti-Cas9-NLS-FLAG-2A-EGFP (lentiCas9-EGFP), was generated by subcloning Cas9 into a lentiviral vector.

Pooled Guide-Only Library Cloning and Viral Production

The Cas9-GFP KPD cell line was transduced at a MOI of ~0.4 with lentivirus produced from a genome-wide lentiviral mouse CRISPR knockout guide-only library (Sanjana et al., 2014) containing 67,405 sgRNAs (mGeCKOa, Addgene 1000000053) with at least 400-fold representation (cells per construct) in each infection replicate. A detailed viral production and infection protocol can be found in Extended Experimental Procedures.

Animal Work Statement

All animal work was performed under the guidelines of the MIT Division of Comparative Medicine, with protocols (0411-040-14, 0414-024-17, 0911-098-11, 0911-098-14, and 0914-091-17) approved by the MIT Committee for Animal Care, and were consistent with the Guide for the Care and Use of Laboratory Animals, National Research Council, 1996 (institutional animal welfare assurance no. A-3125-01).

Mice, Tumor Transplant, and Metastasis Analysis in the Primary Screen

Untransduced or mGeCKOa-transduced Cas9-GFP KPD cells were injected subcutaneously into the right side flank of *Nu/Nu* mice at 3×10^7 cells per mouse. Transplanted primary tumor sizes were measured by caliper. At 6 weeks post-transplantation, mice were sacrificed and several organs (liver, lung, kidney, and spleen) were dissected for examination of metastases under a fluorescence stereoscope.

Mouse Tissue Collection

Primary tumors and other organs were dissected manually. For molecular biology, tissues were flash frozen with liquid nitrogen and ground in 24-well polyethylene vials with metal beads in a GenoGrinder machine (OPS Diagnostics). Homogenized tissues were used for DNA/RNA/protein extractions using standard molecular biology protocols. Tissues for histology were then fixed in 4% formaldehyde or 10% formalin overnight, embedded in paraffin, and sectioned at 6 μm with a microtome as described previously (Chen et al., 2014). Slices were subjected to H&E staining as described previously (Chen et al., 2014).

Genomic DNA Extraction from Cells and Mouse Tissues

Genomic DNA from cells and tissues (primary tumors and lungs) was extracted using a homemade modified salt precipitation method similar to the Puregene (QIAGEN/Gentra) procedure. The sgRNA cassette was amplified and prepared for Illumina sequencing as described previously (Shalem et al., 2014). A detailed readout protocol can be found in Extended Experimental Procedures.

Individual Gene and MicroRNA Validation

Six sgRNAs per protein-coding gene and four sgRNAs per microRNA gene were chosen for validation using individual sgRNAs (Table S4). For protein-coding genes, we cloned both the three sgRNAs from the mGeCKOa library and three additional sgRNAs to target each gene. For microRNAs, we used all four sgRNAs from the mGeCKOa library.

Validation and Control Minipool Synthesis and In Vivo Transplantation

Validation and control minipools (Table S5) were synthesized using array oligonucleotide synthesis (CustomArray) and transduced at >1,000-fold representation in Cas9-GFP KPD cells. After 7 days in culture, Cas9-GFP KPD cells transduced with the validation minipool or control minipool were injected subcutaneously into the right side flank of *Nu/Nu* mice at 3×10^7 cells per mouse with five replicate mice. After 5 weeks, mice were sacrificed, and primary tumors and lungs were dissected.

ACCESSION NUMBERS

Genomic sequencing data have been deposited in the NCBI Sequence Read Archive under accession number PRJNA273894. Plasmids and pooled libraries have been deposited in Addgene (LentiCas9-EGFP: 63592, Metastasis Validation Minipool library: 63594, Mouse Non-targeting Control Minipool: 63595).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, seven tables, and a dataset and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.02.038.

AUTHOR CONTRIBUTIONS

S.C, N.E.S., O.S., F.Z., and P.A.S. conceived and designed the study. S.C., N.E.S., and K.Z. performed all screening and validation experiments. S.C., N.E.S., O.S., and D.A.S. analyzed the data. K.L., J.S., R.W., and H.L. designed the CTC chip and performed CTC analysis. X.S. and J.Q.P. performed western blots. S.C., N.E.S., F.Z., and P.A.S. wrote the manuscript with the input from all authors. P.A.S. and F.Z. supervised the work.

ACKNOWLEDGMENTS

We thank R. Weinberg and R. Kerbel for critically reading the manuscript; the entire P.A.S. lab and F.Z. lab, L. Cong, T. Kelly, X. Ni, M. Nobel, J. Boehm, A. Tsherniak, S. Levine, M. Cornwall-Brady, S. Malstrom, M. Jennings, E. Vasile, C. Whittaker, K. Cormier, R. Bronson, and colleagues at the Koch Institute, Broad Institute, McGovern Institute, and Department of Biology for technical assistance and/or discussion; the Swanson Biotechnology Center for technical support (Genomics, Animal Imaging and Preclinical Testing, Bioinformatics and Computing, Microscopy, Flow Cytometry, Microscopy, and Histology, in particular). This work is supported by grants to P.A.S., including a United States Public Health Service grant R01-CA133404 from the NIH, an MIT-Harvard Center for Cancer Nanotechnology Excellence Grant U54 CA151884 from the National Cancer Institute, a generous gift from the Marie D. and Pierre Casimir-Lambert Fund, an SkTech/MIT Initiative Grant from the Skolkovo Foundation, and the Koch Institute Support (core) grant P30-CA14051 from the National Cancer Institute. F.Z. is supported by the NIH through NIMH grant 5DP1-MH100706) and NIDDK grant 5R01-DK097768; a Waterman Award from the National Science Foundation; the Keck, New York Stem Cell, Damon Runyon, Searle Scholars, Merkin, and Vallee foundations; and Bob Metcalfe. F.Z. is a New York Stem Cell Foundation Robertson Investigator. S.C. is a Damon Runyon Cancer Research Fellow (DRG-2117-12) and also supported by the Dale Frey Award for Breakthrough Scientists. N.E.S. is supported by a Simons Center for the Social Brain Postdoctoral Fellowship and NIH NHGRI award K99-HG008171. O.S. is a fellow of the Klarman Cell Observatory. R.W. is supported by NIH grant U54 CA151884. D.A.S. is supported by a NSF Graduate Research Fellowship. H.L. is supported by Department of Defense grant OCRP W81XWH-14-1-0279. CRISPR reagents (plasmids and libraries) are available to the academic community through Addgene. F.Z. is a co-founder of Editas Medicine and a scientific advisor for Editas Medicine and Horizon Discovery.

Received: December 30, 2014
Revised: February 3, 2015
Accepted: February 18, 2015
Published: March 5, 2015

REFERENCES

Aceto, N., Bardia, A., Miyamoto, D.T., Donaldson, M.C., Wittner, B.S., Spencer, J.A., Yu, M., Pely, A., Engstrom, A., Zhu, H., et al. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell *158*, 1110–1122.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551–2561.

Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

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

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBiol. cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discovery *2*, 401–404.

Chen, S., Xue, Y., Wu, X., Le, C., Bhutkar, A., Bell, E.L., Zhang, F., Langer, R., and Sharp, P.A. (2014). Global microRNA depletion suppresses tumor angiogenesis. Genes Dev. *28*, 1054–1067.

Cheng, Z., Ma, R., Tan, W., and Zhang, L. (2014). MiR-152 suppresses the proliferation and invasion of NSCLC cells by inhibiting FGF2. Exp. Mol. Med. *46*, e112.

Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. *10*, 726–737.

Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res. *42*, 6091–6105.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Francia, G., Cruz-Munoz, W., Man, S., Xu, P., and Kerbel, R.S. (2011). Mouse models of advanced spontaneous metastasis for experimental therapeutics. Nat. Rev. Cancer *11*, 135–141.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, p11.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. Cell *153*, 17–37.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. USA *109*, E2579–E2586.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. Cell *159*, 647–661.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Heimann, R., and Hellman, S. (1998). Aging, progression, and phenotype in breast cancer. J. Clin. Oncol. *16*, 2686–2692.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Huang, S., Hölzel, M., Knijnenburg, T., Schlicker, A., Roepman, P., McDermott, U., Garnett, M., Grernrum, W., Sun, C., Prahallad, A., et al. (2012). MED12 controls the response to multiple cancer drugs through regulation of TGF-β receptor signaling. Cell *151*, 937–950.

Jesien-Lewandowicz, E., Jesionek-Kupnicka, D., Zawlik, I., Szybka, M., Kulczycka-Wojdala, D., Rieske, P., Sieruta, M., Jaskolski, D., Och, W., Skowronski,

W., et al. (2009). High incidence of MGMT promoter methylation in primary glioblastomas without correlation with TP53 gene mutations. Cancer Genet. Cytogenet. *188*, 77–82.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Kaczmarczyk, G., Lewandowski, R., Trautsolt, W., Ziółkowski, A., and Kozielski, J. (2012). Cytological examination of pleural cavity lavage accompanied by the study of gene promoter hypermethylation of p16 and O6-methylguanine-DNA-methyltransferase genes in diagnostics of non-small cell lung cancer metastatic changes into pleura. Contemp Oncol (Pozn) *16*, 322–327.

Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, Mdel.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. *32*, 267–273.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Korczak, B., Robson, I.B., Lamarche, C., Bernstein, A., and Kerbel, R.S. (1988). Genetic tagging of tumor cells with retrovirus vectors: clonal analysis of tumor growth and metastasis in vivo. Mol. Cell. Biol. *8*, 3143–3149.

Kumar, M.S., Pester, R.E., Chen, C.Y., Lane, K., Chin, C., Lu, J., Kirsch, D.G., Golub, T.R., and Jacks, T. (2009). Dicer1 functions as a haploinsufficient tumor suppressor. Genes Dev. *23*, 2700–2704.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Lohr, J.G., Adalsteinsson, V.A., Cibulskis, K., Choudhury, A.D., Rosenberg, M., Cruz-Gordillo, P., Francis, J.M., Zhang, C.Z., Shalek, A.K., Satija, R., et al. (2014). Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. Nat. Biotechnol. *32*, 479–484.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

McClatchey, A.I., Saotome, I., Mercer, K., Crowley, D., Gusella, J.F., Bronson, R.T., and Jacks, T. (1998). Mice heterozygous for a mutation at the Nf2 tumor suppressor locus develop a range of highly metastatic tumors. Genes Dev. *12*, 1121–1133.

McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014). Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell *156*, 1298–1311.

Nguyen, D.X., Bos, P.D., and Massagué, J. (2009). Metastasis: from dissemination to organ-specific colonization. Nat. Rev. Cancer *9*, 274–284.

Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. *14*, 8096–8106.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. *39*, 9275–9282.

Schiano, C., Casamassimi, A., Rienzo, M., de Nigris, F., Sommese, L., and Napoli, C. (2014). Involvement of Mediator complex in malignancy. Biochim. Biophys. Acta *1845*, 66–83.

Schramek, D., Sendoel, A., Segal, J.P., Beronja, S., Heller, E., Oristian, D., Reva, B., and Fuchs, E. (2014). Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. Science *343*, 309–313.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shao, D.D., Xue, W., Krall, E.B., Bhutkar, A., Piccioni, F., Wang, X., Schinzel, A.C., Sood, S., Rosenbluh, J., Kim, J.W., et al. (2014). KRAS and YAP1 converge to regulate EMT and tumor survival. Cell *158*, 171–184.

Tang, J.T., Wang, J.L., Du, W., Hong, J., Zhao, S.L., Wang, Y.C., Xiong, H., Chen, H.M., and Fang, J.Y. (2011). MicroRNA 345, a methylation-sensitive microRNA is involved in cell proliferation and invasion in human colorectal cancer. Carcinogenesis *32*, 1207–1215.

Tano, K., Shiota, S., Collier, J., Foote, R.S., and Mitra, S. (1990). Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine. Proc. Natl. Acad. Sci. USA *87*, 686–690.

Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: molecular insights and evolving paradigms. Cell *147*, 275–292.

Vanharanta, S., and Massagué, J. (2013). Origins of metastatic traits. Cancer Cell *24*, 410–421.

Waghorne, C., Thomas, M., Lagarde, A., Kerbel, R.S., and Breitman, M.L. (1988). Genetic evidence for progressive selection and overgrowth of primary tumors by metastatic cell subpopulations. Cancer Res. *48*, 6109–6114.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Weinberg, R.A. (2007). The Biology of Cancer (Garland Science).

Yokota, J., Nishioka, M., Tani, M., and Kohno, T. (2003). Genetic alterations responsible for metastatic phenotypes of lung cancer cells. Clin. Exp. Metastasis *20*, 189–193.

Zender, L., Xue, W., Zuber, J., Semighini, C.P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J.M., Schirmacher, P., et al. (2008). An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. Cell *135*, 852–864.

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature *509*, 487–491.