

UCSF

UC San Francisco Previously Published Works

Title

Combinatorial genetics in liver repopulation and carcinogenesis with a in vivo CRISPR activation platform

Permalink

<https://escholarship.org/uc/item/8wd4g8cq>

Journal

Hepatology, 68(2)

ISSN

0270-9139

Authors

Wangenstein, Kirk J

Wang, Yue J

Dou, Zhixun

et al.

Publication Date

2018-08-01

DOI

10.1002/hep.29626

Peer reviewed



Published in final edited form as:

*Hepatology*. 2018 August ; 68(2): 663–676. doi:10.1002/hep.29626.

## Combinatorial genetics in liver repopulation and carcinogenesis with a novel *in vivo* CRISPR activation platform

Kirk J. Wangenstein<sup>1,2</sup>, Yue J. Wang<sup>2</sup>, Zhixun Dou<sup>2,3</sup>, Amber W. Wang<sup>2</sup>, Elham Mosleh-Shirazi<sup>2</sup>, Max A. Horlbeck<sup>4</sup>, Luke A. Gilbert<sup>4</sup>, Jonathan S. Weissman<sup>4</sup>, Shelley L. Berger<sup>2,3</sup>, and Klaus H. Kaestner<sup>2</sup>

<sup>1</sup>Department of Medicine, Division of Gastroenterology, University of Pennsylvania, Philadelphia, PA, 19104, USA

<sup>2</sup>Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104, USA

<sup>3</sup>Epigenetics Institute, Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA, 19104, USA

<sup>4</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, 94158, USA

### Abstract

CRISPR/Cas9 activation (CRISPRa) systems have enabled genetic screens in cultured cells lines to discover and characterize drivers and inhibitors of cancer cell growth. We adapted this system for use *in vivo* to assess whether modulating endogenous gene expression levels can result in functional outcomes in the native environment of the liver. We engineered the *dCas9*<sup>+</sup> mouse, a Cre-inducible CRISPRa system for cell type-specific gene activation *in vivo*. We tested the capacity for genetic screening in live animals by applying CRISPRa in a clinically relevant model of liver injury and repopulation. We targeted promoters of interest in regenerating hepatocytes using multiple single guide RNAs (gRNAs), and employed high-throughput sequencing to assess enrichment of gRNA sequences during liver repopulation, and to link specific gRNAs to the initiation of carcinogenesis. All components of the CRISPRa system were expressed in a cell type-specific manner and activated endogenous gene expression *in vivo*. Multiple gRNA cassettes targeting a proto-oncogene were significantly enriched following liver repopulation, indicative of enhanced cell division of cells expressing the proto-oncogene. Furthermore, hepatocellular carcinomas developed containing gRNAs that activated this oncogene, indicative of cancer initiation events. Furthermore, we employed our system for combinatorial cancer genetics *in vivo*, as we found that while clonal hepatocellular carcinomas were dependent on the presence of the oncogene-inducing gRNAs, they were depleted for multiple gRNAs activating tumor suppressors.

**Conclusion**—The *in vivo* CRISPRa platform developed here allows for parallel and combinatorial genetic screens in live animals. This approach enables screening for drivers and suppressors of cell replication and tumor initiation.

## Keywords

CRISPRa; dCas9; FAH; Hereditary Tyrosinemia; Genetic Screen; Hepatocellular Carcinoma

---

Advances in sequencing technology have catalogued gene expression levels, epigenetic states, and transcription factor binding sites of many cell types in human and mouse(1-4). Despite these efforts, the function of a large number of mammalian genes and in particular their combinatorial effects remain unknown. Candidate gene approaches are labor intensive, while elucidating epistatic gene interactions is difficult or impossible to perform using conventional mouse genetic methods. For these reasons, functional and combinatorial assays that can predictably alter the expression levels of multiple genes in specific cell types, and can link alterations to phenotypes in live animals, will be extremely useful for the annotation of gene function. Furthermore, an *in vivo* platform that could screen and assess genetic interactions would greatly expand our knowledge of pathophysiology in the native environment of the organ of interest.

The CRISPR/Cas9 system from *Streptococcus pyogenes* has been adapted to study genetics in eukaryotes, and has facilitated genetic manipulation of specific loci in numerous cell lines and *in vivo* to generate genetically modified organisms. The native form of Cas9 binds to single guide RNA (gRNA) sequences, which direct the protein to complementary DNA elements where the endonuclease domain cleaves the targeted locus. Whole genome loss-of-function screens using the native form of Cas9 have been performed in cell lines, which have been transplanted into mice to study tumorigenesis(5). Cas9 has also been used to induce tumorigenesis by targeting up to ten different tumor suppressors *in vivo* in the mouse liver(6, 7). These studies delivered plasmids containing gRNA expression cassettes by hydrodynamic injection, which leads to DNA delivery and transient expression in roughly 20% of hepatocytes(8). However, transient expression of the wildtype Cas9 together with gRNAs results in permanent disruptions in the genome. Thus, although useful to model and compare a limited number of targeted mutations, scale-up of these systems to include large pools of gRNAs has been limited by difficulties in linking specific gRNAs to observed phenotypes, as the gRNA coding sequences themselves are lost from greater than 95% of induced tumors(6). These systems are further limited by the cytotoxic disruption of essential genes or from genomic stress induced by DNA breaks(9, 10). Thus, new *in vivo* screening methods are needed that are scalable and allow observed phenotypes to be linked to specific gRNAs.

CRISPR/Cas9 activation (CRISPRa) combines a catalytically dead, mutant Cas9 (dCas9) with transcriptional activators to enable gRNA-mediated, locus-specific gene activation(11-21). Rather than inducing DNA breaks, CRISPRa recruits RNA polymerase II in a site-specific manner to promote the expression of target genes, allowing gRNAs to be linked to specific outcomes. These ‘gain-of-function’ systems have been used to screen for genetic drivers of a number of biological processes in human cell lines(22).

CRISPRa systems are now capable of producing several log-fold gene activation by (1) linking multiple different types of transcriptional activator domains to dCas9(19, 22), or (2) by fusing dCas9 to a repeating peptide array known as the ‘SunTag’ domain(20, 21). The

SunTag domain allows for the docking of up to ten copies of the VP64 transcriptional activator to the intended target(20). CRISPRa is scalable genome-wide and has been employed in cell lines from multiple species(23). Recently, a CRISPRa system was used for a genetic screen for drug resistance in an *ex vivo* mouse leukemia model system; however, this system employed cancer cell lines and therefore did not examine tumor initiation *in vivo*(24). To our knowledge, CRISPRa has not yet been utilized *in vivo* in a whole-animal vertebrate system.

Herein, we developed an *in vivo* CRISPRa screening system, and apply it to a liver injury and repopulation model, the *Fah*<sup>-/-</sup> mouse model of Hereditary Tyrosinemia(25-27). *Fah*<sup>-/-</sup> mice undergo conditional, lethal hepatocyte injury from the accumulation of toxic metabolites of tyrosine. The injury is controlled by the administration of nitisinone, a drug that blocks an upstream enzyme to prevent liver injury(25). Hydrodynamic tail vein injection of plasmids that contain a *Fah* cDNA transgene, followed by removal of nitisinone, leads to repopulation of the *Fah*<sup>-/-</sup> liver with hepatocytes that stably express FAH<sup>26, 27</sup>. In addition, any 'cargo' genes on the plasmids will be expressed in repopulating hepatocytes as well(27, 28).

In this study, we applied CRISPRa to conduct parallel genetic screens during liver repopulation, and demonstrate the intended activation of target gene expression. We find that our system allows for genetic epistasis experiments, showing that upregulation of an endogenous oncogene locus promotes cell proliferation and initiation of carcinogenesis, and conversely that targeting of tumor suppressors inhibits tumorigenesis *in vivo*. We present CRISPRa as a valuable resource for scalable combinatorial genetic screens in mice.

## EXPERIMENTAL PROCEDURES

### Plasmid vectors

We obtained the dCas9-SunTag sequence used previously in cell lines (pHRdSV40-dCas9-10xGCN4\_v4-P2A-BFP, Addgene Cat. #60903)(20), and PCR-amplified the dCas9-SunTag sequence using primers dCas9\_FseI\_F (5' - AGAAGGCCGGCCTACGCGGCCACCAT-3') and dCas9\_FseI\_R (5' - GCATGGCCGGCCAAATTTTGTAAATCCAGAGGTTGA-3'), and inserted it into the *FseI* site of the *Rosa26* targeting vector Ai9 (Addgene Cat. #22799) to construct pAi9-dCas9.

A mouse codon-optimized transcriptional activator (TA) sequence was based on reference (20), except that the GFP sequence was removed. The cDNA was designed *in silico* and synthesized as a gene fragment, together with a downstream poly-adenylation sequence and a U6 promoter for gRNA synthesis (Genewiz Corp, NJ), and inserted this into plasmid pKT2/*Fah*-Cag//SB(27) to construct pKT2/*Fah*-TA//SB. The U6 promoter has two BspQI restriction sites downstream, which allows for sticky-end ligation of annealed oligonucleotides or digested PCR products containing gRNA sequences.

Based on a published computational design (20), we obtained oligonucleotides for cloning of gRNA expression constructs targeting *Myc*, *Tnfrsf1a*, or controls (Table S1), and for *Slc7a11* and *Tp53* (Table S2). We annealed the oligos at a concentration of 0.1  $\mu$ M in T4

DNA ligase buffer (Invitrogen). The ends were phosphorylated with polynucleotide kinase (New England Biolabs) for 60 minutes at 37°C, followed by 15 minutes heat inactivation at 65°C. The annealed oligonucleotides were pooled at a concentration of 0.1 nM each, and 0.09 ng of this pool was ligated to 50 ng of BspQI-digested plasmid pKT2/FAH-TA//SB or plasmid pKT2/Fah-U6//SB, which were then used to transform *E. coli*. Finally, we pooled up to 20,000 bacterial clones (for up to 500X average coverage per gRNA) and prepared plasmid DNA to generate TALib1 and TALib2, which represent pools of all possible plasmid clones (Sigma Maxiprep kit). For quality control, we picked and sequenced 20-46 random bacterial clones from each library, all of which contained the expected gRNA sequences. We analyzed the distribution of the gRNAs cloned in the plasmids and found that it was not different than expected for a random distribution.

### Derivation of the dCas9<sup>+</sup> mouse strain

The vector pAi9-dCas9 was electroporated into mouse embryonic stem (ES) cells, and Neomycin selection was performed to obtain putative homologous recombinant clones. We used PCR and Southern Blotting to identify ES cell clones with the correct integration of the construct, and performed karyotyping of ES cell clones to screen for correct chromosome number (data not shown). Correctly-targeted ES cells were then injected into hybrid B6SJL/F1 mouse blastocysts, followed by implantation into pseudopregnant female mice. The offspring were genotyped, and 14/16 mice were found to be chimeric for the targeted ES cells. Genotyping was performed with primers Rosa\_Com (5' - GAGGGGAGTGTGCAATACCT-3'), CAGenhR (5' - CAGGCGGGCCATTTACCGTAAG-3'), and ROSA-R (5' - GGAGCGGGAGAAATGGATATG-3'), yielding a specific PCR product at 251 bp for the gene replacement allele and 501 bp for the wildtype allele. The dCas9<sup>+</sup> mice were crossed with *Fah*<sup>-/-</sup> mice to obtain *Fah*<sup>-/-</sup>;dCas9<sup>+</sup> mice. All CRISPRa experiments were conducted using FAH-null, dCas9 heterozygous (*Fah*<sup>-/-</sup>;dCas9<sup>+</sup>) mice. dCas9<sup>+/+</sup> mice are being deposited with the Mutant Mouse Regional Resource to facilitate distribution to the research community.

### Mice

*Fah*<sup>-/-</sup> and *Fah*<sup>-/-</sup>;dCas9<sup>+</sup> mice were maintained on 7.5 µg/ml nitisinone (SOBI, Sweden) until the time of injection with plasmid libraries at 8-12 weeks of age. One week prior to injection of the plasmids, the mice were tail vein-injected with 10<sup>11</sup> particles of hepatocyte-tropic serotype 8 adeno-associated virus with a hepatocyte-specific thyroxin-binding globulin (TBG) promoter driving Cre expression (AAV-Cre) in 100 µl saline (AV-8-PV1091, Penn Vector Core). Plasmid pools (10 µg TALib1 or TALib2) were hydrodynamically injected into the tail vein as described previously(27, 28). Nitisinone was removed after the injection, and mice were euthanized one or two months later.

### Immunohistochemistry Staining

Liver tissues were fixed overnight in 4% paraformaldehyde, and then paraffin-sectioned (Penn Digestive Diseases Center Morphology Core). Immunohistochemistry was performed using standard protocols as described previously(27, 28). For FAH staining we employed rabbit anti-FAH (ab81087, Abcam), and goat anti-rabbit for secondary antibody (Vector

Labs). The dCas9 has a hemagglutinin (HA) tag, therefore we could stain tissues with a mouse anti-HA antibody (H3663, Sigma-Aldrich), and rabbit anti-mouse IgG1 secondary antibody (SAB3701173, Sigma-Aldrich). For MYC we used a rabbit anti-MYC antibody (sc764, Santa Cruz Biotechnology), and goat anti-rabbit as the secondary antibody (Vector Labs). Staining was performed similarly using antibodies to YAP1 (4912, Cell Signaling Technologies), AFP (sc8108, Santa Cruz Biotechnology), OPN (AF808, R&D Systems), and Ki67 (RM9106, Thermo Scientific).

### Western blotting

Tissues were lysed in buffer containing 20 mM Tris, pH 7.5, 137 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 1% NP-40, 10% glycerol, supplemented with cOmplete™ Mini Protease Inhibitor Cocktail (Thermo Scientific) and benzonase (Novagen) at 12.5 U/ml. The lysates were rotated at 4°C for 30 minutes. The lysates were centrifuged at 16,000 g for 10 minutes, and 100 µg protein was subjected to electrophoresis using NuPAGE 4-12% Bis-Tris precast gels (Life Technologies). After transferring to nitrocellulose membrane, Ponceau S (Sigma) was used to stain the membrane, and then washed away by TBS supplemented with 0.1% Tween 20. 5% non-fat milk in wash buffer was used to block the membrane at room temperature for 1 hour. Primary antibody (HA, Abcam ab91110) was diluted in 5% BSA in wash buffer, and incubated at 4°C overnight. The membrane was washed three times with wash buffer, each for 10 minutes, followed by incubation of HRP-conjugated secondary antibodies at room temperature for 1 hour, in 5% non-fat milk in wash buffer. The membrane was washed again three times, and imaged by film or by GE Amersham Imager 600.

### High throughput sequencing

DNA was extracted from approximately 400 mg fresh liver tissue per mouse using a column purification kit according to the manufacturer's recommendations (Machery-Nagel Corp). Illumina sequencing libraries were prepared from the genomic DNA with two rounds of PCR of the gRNA sequences. The first round of PCR was performed using 5-10 µg of genomic DNA (from three *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice and one *Fah*<sup>-/-</sup> control mouse repopulated for four weeks with TALib1) or 5 ng TALib1 (three replicates) and primers 'KWgRNA\_F' (5'-AATGATACGGCGACCACCGAGATCTACACTCTTCCCTTGTGGAAAGGACGAAACACCG-3') plus 'KWgRNA\_R' (5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGAGGGTTAATTGATATCGCTAGC-3'), with 25 cycles of amplification and an annealing temperature of 58°C. The resultant PCR product was purified with Ampure beads (Beckman Coulter, Inc.), or by gel extraction (Machery-Nagel Corp.). The second round of PCR was performed with primer 'Primer 1.1' (5'-AATGATACGGCGACCACCGAGATCT-3') plus Illumina standard index sequencing primers (New England Biolabs). Sequencing was performed on an Illumina MiSeq or HiSeq machine with custom primer 'KWgRNAcustSEQ\_F' (5'-TTTCCCTTGTGGAAAGGACGAAACACCG-3'). The Fastq files were de-multiplexed to give raw read counts for each of the gRNA for each sample. The proportions for each of the gRNA sequences for each of the samples were adjusted for the change in median value of the control gRNA (to account for a drop in the proportion of the control gRNA due to a rise in the proportion of *Myc* gRNA), then were normalized to the average proportions of each

gRNA in TALib1 or TALib2. In TALib1 and TALib2, one gRNA sequence of each type (3 total, targeting control, *Myc* and *Tnfrsf1a*) was more than ten-fold underrepresented in the raw sequence results for all experimental samples, so they were excluded from further analysis as outliers. These likely represented inefficient cloning from the oligonucleotide pools. Therefore, 37 gRNA sequences, 19 for *control*, nine for *Myc*, and nine for *Tnfrsf1a*, were included in the analysis for TALib1 (Table S1). Similarly, a total of 57 gRNAs were met the criteria for inclusion in the analysis for TALib2, including 20 *control*, nine *Myc*, ten *Tnfrsf1a*, nine *Slc7a11*, and nine *Tp53* gRNAs (Table S2).

### RNA extraction and quantitative RT-PCR

Total RNA was extracted from liver or tumor tissue using the RNeasy kit (Qiagen). Superscript II reverse transcriptase was used for generating complementary DNA (Life Technologies). Polymerase chain reaction (PCR) was performed using SYBRGreen qPCR Master Mix on a Mx3000 PCR cycler (Agilent Technologies). Each reaction was performed in triplicate and normalized relative to the ROX reference dye, and median cycle threshold values were used for subsequent analyses. TATA-box binding protein (Tbp) was used to normalize messenger RNA (mRNA) expression level. The following primers were used: *Myc* (forward: 5'-GCTGTTTGAAGGCTGGATTTC-3'; reverse: 5'-GATGAAATAGGGCTGTACGGAG-3') Tbp (forward: 5'-CCCCTTGACCCTTCACCAAT-3'; reverse: 5'-GAAGCTGCGGTACAATTCCAG-3').

### Statistics

ANOVA was performed for the gRNA prevalence values for TALib1, and *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mice, and showed a value of  $P = .00016$  for a difference between the *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> values, as a group, as compared to TALib1 input. We performed paired two-tailed T-tests to compare the difference in the corrected prevalence values for the gRNAs in TALib1 and TALib2 for groups of gRNAs: *control*, *Myc*, *Tnfrsf1a*, *Slc7a11*, and *Tp53*. *Myc* mRNA levels for the treatment groups were compared using a T-test. Chi-squared tests, followed by Bonferroni adjustment for multiple comparisons, were performed to compare linkage of tumors to gRNAs targeting *Myc*, *Tnfrsf1a*, *Slc7a11*, and *Tp53* with *control* gRNAs, using a cut-off value for tumor linkage of an individual gRNAs of prevalence greater than 0.05 or 0.001. Both cut-off values gave highly significant results showing enrichment of *Myc* and *Slc7a11* gRNAs and depletion of *Tnfrsf1a* and *Tp53* gRNAs compared to *control* gRNAs.

## RESULTS

### Derivation of a tissue-specific CRISPRa platform in mice

To achieve target gene activation, the CRISPRa system needs three components to form a complex at a genomic locus: a gRNA complementary to the locus, the dCas9 protein, and a transcriptional activator (TA) protein that binds to the 'SunTag' domain of dCas9 (Fig 1a).

We first derived mice containing a *Rosa26* gene replacement allele encoding a nuclease-deficient *dCas9* gene fused to the 'SunTag' domain(20), with an upstream floxed stop cassette (*loxP*-stop-*loxP*), and termed these gene replacement mice *Gt(ROSA)26Sor<sup>tm1(CAG-dCas9-SunTag)Khk</sup>*, or '*dCas9*<sup>+</sup>' for short (Fig. 1b). PCR genotyping

confirmed the presence of the *dCas9*<sup>+</sup> allele (Fig. 1c). Mice homozygous for the allele (*dCas9*<sup>+/+</sup>) were fertile and phenotypically indistinguishable from their control littermates.

Next, *dCas9*<sup>+</sup> mice were tested for the activity of the floxed stop cassette and for tissue-specific dCas9 protein expression. In order to remove the floxed stop cassette from the *dCas9* allele in hepatocytes, an adeno-associated virus serotype 8 expressing Cre with a hepatocyte-specific thyroxin-binding globulin (TBG) promoter (AAV-Cre)(29, 30) was injected intravenously into adult *dCas9*<sup>+</sup> mice. We found that *dCas9* protein was expressed in the liver of mice injected with AAV-Cre, but was not present in the heart, pancreas, or kidney; nor was it detectable in the liver of uninjected *dCas9*<sup>+</sup> mice (Fig. 1d). Immunohistochemical staining of liver sections confirmed the expression of predominantly nuclear dCas9 in hepatocytes (Fig. 1e). Thus, our *dCas9*<sup>+</sup> mouse exhibits tissue-specific and temporal control of expression in the liver, and in any other mouse cell type for which a Cre-driver is available.

### CRISPRa screen during liver repopulation

Having established that the dCas9 protein is expressed specifically within hepatocytes upon injection of AAV-Cre, we next sought to facilitate CRISPRa-linked liver repopulation assays by crossing *dCas9*<sup>+</sup> mice to *Fah*<sup>-/-</sup> mice to produce *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mice. To prevent liver injury, these mice were provided with continuous nitisinone therapy until *Fah*-encoding plasmids were injected at 8-12 weeks of age.

To introduce the remaining components of the CRISPRa system, we derived *Sleeping Beauty* transposon-containing plasmids with gRNA, TA, and *Fah* expression cassettes. After pooling, these plasmids were hydrodynamically injected into *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mice (See schema Fig 1a). The TA protein is a fusion of the VP64 domain, a transcriptional activator derived from Herpes Simplex Virus, with an antibody fragment that avidly binds to the ‘SunTag’ epitope array, such that up to ten copies are loaded onto each dCas9 protein(20, 21). The design of the gRNA expression construct facilitates high-throughput sequencing for precise quantification of the read counts for each of the gRNA sequences within libraries(13, 20). We used previously validated computational prediction tools to design gRNAs targeted to the transcription start sites of selected target genes(31). Our initial plasmid library, which we term ‘TALib1’, included ten gRNAs targeting *Myc*, (Fig. 2a, Table S1), ten targeting *Tnfrsf1a* (Fig. S1, Table S1), and twenty non-targeting control gRNAs (Table S1)(31). Thus, TALib1 is a pool of 40 plasmids that express *Fah*, the TA construct, and a unique gRNA.

To test our dCas9 activation system, *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mice were injected with AAV-Cre to activate expression of the dCas9 protein in hepatocytes. After 72 hours, TALib1 was hydrodynamically injected into either *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> or *Fah*<sup>-/-</sup> mice, and nitisinone was withdrawn to initiate liver injury and repopulation. Four weeks post-injection, mice were euthanized and examined for CRISPRa activity during liver repopulation.

Western blot analysis detected dCas9 and TA protein in livers tissue, confirming that the CRISPRa components were present in the liver one month after injection of AAV-Cre and the plasmid library (Fig 2b). Next, we determined whether the target gene *Myc* was activated by the CRISPRa system, and whether the localization of MYC protein coincided with

repopulation nodules. FAH and MYC protein expression and localization were examined by immunohistochemistry of serial paraffin sections of liver tissue (Fig 2c). Nodules of FAH-positive hepatocytes were found in the livers, consistent with the clonal expansion of stably transduced hepatocytes seen with *Sleeping Beauty*-mediated gene therapy in *Fah*<sup>-/-</sup> mice (Fig 2c, middle column) (27, 28, 32). MYC protein was strongly positive within the nuclei of hepatocytes in all three *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice injected with TALib1, but not in *Fah*<sup>-/-</sup> control mice. MYC expression corresponded with FAH-positive nodules, indicating a robust and specific activation of expression (Fig. 2c, right column). Importantly, not all FAH-positive nodules displayed upregulation of MYC, consistent with a gRNA-specific effect, as only 10 out of the 40 gRNAs in the TALib1 pool were targeted to the *Myc* locus (Fig. 2c, yellow arrows). Thus, our CRISPRa system successfully activates target gene expression in repopulating hepatocytes *in vivo*.

### Promotion of liver repopulation by CRISPRa induction of *Myc*

Having confirmed that CRISPRa can induce MYC protein expression, we next interrogated whether there was a growth advantage imparted on hepatocytes with activated *Myc* by determining changes in gRNA sequence prevalence over the course of liver repopulation. An increase in gRNA prevalence represents a functional promotion of hepatocyte clonal expansion. To this end, we recovered DNA from the repopulated liver and performed high throughput sequencing of the gRNA sequences to assess the prevalence of each gRNA as compared to the input plasmid pool for TALib1. We found significant enrichment for gRNAs targeting *Myc* activation as compared to control gRNAs (Fig. 3a). There was no significant change in prevalence for gRNAs targeting *Tnfrsf1a*. We found that *Myc* gRNAs were enriched up to 10-fold compared to a *Fah*<sup>-/-</sup> control mouse or the input plasmid pool (Fig. 3b).

To confirm our findings and to assay additional target genes, we engineered a second plasmid pool that included gRNAs to activate *Slc7a11* and *Tp53* (10 gRNAs each), (TALib2; Table S2). *Slc7a11* was selected for further study based on our results showing that this gene, which functions in glutathione metabolism, is massively upregulated in repopulating hepatocytes in *Fah*<sup>-/-</sup> mice (Wang, A., *et al.*, manuscript submitted elsewhere). *Tp53* was selected as a potential negative regulator of carcinogenesis(33, 34). As with the above experiment, *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice were first injected with AAV-Cre to activate dCas9 expression, followed by hydrodynamic injection of TALib2 and withdrawal of nitisinone. The mice were allowed to repopulate for one to two months. Examination of bulk liver tissue for gRNA prevalence changes at two months using high throughput sequencing again showed significant enrichment of multiple gRNAs targeting *Myc* (Fig. 3c). We conclude that our *in vivo* dCas9 activation system activated the *Myc* locus in hepatocytes and accelerated repopulation by these cells, as measured by enrichment in the corresponding gRNA sequences compared to the control gRNAs.

### CRISPRa induction of *Myc* expression leads to liver carcinogenesis

We observed that mice treated with either TALib1 and TALib2 developed one or more liver tumors following repopulation (Fig 4a). In one of the three of the *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice treated with TALib1 we detected a tumor of compact, dysplastic-appearing cells that were

strongly positive for FAH and MYC (Fig 4b, top row), indicating that CRISPRa activation of the endogenous *Myc* locus was linked to the initiation of hepatocyte carcinogenesis. The tumor tissue was highly proliferative, as indicated by a large number of Ki-67-positive cells, and stained positive for markers of aggressive hepatocellular carcinoma including alpha-fetoprotein (AFP), osteopontin (OPN), and yes-associate protein 1 (YAP1) (Fig 4b, bottom row). With TALib2, two out of three mice euthanized at one month and three out of three mice euthanized at two months were found to have developed hepatic tumors. No tumors were observed in control *Fah*<sup>-/-</sup> repopulated with the library, as reported previously for *Fah*<sup>-/-</sup> repopulated with FAH plasmids(27, 32).

To quantify the degree of *Myc* activation, we performed quantitative RT-PCR on liver tissue from a C57B6 wildtype mouse, from control *Fah*<sup>-/-</sup> mice repopulated with *Fah*-expressing plasmids, from *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice repopulated with TALib1 or TALib2, and from liver tumors of *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice treated with TALib2 (Fig. 4c). The results showed an average 23-fold induction of *Myc* mRNA in bulk liver tissue from mice repopulated with the CRISPRa library ( $p < 0.05$ ), and more than 300-fold induction of *Myc* mRNA in tumors ( $p < 0.05$ ).

Next, we extracted genomic DNA from liver tumors and performed high throughput sequencing of the gRNA sequences to determine which of them were linked to each tumor. We found that all tumors contained linked *Myc* gRNAs, indicating that tumors could not form without CRISPRa activation of the *Myc* locus (Figure 4e). The tumors averaged 4.6 unique gRNA sequences (range 2-7), and the proportions of each of the gRNAs linked to the tumors were consistent with clonality (Fig. 4e). For example, the CRISPRa tumor pie chart shown in Fig. 4d illustrates a tumor that contained 7 unique gRNA insertion events, including two separate events for the same *Myc* gRNA, which resulted in twice the number of sequencing reads for this gRNA compared to the other gRNA sequences (Fig. 4d, right panel). Chi-squared testing indicated that, compared to the control gRNA sequences, there was a highly significant enrichment in *Myc* and *Slc7a11* gRNA sequences in the tumors, while *Tp53* and *Tnfrsf1a* gRNAs were significantly depleted ( $p < 0.001$  for all comparisons).

To conclude, our results demonstrate that CRISPRa can initiate hepatocellular carcinomas in our mouse liver injury model by activating high levels of endogenous *Myc* expression. Importantly, because the initial hydrodynamic injection results in multiple *Sleeping Beauty*-mediated integration events in each hepatocyte clone (4.6 on average), or system allows for the assessment of combinatorial effects in each clonal tumor. Thus, we discovered that *Myc*-driven hepatic tumorigenesis is strongly inhibited by simultaneous activation of *Tp53* and *Tnfrsf1a*, as tumors were significantly depleted for gRNA sequences that target these genes ( $p < 0.001$  for each comparison). This confirms the capacity of our CRISPRa system for performing genetic screening and epistasis assays *in vivo*.

## DISCUSSION

The results presented here demonstrate functional, parallel and combinatorial genetic screening *in vivo* using a CRISPRa platform in mice. Multiple independent gRNAs activated *Myc* expression to give a robust growth phenotype in a liver repopulation model. The

phenotype is analogous to what we previously observed with Myc cDNA overexpression(28). Indeed, it was previously shown that liver-directed Myc cDNA overexpression in transgenic animals resulted in hepatocellular carcinoma by 12-15 months(35). In our mice, carcinoma development occurred much more rapidly, likely due to the liver injury and repopulation that occurs in *Fah*<sup>-/-</sup> mice, which requires multiple rounds of hepatocyte cell division, and due to the high level of activation of Myc expression achieved in our system.

Myc expression was the result of gRNAs that directed dCas9 and VP64 to the transcriptional start site of *Myc*. Myc is induced in many, if not most, human cancers. In addition, it is upregulated during liver regeneration after 2/3 partial hepatectomy(36, 37) and liver repopulation in *Fah*<sup>-/-</sup> mice (Wang A. et al., manuscript submitted elsewhere). Myc expression levels are controlled by transcriptional activators such as by BRD4(38, 39), by phosphorylation and subsequent proteosomal degradation(40), and by interaction with the protein AURKA(41). Our results demonstrate that the level of transcription from the endogenous *Myc* locus can be modulated using the Cas9 activation system *in vivo*, and that enhanced expression leads to a dramatic physiological impact on repopulation and carcinogenesis. Thus, Myc expression levels – normally tightly controlled in repopulating hepatocytes – can be significantly altered by the CRISPRa platform.

In contrast to the effect of Myc on promoting repopulation, we did not observe significant changes in gRNAs directed to *Tnfrsf1a*, *Slc7a11*, or *Tp53* loci during liver repopulation. However, both *Myc* and *Slc7a11* gRNAs were significantly enriched in tumors, whereas *Tnfrsf1a* and *Tp53* were strikingly depleted. The results suggest that activation of these loci had a strong epistatic impact on Myc-dependant tumorigenesis. Previously, we had found that TNFR1 cDNA overexpression with a strong chicken beta-actin promoter blocked liver repopulation in *Fah*<sup>-/-</sup> mice, which was not observed with the CRISPRa platform(28). It is possible that the level of expression of TNFR1 with the cDNA was greater than with dCas9 system used here. Another possibility for the discrepancy is that the Myc and TNFR1 upregulation overlapped within most hepatocytes in the present study, as hepatocytes took up more than four gRNAs on average, such that the positive effect on growth directed by Myc overrode or masked the negative effects of TNFR1 upregulation. Importantly, gRNAs activating *Tnfrsf1a* blocked Myc-driven tumorigenesis, suggesting that targeting of this cell surface receptor could be used as a therapy to prevent or treat hepatocellular carcinomas in humans.

Future experiments with larger, more complex libraries will be necessary to fully delineate which genes can drive liver repopulation in the setting of injury and tumorigenesis. Thus, we are currently developing screens in which all genes in specific pathways can be interrogated simultaneously, by targeting all relevant gene promoters with guide RNAs. The size of the pooled libraries that can be screened depends on the efficiency of stable integration of transposons expressing single gRNAs. Hydrodynamic injection of *Sleeping Beauty* transposons into *Fah*<sup>-/-</sup> mice leads to stable integration of the transposon in up to 1 in 100 hepatocytes(32). Given that an adult mouse has roughly 10<sup>8</sup> hepatocytes, up to 10<sup>6</sup> unique clonal events can be analyzed in a single animal. Therefore, we estimate that a pooled plasmid library with a complexity of 10,000 unique gRNAs targeting 2,000 genes with five

gRNAs each could achieve an average of 100-fold representation in a single mouse, enabling high throughput pathway and gene set analyses for biological processes in the liver.

Our data on the epistatic relationship between Myc and Tp53, demonstrating that Myc-driven tumors that also express high levels of p53 are at a selective disadvantage, establishes that our system can be used to analyze epistatic relationships between the genes that are part of the screen. In the future, larger libraries of gRNA could be employed to determine which genes promote and which inhibit Myc-driven tumors. These data will be complementary to transcriptome profiling of liver tumors, which by definition captures all changes in gene expression profile in HCC, whether they are primary drivers or secondary consequences of the neoplastic process.

We demonstrate that our dCas9 allele has exquisite tissue specificity that is dependent on the removal of a stop cassette by Cre. We envision that our system will be useful for performing additional tissue-specific genetic screens, as dCas9 can be activated in any tissue of interest. We showed hepatocyte specificity with an AAV-Cre virus, but viruses with other cell type-specific tropisms could be used to target other tissues, such as the hematopoietic system, heart or brain. Another method for tissue-specific activation is with transgenic mice harboring tissue-specific promoters driving Cre expression, such as Villin-Cre for intestine-specific expression(42). The TA and gRNA sequences could be provided to target tissues with lentiviral or electroporation techniques, or in lipid formulations. The *dCas9<sup>+</sup>* mouse strain could also be used for performing screens on cultured primary cells such as mouse embryonic fibroblasts (MEFs), stem cells, blood cells, hepatocytes, etc.

The types of screens that will be able to be performed using mouse tissues with our system are limitless. Growth and tumorigenesis assays *in vivo* are especially promising. These data could be integrated with expression analysis to identify true ‘drivers’ of tumorigenesis – genes that are overexpressed and contribute to tumorigenesis – as opposed to ‘passenger’ genes – genes that become overexpressed in tumors but have no effect when the expression is targeted. The ‘driver’ genes will make for excellent targets for cancer prevention or treatment.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Drs. Jonathan Schug and Shilpa Rao, for their services and advice with DNA sequencing analysis. We thank Noam Erez and Elizabeth Howell for technical support, and Drs. Adam Zahm and Yitzhak Reizel for editing and helpful discussions.

**Financial Support:** This work was supported by the National Institutes of Health (R01-DK102667, to K.H.K. and K08-DK106478 and T32-DK007066 to K.J.W.), the American Gastroenterological Association Elsevier Pilot Award (K.J.W.), the Functional Genomics Core of the Penn Diabetes Research Center (P30-DK19525) and the Molecular Pathology and Imaging Core of the Penn Center for Molecular Studies in Digestive and Liver Disease (P30-DK50306, pilot grant to K.J.W.).

## List of Abbreviations

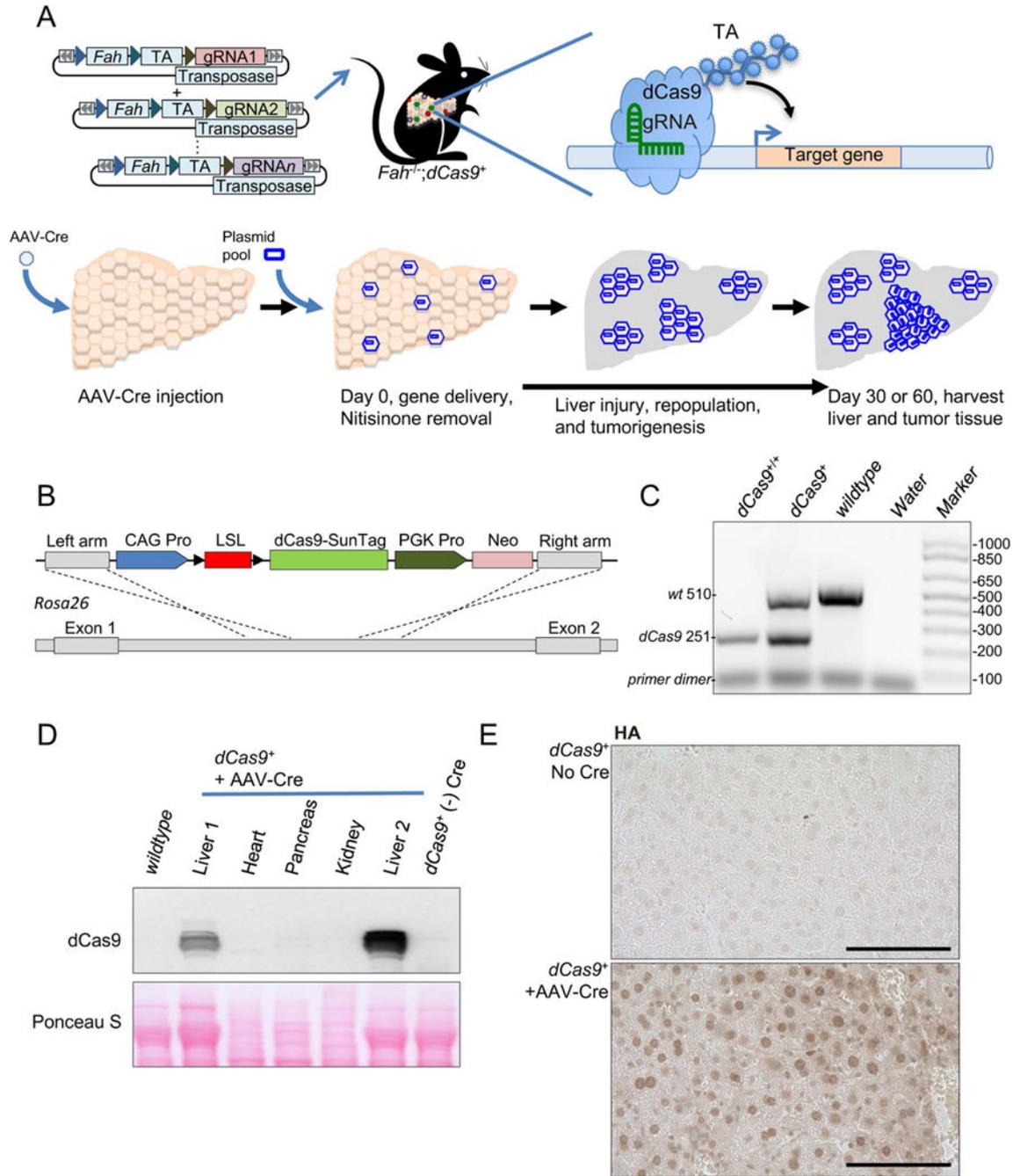
<b>CRISPRa</b>	CRISPR/Cas9 activation
<b>gRNA</b>	guide RNA
<b>dCas9</b>	catalytically dead Cas9
<b>HA</b>	Hemagglutinin Tag
<b>TA</b>	transcriptional activator
<b>Tbp</b>	TATA-box binding protein
<b>mRNA</b>	messenger RNA
<b>TBG</b>	thyroxin-binding globulin
<b>AAV-Cre</b>	serotype 8 adeno-associated virus with thyroxin-binding globulin promoter driving Cre expression
<b>AFP</b>	alpha-fetoprotein
<b>OPN</b>	osteopontin
<b>YAP1</b>	yes-associate protein 1

## References

1. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, et al. Landscape of transcription in human cells. *Nature*. 2012; 489:101–108. [PubMed: 22955620]
2. Consortium F, the RP, Clst. Forrest AR, Kawaji H, Rehli M, Baillie JK, et al. A promoter-level mammalian expression atlas. *Nature*. 2014; 507:462–470. [PubMed: 24670764]
3. Zhang Y, An L, Yue F, Hardison RC. Jointly characterizing epigenetic dynamics across multiple human cell types. *Nucleic Acids Res*. 2016
4. Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res*. 2013; 23:555–567. [PubMed: 23325432]
5. Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*. 2015; 160:1246–1260. [PubMed: 25748654]
6. Weber J, Ollinger R, Friedrich M, Ehmer U, Barenboim M, Steiger K, Heid I, et al. CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proc Natl Acad Sci U S A*. 2015; 112:13982–13987. [PubMed: 26508638]
7. Xue W, Chen S, Yin H, Tammela T, Papagiannakopoulos T, Joshi NS, Cai W, et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature*. 2014; 514:380–384. [PubMed: 25119044]
8. Liu Y, Thor A, Shtivelman E, Cao Y, Tu G, Heath TD, Debs RJ. Systemic gene delivery expands the repertoire of effective antiangiogenic agents. *J Biol Chem*. 1999; 274:13338–13344. [PubMed: 10224095]
9. Huang LC, Clarkin KC, Wahl GM. Sensitivity and selectivity of the DNA damage sensor responsible for activating p53-dependent G1 arrest. *Proc Natl Acad Sci U S A*. 1996; 93:4827–4832. [PubMed: 8643488]
10. Jackson SP. Sensing and repairing DNA double-strand breaks. *Carcinogenesis*. 2002; 23:687–696. [PubMed: 12016139]

11. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* 2013; 23:1163–1171. [PubMed: 23979020]
12. Farzadfard F, Perli SD, Lu TK. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth Biol.* 2013; 2:604–613. [PubMed: 23977949]
13. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 2013; 154:442–451. [PubMed: 23849981]
14. Hu J, Lei Y, Wong WK, Liu S, Lee KC, He X, You W, et al. Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. *Nucleic Acids Res.* 2014; 42:4375–4390. [PubMed: 24500196]
15. Kearns NA, Genga RM, Enuameh MS, Garber M, Wolfe SA, Maehr R. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development.* 2014; 141:219–223. [PubMed: 24346702]
16. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods.* 2013; 10:977–979. [PubMed: 23892898]
17. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* 2013; 31:833–838. [PubMed: 23907171]
18. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods.* 2013; 10:973–976. [PubMed: 23892895]
19. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, PRI E, Lin S, et al. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods.* 2015; 12:326–328. [PubMed: 25730490]
20. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell.* 2014; 159:647–661. [PubMed: 25307932]
21. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell.* 2014; 159:635–646. [PubMed: 25307933]
22. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature.* 2015; 517:583–588. [PubMed: 25494202]
23. Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, Haque SJ, et al. Comparison of Cas9 activators in multiple species. *Nat Methods.* 2016
24. Braun CJ, Bruno PM, Horlbeck MA, Gilbert LA, Weissman JS, Hemann MT. Versatile in vivo regulation of tumor phenotypes by dCas9-mediated transcriptional perturbation. *Proc Natl Acad Sci U S A.* 2016
25. Grompe M, Lindstedt S, al-Dhalimy M, Kennaway NG, Papaconstantinou J, Torres-Ramos CA, Ou CN, et al. Pharmacological correction of neonatal lethal hepatic dysfunction in a murine model of hereditary tyrosinaemia type I. *Nat Genet.* 1995; 10:453–460. [PubMed: 7545495]
26. Overturf K, Al-Dhalimy M, Tanguay R, Brantly M, Ou CN, Finegold M, Grompe M. Hepatocytes corrected by gene therapy are selected in vivo in a murine model of hereditary tyrosinaemia type I. *Nat Genet.* 1996; 12:266–273. [PubMed: 8589717]
27. Wangensteen KJ, Wilber A, Keng VW, He Z, Matise I, Wangensteen L, Carson CM, et al. A facile method for somatic, lifelong manipulation of multiple genes in the mouse liver. *Hepatology.* 2008; 47:1714–1724. [PubMed: 18435462]
28. Wangensteen KJ, Zhang S, Greenbaum LE, Kaestner KH. A genetic screen reveals Foxa3 and TNFR1 as key regulators of liver repopulation. *Genes Dev.* 2015; 29:904–909. [PubMed: 25934503]
29. Yanger K, Zong Y, Maggs LR, Shapira SN, Maddipati R, Aiello NM, Thung SN, et al. Robust cellular reprogramming occurs spontaneously during liver regeneration. *Genes Dev.* 2013; 27:719–724. [PubMed: 23520387]

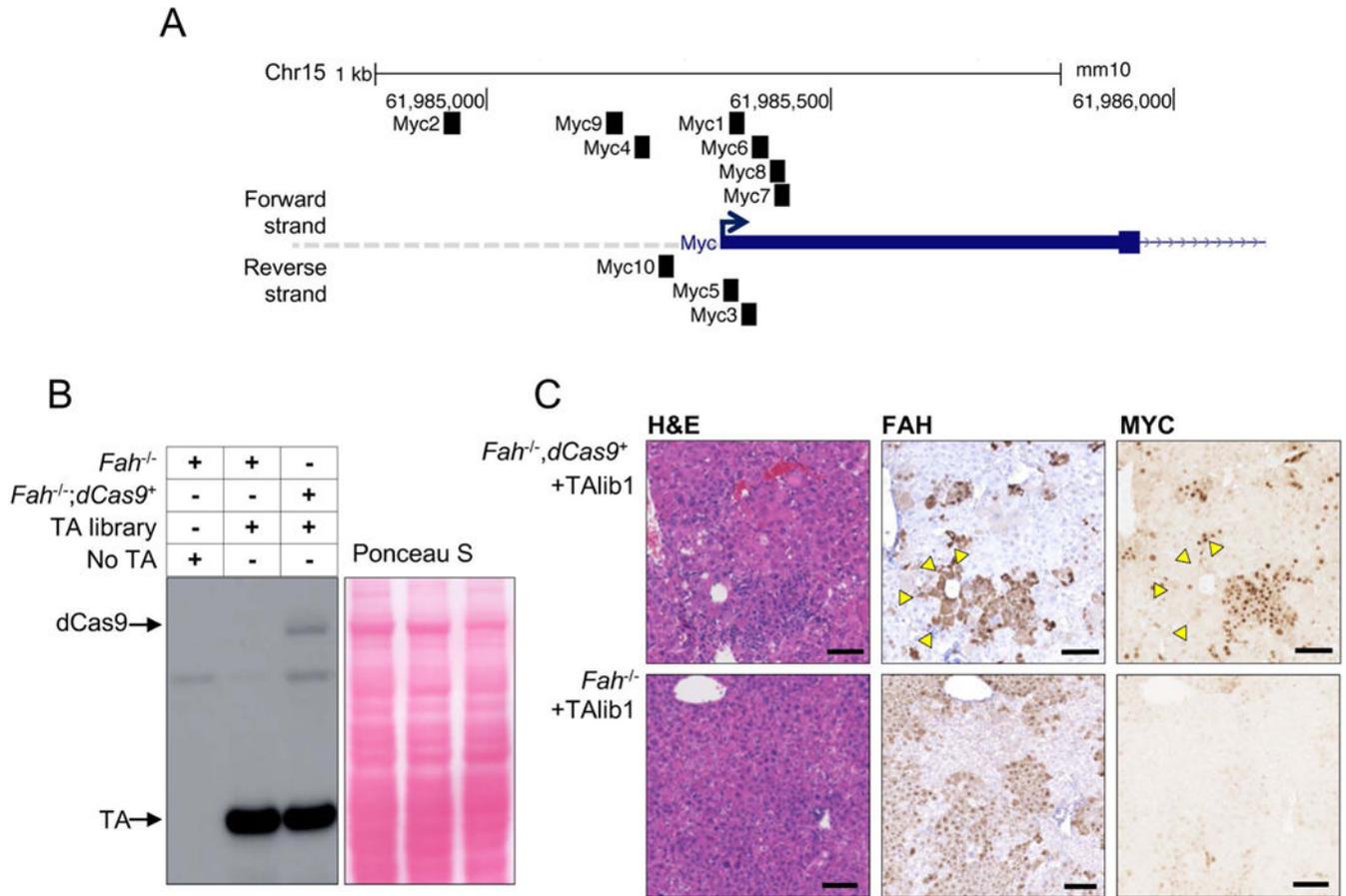
30. Shin S, Wangensteen KJ, Teta-Bissett M, Wang YJ, Mosleh-Shirazi E, Buza EL, Greenbaum LE, et al. Genetic Lineage Tracing Analysis of the Cell of Origin of Hepatotoxin-Induced Liver Tumors in Mice. *Hepatology*. 2016
31. Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, Fields AP, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife*. 2016; 5
32. Montini E, Held PK, Noll M, Morcinek N, Al-Dhalimy M, Finegold M, Yant SR, et al. In vivo correction of murine tyrosinemia type I by DNA-mediated transposition. *Mol Ther*. 2002; 6:759–769. [PubMed: 12498772]
33. Tornesello ML, Buonaguro L, Tatangelo F, Botti G, Izzo F, Buonaguro FM. Mutations in TP53, CTNNB1 and PIK3CA genes in hepatocellular carcinoma associated with hepatitis B and hepatitis C virus infections. *Genomics*. 2013; 102:74–83. [PubMed: 23583669]
34. Schulze K, Imbeaud S, Letouze E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet*. 2015; 47:505–511. [PubMed: 25822088]
35. Santoni-Rugiu E, Nagy P, Jensen MR, Factor VM, Thorgeirsson SS. Evolution of neoplastic development in the liver of transgenic mice co-expressing c-myc and transforming growth factor- $\alpha$ . *Am J Pathol*. 1996; 149:407–428. [PubMed: 8701981]
36. White P, Brestelli JE, Kaestner KH, Greenbaum LE. Identification of transcriptional networks during liver regeneration. *J Biol Chem*. 2005; 280:3715–3722. [PubMed: 15546871]
37. Taub R. Liver regeneration 4: transcriptional control of liver regeneration. *FASEB J*. 1996; 10:413–427. [PubMed: 8647340]
38. Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison EA, Magoon D, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*. 2011; 478:524–528. [PubMed: 21814200]
39. Delmore JE, Issa GC, Lemieux ME, Rahl PB, Shi J, Jacobs HM, Kastiris E, et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell*. 2011; 146:904–917. [PubMed: 21889194]
40. Amati B. Myc degradation: dancing with ubiquitin ligases. *Proc Natl Acad Sci U S A*. 2004; 101:8843–8844. [PubMed: 15187232]
41. Dauch D, Rudalska R, Cossa G, Nault JC, Kang TW, Wuestefeld T, Hohmeyer A, et al. A MYC-aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. *Nat Med*. 2016
42. Madison BB, Dunbar L, Qiao XT, Braunstein K, Braunstein E, Gumucio DL. Cis elements of the villin gene control expression in restricted domains of the vertical (crypt) and horizontal (duodenum, cecum) axes of the intestine. *J Biol Chem*. 2002; 277:33275–33283. [PubMed: 12065599]



**Figure 1. Derivation of a tissue-specific CRISPRa mouse**

(a) Schema of our approach to activate gene expression in repopulating hepatocytes by injecting *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice with pools of *Sleeping Beauty* transposon plasmids expressing *Fah*, Transcriptional Activator (TA), and guide RNAs (gRNA). The CRISPRa complex is loaded onto the promoter region of target genes (cartoon, right). Below is a schema of the experimental design to first inject AAV-Cre to activate *dCas9* specifically within hepatocytes, followed by hydrodynamic injection of plasmid pools and removal of nitisinone to induce liver injury and repopulation. (b) Schematic of the gene replacement allele

(*dCas9<sup>+</sup>*), which was targeted to the Rosa26 locus. (c) Genotyping for the *dCas9<sup>+</sup>* allele, showing examples of homozygosity and heterozygosity for the replacement allele (*dCas9<sup>+/+</sup>* and *dCas9<sup>+</sup>*, respectively) and the wildtype allele. (d) Western blot using a probe for the HA-epitope of dCas9, demonstrating expression of the protein in the liver of *Fah<sup>-/-</sup>;dCas9<sup>+</sup>* mice injected with AAV-Cre, and absence of expression in heart, pancreas, and kidney tissues, or in control animals (wildtype mice and *Fah<sup>-/-</sup>;dCas9<sup>+</sup>* mice uninjected with AAV-Cre). Ponceau S staining indicates even loading (e) Immunohistochemistry staining with HA antibody of dCas9 protein, showing tissue-specific expression in hepatocytes, and nuclear localization. Scale bar = 100  $\mu$ m. Abbreviations: TA: transcriptional activator, AAV-Cre: adeno-associated virus serotype 8, with TBG promoter driving Cre, HA: Hemagglutinin.

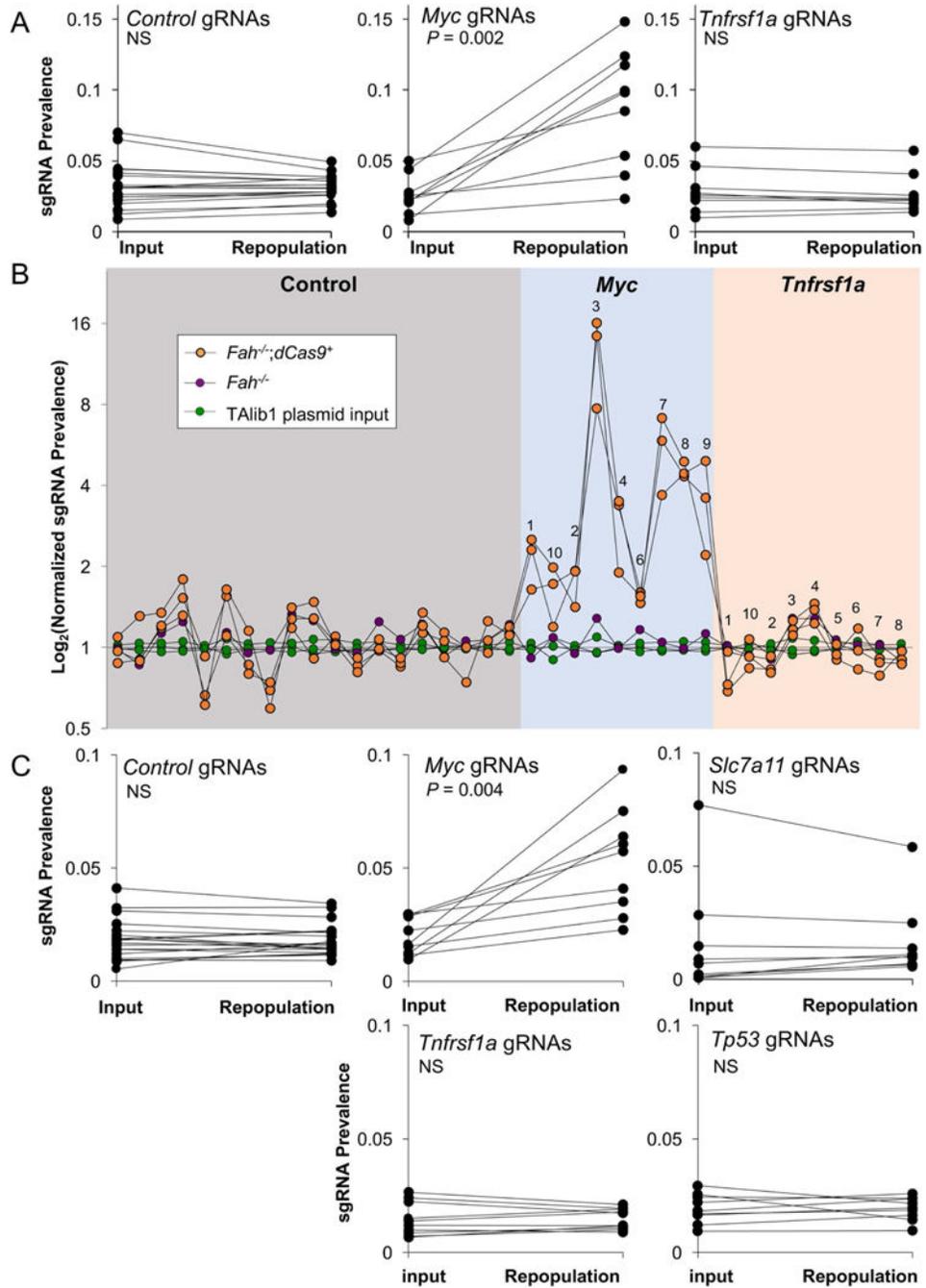


**Figure 2. CRISPRa components co-localize in repopulating hepatocytes and activate the expression of the target *Myc***

(a) Illustration of the gRNA binding sites targeted to the promoter region of the *Myc* locus.

(b) Western blot with an anti-HA antibody demonstrating TA expression and dCas9 expression in the liver of *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice, one month after injection with AAV-Cre and TALib. Ponceau S staining demonstrates even loading.

(c) Serial paraffin sections from *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> or *Fah*<sup>-/-</sup> mice injected with TALib1 were stained with Hematoxylin and Eosin (H&E, left panel), Immunohistochemistry for FAH (middle panel), or MYC (right panel). The staining shows FAH-positive repopulation nodules for both *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice and *Fah*<sup>-/-</sup> mice, but MYC is induced only in *Fah*<sup>-/-</sup>; *dCas9*<sup>+</sup> mice. MYC expression coincides with FAH expression. The yellow arrows mark a repopulation nodule that is not MYC positive, which is expected as the gRNA pool contained a mix of *Myc*, *Tnfrsf1a*, and *control* gRNAs. Scale bar = 100  $\mu$ m. Abbreviations: HA: Hemagglutinin.



**Figure 3. CRISPRa enhances liver repopulation: enrichment of gRNAs targeting *Myc* during liver repopulation**  
 (a) The average adjusted prevalence for *Control* (left), *Myc* (middle) and *Tnfrsf1a* (right) gRNA sequences are graphed. Each average gRNA prevalence value for TALib1 (the input, n = 3 independent replicates) is shown paired with the value in *Fah*<sup>-/-</sup>;dCas9<sup>+</sup> mice (n = 3 biological replicates) after 4 weeks of repopulation. P-values are shown for the differences in the prevalence. (b) Graph of gRNA prevalence normalized to the average prevalence of the control gRNA sequences in TALib1. The green dots indicate the normalized prevalence values for each of the gRNA sequences in TALib1 (three independent replicates, each

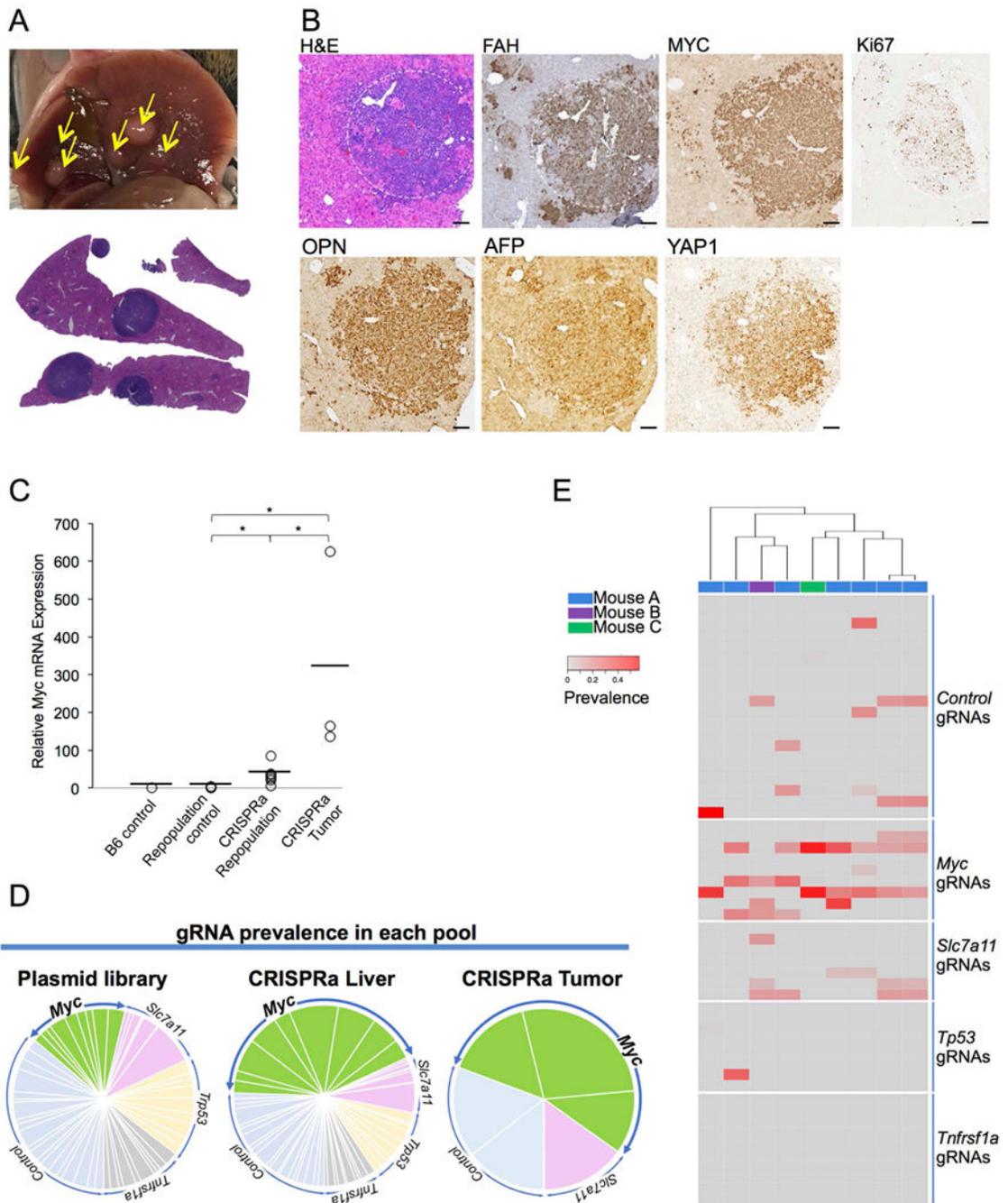
connected with a line). The orange dots show the normalized gRNA prevalence values for three *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mice, one month after liver repopulation (values from each mouse are connected with a line). The purple dots indicate the gRNA prevalence values for a *Fah*<sup>-/-</sup> control mouse. The number above each column indicates the specific gRNA for *Myc* and *Tnfrsf1a*. (c) The average adjusted prevalence values for TALib2 gRNAs. Each average gRNA prevalence is shown for *Control*, *Myc*, *Slc7a11*, *Tnfrsf1a*, and *Tp53* gRNA sequences in the TALib2 plasmid pool (the input, n = 3 independent replicates), paired with the value in *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mice (n = 3 biological replicates) after eight weeks of repopulation. P-values are shown for the differences in the prevalence.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 4. CRISPRa drives carcinogenesis in mice by upregulating the *Myc* locus**  
 (a) Gross image of a representative *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mouse, injected with TALib2 and euthanized after one month of liver repopulation. Below is a corresponding H&E image from this mouse. (b) Serial sections from a *Fah*<sup>-/-</sup>;*dCas9*<sup>+</sup> mouse, one month after injection with TALib1, showing a tumor that is larger in size than the surrounding repopulating hepatocyte nodules, with poorly defined borders (upper left panel). The cells are deeply basophilic and have a high nuclear-to-cytoplasmic ratio. The tumor is FAH-positive and MYC-positive, demonstrating that it is linked to the plasmids that were injected into the

animal. There is a high proliferative index, as indicated by Ki67 staining. Below, immunostaining demonstrates co-localization of OPN, AFP, and YAP1 in the tumor, markers of aggressive hepatocellular carcinoma. Scale bar = 100  $\mu$ m. (c) Quantitative RT-PCR of *Myc* mRNA, normalized to *Tbp* mRNA, demonstrating significant upregulation of *Myc* in CRISPRa-repopulated liver tissue and in CRISPRa-tumors compared to control mice repopulated with *Fah*-expressing plasmids. \* indicates  $P < .05$ . (d) Pie charts showing the numerical proportion of all of the components of TALib2 in the plasmid library pool (left), in bulk, grossly tumor-free liver tissue after eight weeks of repopulation (middle), and in a representative CRISPRa-tumor (right). The results demonstrate enrichment for *Myc* gRNAs over the course of repopulation and tumorigenesis. The CRISPRa tumors are noted to have a quantal distribution of the gRNA prevalence, indicating that the tumor originates from a single clone. In this example, seven gRNA insertion events are predicted to have occurred: two for one of the *Myc* gRNAs (largest sector), two additional unique *Myc* gRNAs, two *Control* gRNAs, and one *Slc7a11* gRNA. (e) Heatmap of the prevalence of each gRNA (rows) in each tumor (columns) collected from mice repopulated with the CRISPRa system. *Myc* gRNAs were found to be linked to all of the tumors. Also, as compared to *control* gRNAs, *Myc* and *Slc7a11* gRNAs are significantly enriched in tumors, whereas *Tp53* and *Tnfrsf1a* gRNAs are significantly depleted ( $p < 0.001$  for each comparison by chi-squared test). Note that the right-most two columns show that a single tumor has an identical gRNA linkage pattern when two different parts of the tumor were sampled, indicating homogeneity in the linkage of gRNAs throughout the tumor.