



Published in final edited form as:

Nat Methods. 2019 May ; 16(5): 405–408. doi:10.1038/s41592-019-0371-5.

***In vivo* profiling of metastatic double knockouts through CRISPR-Cpf1 screens**

Ryan D. Chow^{†,1,2,3,4}, **Guangchuan Wang**^{†,1,2,3}, **Lupeng Ye**^{†,1,2,3}, **Adan Codina**^{1,2,3,5}, **Hyunu Ray Kim**^{1,2,3}, **Li Shen**^{1,2,3}, **Matthew B. Dong**^{1,2,3,4,6}, **Youssef Errami**^{1,2,3}, and **Sidi Chen**^{1,2,3,4,5,6,7,8,*}

¹Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA

²System Biology Institute, Integrated Science & Technology Center, Yale University, West Haven, Connecticut, USA

³Center for Cancer Systems Biology, Integrated Science & Technology Center, West Haven, Connecticut, USA

⁴Yale M.D.-Ph.D. Program, New Haven, Connecticut, USA

⁵Yale Combined Program in the Biological and Biomedical Sciences, New Haven, Connecticut, USA

⁶Yale Immunobiology Program, The Anlyan Center for Medical Research & Education, New Haven, Connecticut, USA

⁷Yale Comprehensive Cancer Center, New Haven, Connecticut, USA

⁸Yale Stem Cell Center, New Haven, Connecticut, USA

Abstract

The genetic interactions influencing metastatic potential have been challenging to investigate systematically. Here we developed MCAP (massively parallel CRISPR-Cpf1/Cas12a crRNA array profiling), an approach for combinatorial interrogation of double knockouts *in vivo*. We designed an MCAP library of 11,934 arrays targeting 325 pairwise combinations of genes implicated in metastasis. By assessing the metastatic potential of the double knockouts in mice, we unveiled a quantitative landscape of genetic interactions driving metastasis.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence: SC (sidi.chen@yale.edu), +1-203-737-3825 (office), +1-203-737-4952 (lab).

Contributions

R.D.C. designed the study, performed experiments, developed statistical algorithms and computational pipelines, and analyzed the data. G.W. and L.Y. performed molecular, cellular, animal experiments and MCAP readout. A.C. optimized *in vivo* screens and assisted with validation experiments. H.R.K. and L.S. assisted with cloning and validation experiments. M.B.D. assisted with flow cytometry experiments. Y.E. assisted with *in vitro* assays. S.C. conceived the study, provided conceptual advice, secured funding and supervised the work. R.D.C. and S.C. wrote the manuscript.

[†] Co-first authors

Competing Financial Interests

The authors have filed a provisional patent related to this work.

Metastasis, the major lethal factor of solid tumors, is a complex multi-step process¹. A systems-level understanding of the genetic interactions influencing metastatic potential is lacking, as library-scale *in vivo* interrogation of double knockouts (DKOs) in mammalian species has been challenging. The type V CRISPR system Cpf1 (also known as Cas12a) has empowered simultaneous genome editing at multiple loci²⁻⁴. As Cpf1 does not require a tracrRNA, multiplexed genome editing can be achieved with a single crRNA array^{3,4}. This characteristic inspired us to develop Cpf1 as a system for interrogating genetic interactions *in vivo*, with substantial advantages in library design, readout and analysis compared to Cas9-based approaches.

We first established a CRISPR-Cpf1 lentiviral system for characterization of double knockouts in a cancer cell line (KPD)^{5,6} (Supplementary Fig. 1). To evaluate the cellular diversity that can be accommodated *in vivo*, we cloned in a library of random 8mers and transplanted 4×10^6 8mer-barcoded cells into *nu/nu* (n = 2) or *Rag1*^{-/-} mice (n = 4). Of the 65,536 possible 8mers, an average of 65,534.5 (99.99%) were recovered in *nu/nu* mice and $64,500.75 \pm 940.58$ (mean \pm s.e.m.) (98.42%) in *Rag1*^{-/-} mice, 12 days post-transplant (Supplementary Fig. 2).

We then sought to develop **M**assively parallel **C**pf1 crRNA **A**rray **P**rofilng (MCAP), an approach for high-throughput screening of DKOs. We focused on genes significantly mutated in a human metastasis cohort (MET-500)⁷ and the top hits from a single knockout (SKO) metastasis screen in mice⁵ (Fig. 1a and Supplementary Table 1). We selected 4 crRNAs for each of the 26 metastasis driver candidates. Compiling these 104 gene-targeting crRNAs and 52 non-targeting control (NTC) crRNAs, we designed a metastasis-focused MCAP library (MCAP-MET) composed of 1,326 NTC-NTC arrays, 5,408 SKO arrays and 5,200 DKO arrays, for a total of 11,934 dual-crRNA arrays (Supplementary Table 2). In the MCAP-MET library, each gene pair is represented by 16 DKO constructs, while each gene is represented by 208 SKO constructs. Additionally, we appended a random 10mer barcode for clonal analyses. Deep sequencing confirmed complete coverage of the library, and analysis of the 10mer barcodes revealed the diversity of barcoded crRNA-arrays (BC-arrays) (n = 774,295) (Supplementary Fig. 3a-d).

We generated lentiviral pools from the MCAP-MET plasmid library and infected Cpf1⁺ KPD cells (Fig. 1b). 7 and 14 days following transduction, we sequenced the crRNAs in the cell pool and found strong concordance with the plasmid library (Supplementary Fig. 3e). In each cell pool we recovered $172,427 \pm 2,591$ (mean \pm s.e.m.) unique BC-arrays. To map the metastatic potential of the MCAP-MET library, we injected the cell pool subcutaneously into *nu/nu* mice (4×10^6 cells per mouse, ~ 350 x coverage) (n = 10). At this coverage, each BC-array is represented by an average of ~ 23 cells upon injection. After 6 weeks, we collected the primary tumors (n = 10) and lung lobes (n = 37), and performed crRNA array sequencing. Using the BC-array data, we assessed the dynamics of selection in our metastasis model. We chose a 0.001% cutoff by considering the distribution of BC-array frequencies in cell samples and quantified the number of “clones” (approximated by BC-arrays) per sample, finding clear evidence of progressive selection as the cell pools formed primary tumors and lung metastases (Supplementary Fig. 4). These results were consistent at a 0.01% frequency cutoff (Supplementary Fig. 5). Collectively, the clone-level analyses

illustrate the progressive selection pressures on the cells as they form primary tumors and metastasize to the lung.

We next considered the data in terms of the 11,934 dual-crRNA arrays (Supplementary Fig. 6 and Supplementary Table 3). Utilizing the 1,326 NTC-NTC arrays as an empirical null distribution, we identified crRNA arrays enriched at a false discovery rate (FDR) < 0.5% in each sample (Supplementary Fig. 7a). We tabulated the percentage of arrays for a given genetic perturbation that were enriched in at least one sample (Supplementary Fig. 7b–c). No single genes had more than 40% of their SKO arrays enriched in lung metastases. In contrast, 62.5% of all arrays targeting the *Nf2_Rb1* pair were enriched in at least one lung metastasis, with 56.25% of arrays enriched for *Nf2_Pten* and *Nf2_Trim72* (Supplementary Fig. 7d–f).

We quantitatively determined the metastatic potential of the various perturbations represented in the MCAP-MET library (Supplementary Fig. 8a–c). To identify specific perturbations exhibiting strong selection *in vivo*, we averaged the crRNA arrays for each SKO or DKO condition on a sample-by-sample basis, then aggregated the data by sample type. In order to pinpoint the perturbations with the strongest selective advantage out of the entire MCAP-MET library, we used all targeting genes/pairs for linear regression modeling. The top gene pairs favored in primary tumors relative to cell pools (outlier test, adjusted $p < 0.05$) included *Nf2_Trim72*, *Nf2_Chd1*, *Nf2_Pten*, *Nf2_Arid1b*, *Nf2_Kdm6a*, and *Nf2_Rb1* (Fig. 1c). A similar set of gene pairs were enriched in lung metastases compared to cell pools (Supplementary Fig. 8d). Comparing primary tumors to lung metastases, *Nf2_Trim72* and *Nf2_Chd1* emerged as the top metastasis-driving mutation pairs (Fig. 1d).

Our analyses suggested that certain gene pairs may be synergistic in promoting metastasis. To identify such mutation combinations, we first identified gene pairs that were significantly more abundant than their respective single gene counterparts (two-sided Wilcoxon rank sum test, adjusted $p < 0.05$). Since the effects of a mutation combination may simply be additive rather than synergistic, we calculated a synergistic coefficient ($\text{SynCo} = \text{DKO}_{\text{NM}} - \text{SKO}_{\text{N}} - \text{SKO}_{\text{M}}$) for each gene pair (Fig. 2a and Supplementary Table 4). Collectively, we found 6 DKOs that were significantly more abundant than the corresponding SKOs and with a $\text{SynCo} > 0$: *Nf2_Trim72*, *Chd1_Nf2*, *Chd1_Kmt2d*, *Jak1_Kmt2c*, *Kmt2d_Pten*, and *Nf1_Pten* (Fig. 2b–d and Supplementary Fig. 9). These data were summarized as a library-wide map of the selective advantage of each DKO relative to the corresponding SKOs (Supplementary Fig. 10 and Supplementary Table 5). Some of these synergistic interactions are recapitulated in human cohorts (Supplementary Fig. 11 and Supplementary Table 6).

We then sought to validate the metastatic potential of the strongest gene pair identified in the screen, *Nf2_Trim72*. After cloning in 5 different dual-crRNA arrays with combinations of *Rosa26*-targeting crRNAs or the top-performing *Nf2* and *Trim72* crRNAs from the screen (*Rosa26+Rosa26*, *Nf2+Rosa26*, *Trim72+Rosa26*, *Nf2+Trim72*, or *Trim72+Nf2*), we assessed mutation efficiency 7 days following lentiviral transduction (n = 5 infection replicates each) and confirmed that array configuration does not influence mutation efficiency (Fig. 3a and Supplementary Fig. 12a). To exclude the possibility that the *Nf2_Trim72* gene pair may have undergone positive selection *in vitro* prior to injection, we

characterized the EdU incorporation of KPD cells expressing Rosa26+Rosa26, Nf2+Rosa26, Trim72+Rosa26, or Nf2+Trim72 dual-crRNA arrays, finding no significant differences (n = 3 cell replicates each) (Supplementary Fig. 12b–c).

To interrogate the metastatic potential of the *Nf2_Trim72* gene pair, we first performed *in vitro* Matrigel invasion assays (n = 3 independent experiments), finding that Nf2+Trim72 cells were more invasive compared to Rosa26+Rosa26, Nf2+Rosa26, or Trim72+Rosa26 cells (Supplementary Fig. 12d–e). We then proceeded to validate the *Nf2_Trim72* gene pair *in vivo*, transplanting 1.8×10^6 cells into *nu/nu* mice (n = 8 mice for each condition). Primary tumors in the Nf2+Trim72 group grew significantly larger than Nf2+Rosa26, Trim72+Rosa26, or Rosa26+Rosa26 tumors (Fig. 3b). We followed the development of metastasis by luciferase live imaging, and 28 days following the initial transplantation, we harvested the primary tumors and lungs (Supplementary Fig. 13). Mice bearing Nf2+Trim72 tumors had significantly more metastatic lung nodules than mice bearing Nf2+Rosa26, Trim72+Rosa26, or Rosa26+Rosa26 tumors (Fig. 3c). Collectively, these data point to specific mutation combinations with heightened metastatic potential *in vivo*, and highlight the power of MCAP for high-throughput interrogation of genetic interactions in challenging biological systems.

Several high-throughput double perturbations have been performed in mammalian cells using RNA interference (RNAi) or CRISPR-Cas9 technologies^{8–16}. However, the dependence of Cas9 on a trans-activating crRNA (tracrRNA) predicates the need for multiple sgRNA cassettes when performing combinatorial knockouts, thus complicating library design, cloning, readout, and analysis. In comparison, MCAP offers a streamlined approach for double or even higher-order knockout/perturbation screens, with the potential for sequential screens using invertible dual-crRNA arrays¹⁷. A remaining challenge that limits the broader utility of MCAP is the mutation efficiency of Cpf1, as it necessitates positive selection screens using redundant library designs with several independent constructs representing each perturbation. Of note, progress has been made towards predicting crRNAs that can induce mutations at higher efficiencies^{18,19}, and Cpf1 itself has been engineered to increase its activity and targeting range²⁰.

MCAP can be readily applied to different cell types, biological processes, and disease models and thus represents a tool for mapping genetic interactions in mammalian species *in vivo* with unparalleled simplicity and throughput.

Online Methods

Please also refer to the associated Supplementary Protocol for additional information²¹, as well as the Life Sciences Reporting Summary.

Animal work statements and institutional approval

All experimental work involving recombinant DNA was performed under the guidelines of the Yale University Environment, Health and Safety (EHS) committee under an approved protocol (Chen-rDNA-15–45). All animal work was performed under the guidelines of Yale University Institutional Animal Care and Use Committee (IACUC) with approved protocols

(Chen-2015–20068; Chen-2018–20068), and was consistent with the Guide for Care and Use of Laboratory Animals, National Research Council, 1996 (Institutional Animal Welfare Assurance No. A-3125–01). 6–8 week old mice, both males and females, were used for MCAP screen experiments. For subsequent validation experiments, only female mice were used.

Design of the MCAP-MET library

The top 23 ranked “tumor suppressors” from the human MET500 cohort⁷ were compiled, and combined with 3 top hits from a previous mouse metastasis screen (*Nf2*, *Trim72*, and *Ube2g2*)⁵ for a final set of 26 genes. We then analyzed the complete exon sequences of these 26 genes to extract all possible Cpf1 spacers (i.e., all 20mers beginning with the Cpf1 PAM, 5'-TTTV). Each of these 20mers was then reverse complemented and mapped to the entire mm10 reference genome by Bowtie 1.1.2²², with settings -n 2 -l 18 -p 8 -a -y --best -e 90. After filtering out all alignments that contained mismatches in the final 3 basepairs (corresponding to the Cpf1 PAM) and disregarding any mismatches in the fourth to last basepair, we quantified the number of genome-wide alignments for each crRNA using all 0, 1, and 2 mismatch (mm) alignments. A total mismatch score (MM score) was calculated for each crRNA using the following formula: MM score = 0mm*1000 + 1mm*50 + 2mm*1. We also counted the number of consecutive thymidines in each crRNA, and used the following formula: T score = 100 / (max_consecutive_Thymidines)². We then sorted all the 20nt crRNAs corresponding to each target gene by low MM score and high T score. Finally, the top 4 crRNAs for each gene were chosen. In the event of ties, crRNAs targeting constitutive exons and/or the first exon were prioritized.

52 NTC crRNAs were randomly selected from a pool of random 20mers that did not map to the mouse genome with up to 2 mismatches. In combination with the 104 crRNAs targeting 26 genes, a total of 5,200 DKO, 5,408 SKO, and 1,326 NTC-NTC arrays were designed for a total of 11,934 dual-crRNA arrays (MCAP-MET library). With a total pool of 26 genes, the number of possible unique combinations of two different genes is 325. Each of these 325 gene pairs was represented by 16 DKO arrays, while each single gene condition was represented by 208 SKO arrays. For SKO crRNA arrays, we placed each gene-targeting crRNA in the first position of the crRNA array and toggled the NTC crRNAs through the second position. For each gene pair, the positioning of the crRNAs representing each of the two genes was determined randomly. For each oligo, we appended a degenerate 10mer (10xN) following the U6 termination sequence to serve as a barcode for downstream clonality analysis. After pooled oligo synthesis (CustomArray), we used Gibson cloning to insert the MCAP-MET library into the BsmBI-linearized crRNA expression vector (pLenti-U6-DR-crRNA-Puro-P2A-Firefly luciferase).

Cell lines

A non-small cell lung cancer (NSCLC) cell line^{5,6} (KPD cell line) was transduced with pLenti-EFs-Cpf1-Blast to generate Cpf1-positive cells (KPD-Cpf1). All cell lines were grown under standard conditions using DMEM containing 10% FBS, 1% Pen/strep in a 5% CO₂ incubator.

Lentiviral library production

Briefly, envelope plasmid pMD2.G, packaging plasmid psPAX2, and pLenti-MCAP-MET plasmid were added at ratios of 1:1.5:2, and then polyethyleneimine (PEI) was added and mixed well by vortexing. The solution was left at room temperature for 10–20 min, and then the mixture was added dropwise into 80–90% confluent HEK293FT cells and mixed well by gently agitating the plates. Six hours post-transfection, fresh DMEM supplemented with 10% FBS and 1% Pen/Strep was added to replace the transfection media. Virus-containing supernatant was collected at 48 h and 72 h post-transfection, and was centrifuged at 1500 g for 10 min to remove the cell debris, then aliquoted and stored at –80°C. Virus was titrated by infecting KPD cells at a number of different concentrations, followed by the addition of 3 µg/mL puromycin at 24 h post-infection to select the transduced cells. The viral titers were determined by calculating the ratios of surviving cells 48 or 72 h post infection and the cell count at infection.

Nextera analysis of indels generated by Cpf1

crRNA arrays (crPten-crNf1 and crNf1-crPten) were cloned into the pLenti-U6-DR-crRNA-Puro vector, and virus was generated for transduction of KPD-Cpf1 cells.

Pten spacer = TGCATACGCTATAGCTGCTT

Nf1 spacer = TAAGCATAATGATGATGCCA

Six days after transduction and puromycin selection, genomic DNA was harvested from the cells in culture. The surrounding genomic regions flanking the target sites of crPten and crNf1 were first amplified by PCR using the following primers (5' – 3'):

Pten_F = ACTCACCAGTGTTTAACATGCAGGC

Pten_R = GGCAAGGTAGGTACGCATTTGCT

Nf1_F = AGCAGCTGTCCTGGCTGTTC

Nf1_R = CGTGCACCTCCCTTGTCAGG

PCR conditions: Using Phusion Flash High Fidelity Master Mix (ThermoFisher), the thermocycling parameters were: 98 °C for 2min, 35 cycles of (98 °C for 1s, 62 °C for 5s, 72 °C for 15 s), and 72 °C for 2 min.

Nextera XT library preparation was then performed according to manufacturer protocol with minor modifications. Reads were mapped to the mm10 mouse genome using BWA²³, with the settings `bwa mem -t 8 -w 200`. Indel variants were first processed with Samtools²⁴ with the settings `samtools mpileup -B -q 15 -d 1000000000000`, then input into VarScan v2.3.9²⁵ with the settings `pileup2indel --min-coverage 2 --min-reads 2 --min-var-freq 0.00001`. Variants occurring within a ± 7nt window of the predicted crRNA cut sites were summed to obtain total mutation frequencies.

Evaluation of *in vivo* library diversity in the absence of mutagenesis

We synthesized a library of degenerate 8mers and cloned them into the crRNA expression vector. After lentiviral production, KPD cells were transduced with the 8mer lentiviral

library and selected by puromycin. 4×10^6 KPD-8mer cells were subcutaneously injected in either Rag1^{-/-} or *nu/nu* mice. 12 days post-transplantation, mice were sacrificed and tumors were isolated for genomic preparation and readout.

MCAP in a mouse model of metastasis

Library transduction was performed with three infection replicates at high coverage and low MOI. Briefly, according to the viral titers, MCAP-MET lentiviruses were added to a total of 1×10^8 KPD-Cpf1 cells at calculated MOI of 0.2 and incubated 24 h before replacing the virus-containing media with 3 $\mu\text{g}/\text{mL}$ puromycin containing fresh media to select the virus-transduced cells. Approximately 2.5×10^7 cells confer a $\sim 2,000\times$ library coverage. MCAP-MET library-transduced cells were cultured under the pressure of 3 $\mu\text{g}/\text{mL}$ puromycin for 14 days before injection. MCAP library-transduced KPD-Cpf1 cells were injected subcutaneously into the right flank of *nu/nu* mice at 4×10^6 cells per flank ($\sim 350\times$ coverage per transplant).

Mouse tumor dissection

Mice were sacrificed by carbon dioxide asphyxiation followed by cervical dislocation. Tumors and lungs were manually dissected, then fixed in 10% formalin for 24–96 hours, and transferred into 70% ethanol. Tissues were flash frozen with liquid nitrogen, and ground in 5 mL Frosted polyethylene vial set (2240-PEF) in a 2010 GenoGrinder machine (SPEXSamplePrep). Homogenized tissues were then used for DNA extraction.

Genomic DNA extraction

200–800 mg of frozen ground tissue were re-suspended in 6 mL of NK Lysis Buffer (50 mM Tris, 50 mM EDTA, 1% SDS, pH 8.0) supplemented with 30 μL of 20 mg/mL Proteinase K (Qiagen) in 15 mL conical tubes, and incubated at 55 °C bath overnight. After all the tissues were lysed, 30 μL of 10 mg/mL RNase A (Qiagen) was added, mixed well and incubated at 37 °C for 30 min. Samples were chilled on ice and then 2 mL of pre-chilled 7.5 M ammonium acetate (Sigma) was added to precipitate proteins. The samples were inverted and vortexed for 15–30s and then centrifuged at 4,000 *g* for 10 min. The supernatant was carefully decanted into a new 15 mL conical tube, followed by the addition of 6 mL 100% isopropanol (at a ratio of ~ 0.7), inverted 30–50 times and centrifuged at 4,000 *g* for 10 minutes. At this time, genomic DNA became visible as a small white pellet. After discarding the supernatant, 6 mL of freshly prepared 70% ethanol was added, mixed well, and then centrifuged at 4,000 *g* for 10 min. The supernatant was discarded by pouring; and remaining residues was removed using a pipette. After air-drying for 10–30 min, DNA was re-suspended by adding 200–500 μL of Nuclease-Free H₂O. The genomic DNA concentration was measured using a Nanodrop (Thermo Scientific), and normalized to 1000 ng/ μL for the following readout PCR.

MCAP library readout

MCAP library readout was performed using a 2-step PCR approach. Briefly, in the 1st round PCR, enough genomic DNA was used as template to guarantee coverage of the library abundance and representation. 12 μg of gDNA was used per sample, split over 6 separate

PCR reactions. For the 1st PCR, the sgRNA-containing region was amplified using primers specific to the MCAP vector using Phusion Flash High Fidelity Master Mix (ThermoFisher) with thermocycling parameters: 98 °C for 1 min, 15 cycles of (98 °C for 1s, 60 °C for 5s, 72 °C for 15s), and 72 °C for 1 min.

Fwd AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCG

Rev CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCC

In the 2nd PCR, 1st round PCR products for each biological replicate were pooled, then 2 µL of well-mixed 1st PCR products were used as the template for amplification using sample-tracking barcode primers with thermocycling conditions as 98 °C for 1 min, 15 cycles of (98 °C for 1s, 60 °C for 5s, 72 °C for 15s), and 72 °C for 1 min. The 2nd PCR products were quantified in 2% E-gel EX (Life Technologies) using E-Gel® Low Range Quantitative DNA Ladder (ThermoFisher), then the same amount of each barcoded samples were combined. The pooled PCR products were purified using QIAquick PCR Purification Kit and further QIAquick Gel Extraction Kit from 2% E-gel EX. The purified pooled library was quantified as above. Diluted libraries with 5–20% PhiX were sequenced with HiSeq 4000 systems (Illumina) with 150bp paired-end read length.

MCAP-MET plasmid library readout and analysis

Raw paired-end fastq read files were first merged to single fastq files by PEAR²⁶ with the settings -y 8G -j 8 -v 3. The merged fastq files were then filtered and demultiplexed using Cutadapt²⁷, using two different sets of adapters for extraction of crRNA array sequences or the 10mer barcode. For the crRNA array, we used the following settings: cutadapt --discard-untrimmed -g tcttGTGGAAAGGACGAAACACCg, followed by cutadapt --discard-untrimmed -a TGTAGATTTTTTTT. The trimmed sequences were then mapped to the MCAP-MET library using Bowtie²²: bowtie -v 3 -k 1 -m 1. For the 10mer barcodes, we used the following Cutadapt settings: cutadapt --discard-untrimmed -a aagcttggcgtGGATC, followed by cutadapt --discard-untrimmed -g TACTAAGTGTAGATTTTTTTT. The resultant sequences were quantified to a reference of all possible 10mer sequences. Reads that successfully mapped to both the MCAP-MET library and contained a valid barcode were tabulated.

Processing of MCAP-MET crRNA array abundance in cells and tumors

PEAR-merged²⁶ fastq files were filtered and demultiplexed using Cutadapt²⁷. To remove extra sequences downstream (i.e. 3' end) of the crRNA array sequences, including the DR and U6 terminator, we used the following settings: cutadapt --discard-untrimmed -e 0.1 -a aagcttggcgtGGATCCGATATCa -m 80. As the forward PCR primers used to readout crRNA array representation were designed to have a variety of barcodes to facilitate multiplexed sequencing, we then demultiplexed these filtered reads with the following settings: cutadapt -g file:fbc.fasta --no-trim, where fbc.fasta contained the 12 possible barcode sequences within the forward primers. Finally, to remove extraneous sequences upstream (i.e. 5' end) of the crRNA array spacers, we first used the following settings: cutadapt --discard-untrimmed -e 0.1 -g tcttGTGGAAAGGACGAAACACCg -m 80. Then, we removed the 5' DR as follows: cutadapt --discard-untrimmed -e 0.1 -g TAATTTCTACTAAGTGTAGAT -m

80. The filtered fastq reads were then mapped to the MCAP-MET reference index. To do so, we first generated a Bowtie index of the MCAP-MET library using the bowtie-build command in Bowtie 1.1.2²². Using these bowtie indexes, we mapped the filtered fastq read files using the following settings: bowtie -n 2 -k 1 -m 1 --best. These settings ensured only single-match reads would be retained for downstream analysis. For data processing on the level of barcoded-crRNAs, we utilized the same trimmed fastq files as above, but instead used the barcoded-crRNA plasmid library as the reference index.

Analysis of MCAP crRNA array library representation

Using the resultant mapping output, we quantified the number of reads that had mapped to each crRNA array within the library. We normalized the number of reads in each sample by converting raw crRNA array counts to reads per million (rpm). The rpm values were then subject to \log_2 transformation for certain analyses. Where applicable, linear regression lines and 95% confidence intervals were calculated. For comparing cells, primary tumors, and lung metastases, crRNA array abundances were averaged within sample groups and linear regression was performed using the NTC-NTC arrays as a model for neutral selection. Significant outliers were identified using the *outlierTest* function from the *car* R package, which calculates the studentized residuals of the linear regression and derives the corresponding p-values. The Benjamini-Hochberg procedure was then used to adjust p-values for multiple comparisons. For gene/gene pair analyses, the corresponding SKO and DKO arrays were first averaged together, then aggregated by sample type. Linear regression was performed using all SKO/DKO genotypes, and outliers were identified as above.

Clone-level analysis of MCAP-MET samples

We analyzed the data at the clone level using the barcoded-crRNA abundances. We first converted the counts in each sample to percentages of total reads. We then used two different frequency cutoffs for considering clones: 0.01% and 0.001%. Differences in the number of clones between sample types was assessed by two-sided Wilcoxon rank sum test, and visualized after \log_2 transform. Empirical CDFs were calculated after combining all the clones in a given sample group; statistical differences in clone size distributions was assessed by two-sided Kolmogorov-Smirnov test. The Shannon diversity index was also calculated on each sample with the *vegan* R package; statistical differences were assessed by two-sided Wilcoxon rank sum test.

Enrichment analysis of MCAP-MET genotypes

To identify crRNA arrays that were enriched in individual samples, we utilized the 1,326 NTC-NTC arrays for modeling the empirical null distribution. Enriched crRNA arrays were subsequently called at $FDR < 0.5\%$. These results were aggregated to the single gene/gene pair level, then tabulated across samples. Finally, we counted all of the significant crRNA arrays associated with each genotype.

Identification of synergistic mutation combinations

We defined the synergy coefficient (SynCo) for each gene pair with the following formula: $SynCo = DKO_{NM} - SKO_N - SKO_M$. The DKO_{NM} value is the median \log_2 rpm abundance of

all corresponding DKO crRNA arrays (i.e., crN-crM), while SKO_N and SKO_M values are defined as the median log₂ rpm abundance of all corresponding SKO crRNA arrays. We calculated the SynCo of each gene pair within the lung metastasis samples and further assessed whether the DKO abundances were statistically significantly higher than the corresponding SKO abundances by two-sided Wilcoxon rank sum test. We defined synergistic mutation combinations as gene pairs where 1) the SynCo score was > 0, and 2) the median differential abundances compared to the corresponding SKOs were both > 0.2, with an associated Benjamini-Hochberg adjusted $p < 0.05$ for both comparisons. To generate a library-wide map of the relative selective advantages for each gene pair vs. single gene knockout, we utilized the aggregated gene-level abundances in lung metastasis samples. We compared the abundance of each DKO to its reference SKO, and visualized the data in a heat map. Each column refers to the reference SKO, while each row denotes the modulatory effects of the second KO.

Design of dual-crRNA arrays for validation experiments

Dual-crRNA arrays containing combinations of *Rosa26*-targeting crRNAs or the best-performing *Nf2* and *Trim72* crRNA were designed. The following spacer sequences were used:

```
crRosa26.1 AGGCTATATTTCTGCTGTCT
crRosa26.2 TAGTTCAAAGCTTCTGACAG
crNf2 AAGGCCTCGATCTCCGTCTT
crTrim72 TGCCGTGCCTGCCTGATCCG
```

Insertion of NLS-GFP sequences into the crRNA expression vector

The primary screen experiments were performed using the U6 crRNA expression vector with an EFS promoter driving expression of puromycin and firefly luciferase. For validation experiments, the coding sequences for NLS-EGFP were inserted after puromycin-P2A-luciferase in the crRNA expression vector by Gibson cloning, with a P2A sequence separating the GFP.

Quantification of mutation frequency by T7E1

7 days following lentiviral transduction and puromycin selection, genomic DNA was extracted from the cells. PCR amplification of the genomic regions flanking the *Nf2* or *Trim72* crRNAs was performed using the following primers:

```
Nf2_F: CTCCTGAGGAAACTAGATGCCAACCT
Nf2_R: AAAGCTGTCTGTGGCAGGGTTATTTG
Trim72_F: GAGGAGAGGGCTGGGTATTTGAGAGA
Trim72_R: GCTGCCAAGCAAGGTAGGTAGCTATT
```

PCR conditions: Using Phusion Flash High Fidelity Master Mix (ThermoFisher), the thermocycling parameters were: 98 °C for 2 min, 35 cycles of (98 °C for 1s, 60 °C for 5s, 72 °C for 15 s), and 72 °C for 2 min.

The PCR amplicons were then used for T7E1 assays following the manufacturer protocol. Statistical significance was assessed by two-sided unpaired Welch's t-test.

EdU proliferation assay

To assess proliferation, we used the Click-iT EdU Alexa Fluor 647 Flow Cytometry Assay Kit (ThermoFisher, #C10419). We incubated cells in culture with 10 μ M for 2 hours, followed by fixation, permeabilization, and staining. Cells were then analyzed on a BD FACSAria and the data was processed using FlowJo. Statistical significance was assessed by two-sided unpaired Welch's t-test.

Matrigel invasion assay

For *in vitro* assessment of invasive potential, unsupplemented DMEM was first mixed with standard Matrigel (Corning #356234) on ice using pre-chilled pipette tips to a final concentration of 25% Matrigel. After placing FluoroBlok cell culture inserts with 8 μ m pores into a 24 well plate, 100 μ l of the 25% Matrigel was added onto each insert. The inserts were incubated in the cell culture incubator for 1 hour to solidify the Matrigel. Cultured cells were then resuspended in unsupplemented DMEM at a concentration of 0.5×10^6 cells/ml, and 200 μ l of the cell suspension was gently added on top of the Matrigel layer. Finally, 600 μ l of 10% FBS DMEM was added to each well, underneath the inserts. Invasive cells were quantified using an inverted microscope 24 hours later on the GFP channel. Statistical significance was assessed by two-sided unpaired Welch's t-test.

Luciferase imaging for tracking metastasis

Mice were anesthetized by isoflurane inhalation and imaged for metastasis using an IVIS machine (PerkinElmer) 5 minutes following intraperitoneal injection of firefly d-luciferin potassium salt (150 mg/kg body weight).

Quantification of primary tumors and lung metastases

Mice were anesthetized by isoflurane inhalation and tumor sizes were quantified every 2–3 days by caliper using the formula Volume (mm^3) = $\pi/6 * x * y * z$. Statistical significance was assessed by two-way ANOVA, jointly considering the effect of time and treatment condition. Mice were euthanized at 28 dpi, and lungs were harvested for quantification of lung metastases. Each lung lobe was separately visualized under a dissecting microscope. Lung lobe metastases were quantified on bright-field images with real-time confirmation by GFP expression. Statistical significance was assessed by two-sided unpaired Welch's t-test.

Genomic comparisons of human primary tumors and metastases

Mutation frequencies from the TCGA PanCancer dataset and the MET500 dataset were filtered for the 26 genes represented in the MCAP-MET library. Statistical significance of the Spearman correlation was determined by calculating the t-statistic of the correlation. Identification of gene pairs that were significantly co-mutated was determined by hypergeometric test.

Statistics

All statistical tests are unpaired and two-sided. Details about the statistical tests are described in the corresponding figure legends and methods.

Blinding statement

Investigators were not blinded for sequencing data analysis, tumor engraftment, or organ dissection.

Code availability

Key scripts used to process and analyze the data will be available to academic community upon reasonable request.

Data and resource availability

MCAP data, sequences of oligos, and library design are described in the Methods section and Supplementary Tables. All vectors and libraries have been deposited to Addgene and are available to the academic community. Cell lines and all data supporting this work will be available to the academic community upon reasonable request to the corresponding author. Genomic sequencing data has been deposited to NCBI SRA (PRJNA515306).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all members in the Chen laboratory, as well as various colleagues in Department of Genetics, Systems Biology Institute, Immunobiology Program, BBS Program, MSTP Program, Comprehensive Cancer Center and Stem Cell Center at Yale, for their assistance and scientific discussion. We thank the Center for Genome Analysis, Center for Molecular Discovery, Pathology Tissue Services, Histology Services, High Performance Computing Center, West Campus Analytical Chemistry Core and West Campus Imaging Core and Keck Biotechnology Resource Laboratory at Yale, for technical support.

S.C. is supported by Yale SBI/Genetics Startup Fund, Damon Runyon Dale Frey Award (DFS-13-15), Melanoma Research Alliance (412806, 16-003524), St-Baldrick's Foundation (426685), Breast Cancer Alliance, Cancer Research Institute (CLIP), AACR (499395, 17-20-01-CHEN), The Mary Kay Foundation (017-81), The V Foundation (V2017-022), Ludwig Family Foundation, DoD (W81XWH-17-1-0235), Sontag Foundation (DSA Award), Chenevert Family Foundation, and NIH/NCI (1DP2CA238295-01, 1R01CA231112-01, 1U54CA209992-8697, 5P50CA196530-A10805, 4P50CA121974-A08306). R.D.C. and M.B.D. are supported by the Yale MSTP training grant from NIH (T32GM007205). G.W. is supported by CRI Irvington and RJ Anderson Postdoctoral Fellowships. A.C. is supported by Yale PhD training grant from NIH (T32GM007223).

Main References

1. Lambert AW, Pattabiraman DR & Weinberg RA Emerging Biological Principles of Metastasis. *Cell* 168, 670–691 (2017). [PubMed: 28187288]
2. Schunder E, Rydzewski K, Grunow R & Heuner K First indication for a functional CRISPR/Cas system in *Francisella tularensis*. *Int. J. Med. Microbiol. IJMM* 303, 51–60 (2013). [PubMed: 23333731]
3. Zetsche B et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771 (2015). [PubMed: 26422227]
4. Zetsche B et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol* 35, 31–34 (2017). [PubMed: 27918548]

5. Chen S et al. Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. *Cell* 160, 1246–1260 (2015). [PubMed: 25748654]
6. Kumar MS et al. Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev* 23, 2700–4 (2009). [PubMed: 19903759]
7. Robinson DR et al. Integrative clinical genomics of metastatic cancer. *Nature* 548, 297–303 (2017). [PubMed: 28783718]
8. Wong ASL, Choi GCG, Cheng AA, Purcell O & Lu TK Massively parallel high-order combinatorial genetics in human cells. *Nat. Biotechnol* 33, 952–961 (2015). [PubMed: 26280411]
9. Wong ASL et al. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci* 113, 2544–2549 (2016). [PubMed: 26864203]
10. Roguev A et al. Quantitative genetic-interaction mapping in mammalian cells. *Nat. Methods* 10, 432–437 (2013). [PubMed: 23407553]
11. Han K et al. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol* advance online publication, (2017).
12. Shen JP et al. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* advance online publication, (2017).
13. Horlbeck MA et al. Mapping the Genetic Landscape of Human Cells. *Cell* 0, (2018).
14. Najm FJ et al. Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol* 36, 179–189 (2018). [PubMed: 29251726]
15. Erard N, Knott SRV & Hannon GJ A CRISPR Resource for Individual, Combinatorial, or Multiplexed Gene Knockout. *Mol. Cell* 67, 348–354.e4 (2017). [PubMed: 28732207]
16. Boettcher M et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol* 36, 170–178 (2018). [PubMed: 29334369]
17. Chow RD, Kim HR & Chen S Programmable sequential mutagenesis by inducible Cpf1 crRNA array inversion. *Nat. Commun* 9, 1903 (2018). [PubMed: 29765043]
18. Kim HK et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat. Methods* 14, 153–159 (2017). [PubMed: 27992409]
19. Kim HK et al. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat. Biotechnol* 36, 239–241 (2018). [PubMed: 29431740]
20. Kleinstiver BP et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol* 1 (2019). doi:10.1038/s41587-018-0011-0.
21. Chow RD, Wang G, Ye L, & Chen S In vivo combinatorial knockout screens using CRISPR-Cpf1. *Protocol Exchange*. doi: 10.1038/protex.2019.018.
22. Langmead B, Trapnell C, Pop M & Salzberg SL Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25 (2009). [PubMed: 19261174]
23. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl* 25, 1754–1760 (2009).
24. Li H et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009). [PubMed: 19505943]
25. Koboldt DC et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576 (2012). [PubMed: 22300766]
26. Zhang J, Kobert K, Flouri T & Stamatakis A PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620 (2014). [PubMed: 24142950]
27. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12 (2011).

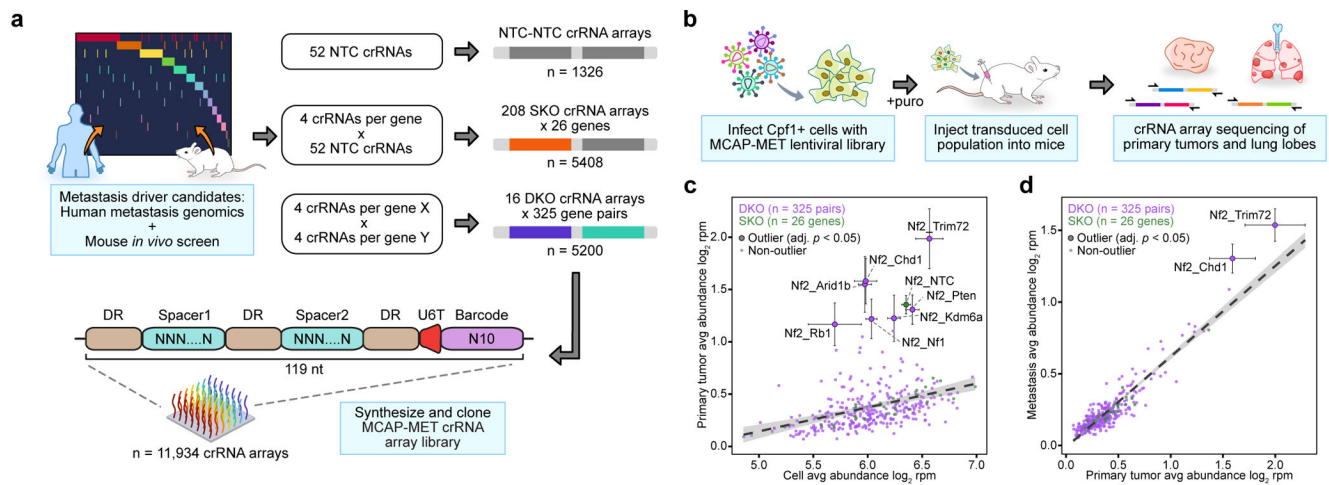


Figure 1: *In vivo* profiling of metastatic double knockouts by massively parallel CRISPR-Cpf1 crRNA array profiling (MCAP)

a. Schematic describing library design for massively parallel CRISPR-Cpf1 crRNA array profiling (MCAP) of metastasis driver combinations. **b.** Experimental design for combinatorial interrogation of metastasis drivers *in vivo*.

c-d. Scatter plot of MCAP-MET single knockout (SKO, $n = 26$ genes) and double knockout (DKO, $n = 325$ gene pairs) abundances in **(c)** cell pools ($n = 6$ cell replicates) vs. primary tumors ($n = 10$ mice) or **(d)** primary tumors vs. lung metastases ($n = 37$ from 10 mice). Data shown in terms of average \log_2 rpm for the indicated sample type, after first averaging the constituent crRNA arrays for each gene/gene pair. The linear regression over the entire library is shown (95% CI shaded in). Significant outliers (two-sided outlier test, adjusted $p < 0.05$) are outlined and enlarged, with s.e.m. error bars.

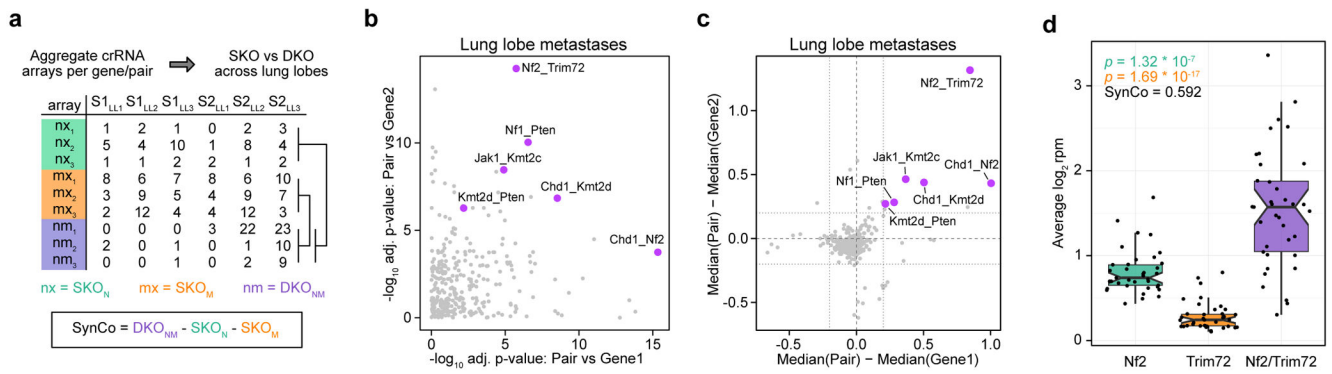


Figure 2: Identification of synergistic mutation combinations

a. Schematic for calculating the synergy coefficient score (SynCo) and identifying synergistic mutation combinations. For a given gene pair NM, the SynCo is defined as $DKO_{NM} - SKO_N - SKO_M$. A positive SynCo value indicates the selective advantage of the gene pair is greater than that of the two individual genes combined. **b-c.** Scatter plot of **(b)** $-\log_{10}$ adjusted p-values for each gene pair (two-sided Wilcoxon rank sum test) or **(c)** median differential abundance compared to the corresponding single genes, in lung metastases ($n = 37$ from 10 mice). Synergistic gene pairs are highlighted in purple. **d.** Tukey boxplots (IQR boxes with $1.5 \times IQR$ whiskers and notched 95% CI of median) detailing the abundances of *Nf2*, *Trim72*, or *Nf2_Trim72* arrays in lung metastases ($n = 37$ from 10 mice), with associated two-sided Wilcoxon rank sum p-values and SynCo scores noted. Statistics are in reference to *Nf2_Trim72* (purple) and colored according to the corresponding SKO conditions (green and orange).

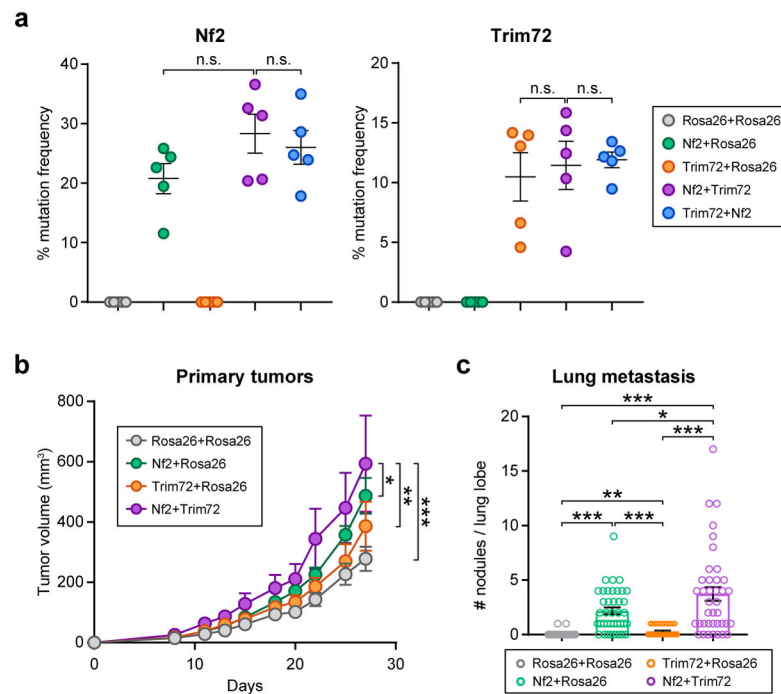


Figure 3: *Nf2* and *Trim72* mutations jointly promote lung metastasis *in vivo*

a. Quantification of T7E1 assays ($n = 5$ infection replicates each) for *Nf2* and *Trim72* (mean \pm s.e.m.). *Nf2* locus: *Nf2*+Rosa6 vs. *Nf2*+*Trim72*, $p = 0.1098$; *Nf2*+*Trim72* vs. *Trim72*+*Nf2*, $p = 0.6110$. *Trim72* locus: *Trim72*+Rosa6 vs. *Nf2*+*Trim72*, $p = 0.7450$; *Nf2*+*Trim72* vs. *Trim72*+*Nf2*, $p = 0.8386$. The order of each crRNA within the array is indicated in the array names (i.e. *Nf2*+*Trim72* vs. *Trim72*+*Nf2*). Statistical significance was assessed by two-sided unpaired Welch's t-test. **b.** Growth curves of primary tumors derived from cells transduced with Rosa26+Rosa26, *Nf2*+Rosa26, *Trim72*+Rosa26, or *Nf2*+*Trim72* crRNA arrays (mean \pm s.e.m.) ($n = 8$ mice for each condition). *Nf2*+*Trim72* vs. *Nf2*+Rosa26, *Trim72*+Rosa26, and Rosa26+Rosa26: $p = 0.0396$, $p = 0.0026$, and $p = 1.483 \times 10^{-5}$ respectively. Statistical significance was assessed by two-way ANOVA. **c.** Quantification of lung metastases in mice bearing Rosa26+Rosa26, *Nf2*+Rosa26, *Trim72*+Rosa26, or *Nf2*+*Trim72* primary tumors at 28 dpi. Data are shown in terms of the number of nodules found in each lung lobe (mean \pm s.e.m.) ($n = 4-5$ lung lobes per mouse, with 8 mice for each condition). *Nf2*+*Trim72* vs. *Nf2*+Rosa26, *Trim72*+Rosa26, and Rosa26+Rosa26: $p = 0.0328$, $p = 4.263 \times 10^{-6}$, and $p = 1.054 \times 10^{-6}$, respectively. *Nf2*+Rosa26 vs. Rosa26+Rosa26 and *Trim72*+Rosa26: $p = 5.091 \times 10^{-8}$ and $p = 8.990 \times 10^{-7}$. *Trim72*+Rosa26 vs. Rosa26+Rosa26, $p = 0.0016$. Statistical significance was assessed by two-sided unpaired Welch's t-test. n.s.: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.