A genome-wide screen in the mouse liver reveals sex-specific and cell non-autonomous regulation of cell fitness

Heather R. Keys† and Kristin A. Knouse†*

†Whitehead Institute for Biomedical Research, Cambridge, MA

*Corresponding author. Email: knouse@wi.mit.edu

†These authors contributed equally to this work
ABSTRACT

Our ability to understand and modulate mammalian physiology and disease requires knowing how all genes contribute to any given phenotype in the organism. Genome-wide screening using CRISPR-Cas9 has emerged as a powerful method for the genetic dissection of cellular processes\textsuperscript{1,2}, but the need to stably deliver single guide RNAs to millions of cells has restricted its implementation to \textit{ex vivo} systems. These \textit{ex vivo} systems cannot reproduce all of the cellular phenotypes observed \textit{in vivo} nor can they recapitulate all of the factors that influence these phenotypes. There thus remains a pressing need for high-throughput functional genomics in a living organism. Here, we establish accessible genome-wide screening in the mouse liver and use this approach to uncover the complete regulation of cellular fitness in a living organism. We discover novel sex-specific and cell non-autonomous regulation of cell growth and viability. In particular, we find that the class I major histocompatibility complex is essential for preventing immune-mediated clearance of hepatocytes. Our approach provides the first comprehensive picture of cell fitness in a living organism and highlights the importance of investigating cellular phenomena in their native context. Our screening method is robust, scalable, and easily adapted to examine diverse cellular processes using any CRISPR application. We have hereby established a foundation for high-throughput functional genomics in a living mammal, enabling unprecedented insight into mammalian physiology and disease.

INTRODUCTION
A complete understanding of the genetic determinants underlying cell fitness and function in mammalian tissues is limited by the availability of tools for high-throughput genetic dissection in living organisms. Genome-wide screening using CRISPR-Cas9 has emerged as a powerful approach for the comprehensive investigation of cellular phenotypes\(^1\), but it requires stable delivery of a single guide RNA (sgRNA) library to tens of millions of cells at a low multiplicity of infection. Transposons and adeno-associated virus (AAV) can deliver sgRNAs \textit{in vivo} with reasonable efficiency\(^3\)^,\(^4\), but transposons often undergo multiple integrations per genome and AAV vectors do not integrate at all, limiting the utility of either of these approaches for genome-wide screening. Lentivirus is a preferable sgRNA delivery method as it allows for stable, single-copy integration in target cells. Unfortunately, genome-scale lentiviral delivery has only been achievable in cells \textit{ex vivo}, thereby restricting genome-wide screening to cell culture systems or cellular transplantation models\(^5\),\(^6\). These \textit{ex vivo} systems cannot always reproduce the cellular phenotypes desired for study and, even when they can, they cannot recapitulate the entirety of extracellular factors that influence these phenotypes \textit{in vivo}. These limitations emphasize the need to bring high-throughput functional genomics into the organism. To satisfy this demand and achieve a comprehensive understanding of cellular fitness in its native context, we sought to enable effective genome-scale sgRNA delivery \textit{in vivo} and establish accessible genome-wide CRISPR screening in a living mammal.

\section*{RESULTS}

\subsection*{Genome-scale sgRNA delivery to the liver}
The mouse liver, comprised of tens of millions of hepatocytes, is a promising tissue for genome-wide screening because it provides cell numbers compatible with genome-scale screening in a single mouse. Moreover, given the liver’s diverse metabolic functions and impressive regenerative capacity, hepatocytes exhibit a broad range of phenotypes—from universal cellular processes to hepatocyte-specific phenomena—that are ripe for genetic dissection. Previous attempts at delivering lentivirus to liver have suffered from poor transduction efficiency and immune-mediated clearance of transduced hepatocytes. However, we hypothesized that intravenously injecting highly concentrated lentivirus into neonatal mice might avoid these pitfalls and achieve efficient, stable transduction. To test this, we generated lentiviruses encoding a non-targeting sgRNA alongside mCherry or mTurq2 reporters and injected varying doses of an equal mixture of these two lentiviruses into postnatal day (PD) one mice (Fig. 1a). We observed a dose-dependent increase in the percentage of transduced hepatocytes, with a dose of $5 \times 10^7$ transduction units (TU) transducing over 75% of hepatocytes (Fig. 1b,c). Using histologic measurements and the transduction frequency of both fluorescent reporters, we estimated that a dose of $5 \times 10^7$ TU transduces approximately 10 million hepatocytes per PD1 liver with an average of just two integration events per cell (Fig. 1c and Extended Data Fig. 1a). This transduction efficiency would afford >200-fold coverage of a 100,000 feature sgRNA library in a single mouse. Importantly, these transduced hepatocytes were distributed uniformly throughout the liver lobule and persisted into adulthood (Fig. 1b and Extended Data Fig. 1b,c). Our lentiviral approach therefore establishes an sgRNA delivery method that is wholly compatible with genome-scale screening in the mouse liver.
We next asked whether we could use this sgRNA delivery approach as the basis for temporally-controlled protein depletion in hepatocytes. We took advantage of commercially available loxP-stop-loxP-Cas9 (LSL-Cas9) mice in which we could induce Cas9 in nearly all hepatocytes by injecting an adeno-associated virus expressing Cre recombinase from the hepatocyte-specific Tbg promoter (AAV-Cre, Extended Data Fig. 1d,e). To evaluate the efficiency and kinetics of protein depletion, we selected two long-lived, non-essential proteins: the mitochondrial enzyme MAO-B (encoded by Maob) and the nuclear lamin Lamin B2 (encoded by Lmnb2). After delivering sgMaob-mCherry or sgLmnb2-mCherry lentivirus to PD1 mice, we injected PBS or AAV-Cre at PD5 and harvested livers at various time points to evaluate protein levels in individual hepatocytes (Fig. 1d). By two weeks after Cas9 induction, MAO-B and Lamin B2 were depleted exclusively in mCherry-positive hepatocytes in mice injected with AAV-Cre (Fig. 1e,f, and Extended Data Fig. 1f,g). Importantly, this combination of lentiviral-mediated sgRNA delivery, AAV-Cre-mediated induction of Cas9, and resulting gene targeting did not induce detectable liver inflammation nor did it affect hepatocyte turnover (Extended Data Fig. 1h-k). This approach therefore offers an effective platform for hepatocyte-specific protein depletion and genetic screening at any point in the animal’s lifetime without obvious collateral perturbation of hepatocyte fitness.

A genome-wide screen in the liver

To enable genome-wide screening for diverse hepatocyte phenotypes, we generated an sgRNA library targeting all genes expressed in the developing, quiescent, and regenerating mouse liver. We performed RNA sequencing on livers at various time
points during mouse development and after liver injury and determined that 13,266 protein-coding genes were expressed (FPKM > 0.3) at one or more time points (Fig. 2a, Extended Data Fig. 2a,b, Supplementary Table 1). We generated an sgRNA library targeting 13,189 of these genes (average of 5 sgRNAs per gene) alongside a previously published set of 6,500 control sgRNAs (~2,000 non-targeting sgRNAs and ~4,500 sgRNAs targeting exonic and intronic regions of control genes) for a total of 71,878 unique sgRNAs (Extended Data Fig. 2c, Supplementary Table 2).

With this method in hand, we undertook a genome-wide screen for hepatocyte fitness (Fig. 2b). To screen for the ability of hepatocytes to both persist and proliferate, we elected to screen over a three-week period in neonatal development when hepatocytes undergo approximately three population doublings to increase liver mass.

We injected 5 x 10^7 TU of our lentiviral library into four female and four male LSL-Cas9 mice at PD1. At PD5, we harvested livers from two males and two females to evaluate the initial library representation. The sgRNA representation in these four livers correlated extremely well (Pearson r = 0.97) with the plasmid library and we detected all sgRNAs, indicating that we can effectively deliver and recover a genome-scale sgRNA library from the neonatal mouse liver (Fig. 2c, Supplementary Table 3). In the remaining mice, we induced Cas9 at PD5 and harvested their livers at PD26 to evaluate the final library representation. We used the Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK) algorithm to identify enriched and depleted genes based on statistical differences in their change in sgRNA abundance at PD26 relative to PD5 (Supplementary Table 3). Using a false discovery rate cutoff of 0.05, we identified 0-6 significantly enriched and 40-386 significantly depleted genes in individual mice,
indicating that our method can detect enriched and depleted genes in a single mouse. We also generated gene-level scores by calculating the median log2 fold change in abundance of sgRNAs targeting a given gene and observed a strong correlation across the four mice (Pearson r = 0.46 to 0.75, Fig. 2d, Extended Data Fig. 2d, Supplementary Table 3). To improve our power to detect enriched and depleted genes, we combined the data from all mice and tested whether the gene-level scores for each gene deviated significantly from those for all genes across the four mice. Using false discovery rate cutoffs of 0.05 and 0.25, we identified, respectively, 30 and 658 significantly enriched genes and 661 and 1,482 significantly depleted genes across all mice (Fig. 2e, Supplementary Table 3). Finally, we calculated a unified gene score representing the median log2 fold change for each gene across all mice (Supplementary Table 3). These gene scores were not biased by gene expression level or protein half-life, reaffirming that long-lived proteins were effectively depleted (Extended Data Fig. 2e,f). Together, these data establish the technical feasibility of genome-wide screening in the mouse liver. Importantly, while screening multiple mice in parallel increases the power to discover significant hits, a single mouse is sufficient to identify significantly enriched and depleted genes.

We next asked whether our screen reliably uncovered regulation of cell fitness. We first confirmed that genes established to be essential in cell culture were highly represented among the depleted genes across mice (Fig. 2f). To evaluate whether our screen could reveal regulation specific to hepatocyte fitness in the liver, we assessed the gene scores for two sets of genes known to affect hepatocyte fitness in vivo: (1) a set of 13 genes established as tumor suppressors in hepatocellular carcinoma
(expected to enrich) and (2) a set of seven genes required for hepatocyte viability
(expected to deplete)\textsuperscript{17} (Supplementary Table 3). Among the tumor suppressor genes, eight of the 13 genes were significantly enriched (FDR < 0.25, Fig. 2g). Among the genes required for hepatocyte viability, all seven genes were significantly depleted (FDR < 0.25, Fig. 2g). These results support our screen as a reliable platform for uncovering genetic regulation of hepatocyte fitness in the liver.

**Sex-specific regulation of hepatocyte fitness**

Screening primary cells in their native context has the unique power to capture all organismal regulation of cell fitness, something that cannot be achieved by screening cell lines in culture. One such advantage is the ability to investigate the influence of biological sex in an otherwise isogenic background. To determine whether genes can shape hepatocyte fitness in a sex-dependent manner, we compared the gene scores for all genes in males versus females. We identified three X-linked genes and nine autosomal genes with sex-specific effects on fitness (Fig. 3a). Two X-linked genes, \textit{Ddx3x} and \textit{Eif2s3x}, both involved in protein synthesis, exhibited the greatest differences between the sexes and were exclusively essential in females. Both of these genes escape X inactivation and have paralogs on the Y chromosome, \textit{Ddx3y} and \textit{Eif2s3y}, with similar function\textsuperscript{18}. Thus, it is possible that disruption of \textit{Ddx3x} and \textit{Eif2s3x} causes a fitness defect in female hepatocytes while male hepatocytes are functionally complemented by the Y chromosome paralogs. Among the remaining genes with sex-specific effects, we were surprised to identify an uncharacterized open reading frame, 1810030O07Rik, also known as \textit{Cxor38}, that when disrupted conferred a fitness
advantage exclusively in females. The Cxorf38 gene is also encoded on the X chromosome and has been shown to escape X inactivation but does not have a known Y chromosome paralog\textsuperscript{19}. It is expressed at low levels across tissues (FPKM ranging from 1-5), with a trend toward higher expression in embryonic versus adult tissues, and is predicted to encode a 320 amino acid protein comprised of a domain of unknown function\textsuperscript{20}. To confirm that Cxorf38 targeting indeed provides a female-specific fitness advantage, we performed in vivo competition assays to evaluate the fitness of hepatocytes depleted of Cxorf38 relative to control hepatocytes (Extended Data Fig. 3, Fig. 3b, top panel). Consistent with our screen results, we observed a significant expansion of Cxorf38-depleted hepatocytes relative to control hepatocytes exclusively in female mice (Fig. 3b, bottom panel). We note that cancer genomics efforts have not predicted Cxorf38 to function as a tumor suppressor gene, suggesting that its function might be limited to embryonic and neonatal development\textsuperscript{21,22}. Our method thus provides a unique opportunity to investigate how biological sex shapes cellular phenotypes and reveals twelve genes that regulate hepatocyte fitness in a sex-specific manner.

**Negative and positive regulation of hepatocyte fitness**

Having established that the majority of genes influence hepatocyte fitness in a sex-independent manner, we next sought to understand broad regulation of hepatocyte fitness by performing gene set enrichment analysis using the unified gene scores across the four mice. No gene sets were identified to be enriched in our screen. However, we noted that over 50% of the top 25 most enriched genes in our screen have been established to act as tumor suppressor genes in at least one context\textsuperscript{21,22}.
Indeed, a set of the top 50 computationally predicted pan-cancer tumor suppressor genes was significantly enriched in our screen\(^\text{21}\), something not typically observed in cell culture screens including those in mouse embryonic stem cells or human hepatocellular carcinoma cell lines\(^\text{23-26}\) (Fig. 3c,d, Supplementary Table 3). One possible explanation for this difference is that cell lines naturally harbor and/or accumulate inactivating mutations in tumor suppressor genes, thereby compromising identification of these genes in subsequent screening\(^\text{27}\). Our screen’s unique ability to, within only a few population doublings, reliably recover tumor suppressors highlights a key advantage of genome-wide screening in primary tissue—the ability to utilize wild-type cells and thereby identify all genetic perturbations that alter a chosen phenotype.

We next turned our focus to genes required for hepatocyte fitness. We identified several gene sets that were significantly depleted in our screen (Fig. 4a, Supplementary Table 4). The gene sets included those previously established as essential for fitness in cell culture, including ribosome, proteasome, spliceosome, and RNA polymerase\(^\text{1,2,12,16,25,26}\). However, we also identified several other gene sets not documented to be essential for cells in culture, including N-glycan biosynthesis, glycosaminoglycan biosynthesis/heparan sulfate, and antigen processing/presentation. Notably, these pathways all play major roles in the presentation or secretion of proteins at the cell surface\(^\text{28,29}\). Our screen thus uncovers requirements for cellular fitness shared by cells in culture but indicates additional, possibly cell non-autonomous, requirements for cell fitness in the organismal context.

**Cell non-autonomous control of hepatocyte fitness**
To identify factors uniquely important for cell fitness in the organismal context, we compared our screen to screens of mouse embryonic stem (ES) cells in culture and human hepatocellular carcinoma (HCC) cell lines in culture. We specifically chose ES cells and HCC cell lines in an effort to control for any species-specific and cell lineage-specific requirements for cell fitness. We identified four gene sets that were preferentially depleted in our screen relative to the cell culture screens: protein export, SNARE interactions in vesicular transport, antigen processing/presentation, and glycosaminoglycan biosynthesis/heparan sulfate (Fig. 4b, Supplementary Table 4). For the protein export and SNARE interactions gene sets, the most depleted genes in each set were depleted across all screens but to a greater extent in our screen (Extended Data Fig. 4a,b). However, for the glycosaminoglycan biosynthesis/heparan sulfate and antigen processing/presentation gene sets, some genes were depleted exclusively in our screen (Fig. 4c,d). Within the glycosaminoglycan biosynthesis/heparan sulfate gene set, Hs2st1 and Ndst1 were uniquely depleted in our screen (Fig. 4c). These two genes encode enzymes involved in the biosynthesis of heparan sulfate, a glycosaminoglycan typically conjugated to plasma membrane or extracellular matrix proteins. Heparan sulfate interacts with a variety of extracellular proteins and thereby modulates both mechanotransduction and signal transduction at the cell surface. Within the antigen processing/presentation gene set, Tap1 and B2m were uniquely and dramatically depleted in our screen, ranking as the 41st and 320th most depleted genes, respectively (Fig. 4d). Both of these genes are involved in—and required for—presentation of antigens at the cell surface by the class I major histocompatibility complex (MHC) pathway. Indeed, within the antigen processing/presentation gene set, eight of the 32
genes attributed to the class I MHC pathway were depleted in our screen whereas none of the 13 genes attributed to the class II MHC pathway exhibited depletion (FDR < 0.25, Extended Data Fig. 4c). Our screen thus uncovered two novel regulators of cell fitness—heparan sulfate and class I MHC—that are uniquely essential for cells in the organism and likely operate through interactions with the extracellular matrix and other cells.

The class I MHC pathway presents intracellular antigens at the cell surface. In this pathway, cytoplasmic proteins are degraded by the proteasome, resulting peptides are transported into the endoplasmic reticulum (ER) via a Tap1/Tap2 heterodimer, and these peptides are then loaded onto the class I MHC complex comprised of a heavy chain (H2) and β2-microglobulin (B2m) (Fig. 4e). At the cell surface, class I MHC can interact with both cytotoxic CD8 T cells and natural killer (NK) cells. The latter interaction can provide a pro-survival role by preventing NK cell cytotoxicity. However, loss of class I MHC alone should not be sufficient to induce NK cell cytotoxicity. Classically, NK cell activation requires both a loss of inhibitory signals, via loss of class I MHC, and presence of activating signals, expressed on the surface of transformed or infected cells. Although our screening approach involves viral infection of hepatocytes, any inflammation resulting from lentiviral and AAV vectors has been shown to resolve within 72 hours and, consistently, we did not observe any inflammation one week after the combination of lentiviral and AAV-Cre infection (Extended Data Fig. 1h,i). It was therefore surprising that loss of class I MHC alone would have such a dramatic impact on hepatocyte fitness. To validate this finding, we performed an in vivo competition between cells depleted of Tap1 and control cells (Extended Data Fig. 4d, Fig. 4f, top...
Consistent with our screen, hepatocytes depleted of Tap1 were lost over time (Fig. 4f, bottom panel). To test whether this fitness defect occurs via NK cell-mediated cytotoxicity, we repeated the competition in the setting of NK cell depletion and observed reduced loss of Tap1-deficient cells (Extended Data Fig. 4e, Fig. 4f, bottom panel). Our results confirm an essential role for class I MHC in hepatocyte viability that acts cell non-autonomously via NK cells. That loss of class I MHC alone is sufficient to induce hepatocyte clearance implies far more rigorous immune surveillance of somatic cells than previously appreciated.

**DISCUSSION**

Herein, we successfully performed, to our knowledge, the first genome-wide CRISPR-Cas9 screen for cell fitness in a living organism. In addition to identifying requirements shared by cells in culture, we also uncovered sex-specific and cell non-autonomous regulation of cell fitness. In particular, we have revealed an essential role for class I MHC in hepatocyte viability. We note that expression of class I MHC at the cell surface is exquisitely sensitive to cellular perturbations and that ER stress of various etiologies can impair its expression. Moreover, obesity and metabolic syndrome are known to induce ER stress in hepatocytes. We speculate that perhaps the rigorous requirement for class I MHC in hepatocytes serves to identify and eliminate stressed cells from the organ and that this surveillance may contribute to the inflammation associated with metabolic syndrome. It will be interesting to determine if this...
requirement extends to other somatic cell types and whether this serves as a stringent quality control mechanism for cells in an organism.

Our screen’s ability to uncover regulation of cell fitness not previously identified in cell culture screens emphasizes the necessity and power of genome-wide screening in the living organism. Importantly, our lentiviral delivery approach coupled with inducible Cas9 can readily be applied to screen diverse hepatocyte phenotypes—from fundamental processes universal to all cells to specialized metabolic and regenerative phenomena unique to hepatocytes—and minimally requires just one mouse. Our method can also be adapted to other CRISPR-based approaches including CRISPR interference/activation and Perturb-Seq\textsuperscript{37}. More broadly, our system establishes the feasibility of genome-wide screening in a living organism and inspires efforts to bring this technology to other organs. Collectively, these diverse applications will bring the experimental tractability once restricted to cell culture to the living organism, enabling unprecedented insight into mammalian physiology and disease.

REFERENCES


**METHODS**

**Animals**

C57BL/6J mice (strain 000664) and LSL-Cas9 mice (strain 026175) were purchased from the Jackson Laboratory. Mice were either singly- or group-housed with a 12-hour light-dark cycle (light from 7 AM to 7 PM, dark from 7 PM to 7 AM) in a specific-pathogen-free animal facility with unlimited access to food and water. To deliver lentivirus, up to 100 µL of lentivirus in PBS was injected into the temporal vein of
postnatal day one mice. For protein depletion tests, mice were injected with 1.25 x 10^7 transduction units (TU) of sgRNA-mCherry lentivirus. For the screen, mice were injected with 5 x 10^7 TU of sgRNA-mCherry lentiviral library. For validating Cxorf38 and Tap1, mice were injected with 1 x 10^7 TU of an equal mixture of sgAAVS1-mTurq2 and sgCxorf38-mCherry or sgTap1-mCherry lentiviruses. To deliver AAV-Cre, a stock solution of AAV8-TBG-Cre (Addgene 107787-AAV8) was diluted in PBS to a total volume of 20 µL and injected intraperitoneally into postnatal day five mice. For protein depletion tests, the screen, and validating hits, mice were injected with 2 x 10^{11} GC of AAV-TBG-Cre. To deplete NK cells, 15 µg/g of anti-NK1.1 (clone PK136, BioXCell) was injected intraperitoneally every three days beginning at postnatal day five. All animal procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care.

Cell lines

The mouse hepatocyte cell line AML12 was purchased from the American Type Culture Collection (ATCC) and cultured in DMEM/F12 medium supplemented with 10% fetal bovine serum, 10 µg/mL insulin, 5.5 µg/mL transferrin, 5 ng/mL selenium, and 40 ng/mL dexamethasone (ThermoFisher Scientific). HEK-293T cells were cultured in DMEM supplemented with 10% fetal bovine serum, 100 units/mL penicillin, and 100 µg/mL streptomycin (ThermoFisher Scientific).

Vector construction
The vector was produced through the following steps: 1) removal of the EFS-NS promoter and Cas9 from the parental vector and insertion of a hepatocyte-specific promoter driving dsRed expression, 2) replacement of dsRed with mCherry or mTurq2, and 3) removal of the puromycin resistance cassette.

To produce pLCv2-opti-stuffer-dsRed-puro, 100 ng of a synthetic gblock encoding the HS-CRM8-TTRmin module upstream of dsRed (Integrated DNA Technologies) and 1 µg of sgOpti (gift from Eric Lander and David Sabatini, Addgene plasmid #85681)39, a lentiCRISPRv2 derivative containing an optimized scaffold (5'-GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTT-3')40 were digested sequentially with NheI and BamHI (New England Biolabs). The vector and fragment were purified using the QIAquick Gel Extraction Kit (Qiagen) and ligated with T4 DNA Ligase (New England Biolabs) in an 11 µL reaction to replace the EFS-NS promoter and Cas9 with the gblock fragment. 2.5 µL of the ligation was used to transform Stbl2 cells (Invitrogen) and DNA was isolated from ampicillin-resistant colonies with the QIAprep Spin Miniprep Kit (Qiagen). Clones were verified by Sanger sequencing (Quintara Biosciences) prior to retransformation and maxiprep using the ZymoPURE II Plasmid Maxiprep Kit (Zymo Research).

HS-CRM8-TTRmin-dsRed:

GAATTCGCTAGCACCAGCAGCCGGGAGGCTGCTGCTGTAATATAACCCAAGG
TCACCCCAGTTATCGGAGGACAAACAGGGGCTAAGTCCACACGCGTGGTACCCT
To construct pLCv2-opti-stuffer-mCherry-puro and pLCv2-opti-stuffer-mTurq2-puro, mCherry and mTurq2 were amplified from pKL028 and mTurquoise2-CMV (gifts from Iain Cheeseman), respectively, for 25 cycles with Q5 HotStart Polymerase (New...
England Biolabs) using the following primers (underlined nucleotides are homologous to the vector):

pLC_EBFP2_F: **GGTTCTAGAGCGCTGCCACC**ATGGTGAGCAAGGGCGAGGAG
pLC_EBFP2_R: **GCCGGATCC**CTTGTACAGCTCGTCCATGCC

Amplicons and pLCv2-opti-stuffer-dsRed-puro were digested with XbaI and BamHI HF (New England Biolabs) and purified, ligated, transformed, and DNA was isolated and sequence verified as above.

To construct pLCv2-opti-stuffer-mCherry and pLCv2-opti-stuffer-mTurq2, a fragment encompassing the WPRE and 3’LTR was amplified from pLCv2-opti-stuffer-mCherry-puro as above using the following primers:

Puro_removal_F: TGAACGCGTTAAGTCGACAATCAACC
Puro_removal_R: TCGAGGCTGATCAGCGGGTTTAAAC

The amplicon and pLCv2-opti-stuffer-mCherry-puro (or -mTurq2-puro) were digested with BsrGI-HF and Pmel (New England Biolabs) and purified as above. NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) was used to assemble 25 ng each of vector and fragment in a 20 µL reaction for 15 min at 50 °C. 50 µL of DH5-alpha cells were transformed with 2 µL assembly mix, and DNA was isolated and sequence verified as described above.
Individual sgRNAs were cloned as previously described\textsuperscript{41}, using the following oligonucleotides (Integrated DNA Technologies):

\begin{verbatim}
sgAAVS1_F: CACCGGGGCCACTAGGGACAGGAT
sgAAVS1_R: AAACATCCTGTCCCTAGTGCCGC
sgMaob_1_F: CACCGACGGATAAAGGATATACTTG
sgMaob_1_R: AAACCAAGTATATCCTTTATCCGTC
sgMaob_2_F: CACCGGGAAAATCATATGCTTCCTCAG
sgMaob_2_R: AAACCTGAAGCATATGATTTTCCC
sgLmnb2_1_F: CACCGAGGTACGGGAGACCCGACGG
sgLmnb2_1_R: AAACCCGTCGGTCTCCCGTACCTC
sgLmnb2_2_F: CACCGCTGCGCACCTACCTACCGT
sgLmnb2_2_R: AAACAGTGTCAGGTATGCGTACGC
sgCxorf38_1_F: CACCGGTCAACCACAAAGTGACAC
sgCxorf38_1_R: AAACGTGATCACTTTTGTGGTGACC
sgCxorf38_2_F: CACCGGTCAACCACAAAGTGACAC
sgCxorf38_2_R: AAACCTCAGTGCAGGATCGACAGAC
sgTap1_1_F: CACCGGGTGCCAACGAGCCACTGAG
sgTap1_1_R: AAACCTCAGTGCAGGATCGACAGAC
sgTap1_2_F: CACCGGGTAGAGAACGAATGACAC
sgTap1_2_R: AAACCTCAGTGCAGGATCGACAGAC
\end{verbatim}
**Lentivirus preparation, titration, and concentration**

HEK-293T cells were seeded at a density of 750,000 cells/mL in 20 mL viral production medium (IMDM supplemented with 20% inactivated fetal serum, Thermo Fisher Scientific). After 24 hours, media was changed to fresh viral production medium. At 32 hours post-seeding, cells were transfected with a mix containing 76.8 µL Xtremegene-9 transfection reagent (Thermo Fisher Scientific), 3.62 µg pCMV-VSV-G<sub>42</sub> (gift from Bob Weinberg, Addgene plasmid #8454), 8.28 µg psPAX2 (gift from Didier Trono, Addgene plasmid #12260), and 20 µg sgRNA plasmid in Opti-MEM (Thermo Fisher Scientific) to a final volume of 1 mL. Media was changed 16 hours later to fresh viral production medium. At 48 hours after transfection, virus was collected and filtered through a 0.45 µm filter, aliquoted, and stored at -80 °C until use.

To determine lentivirus titer, AML12 cells were transduced with a dilution series of lentivirus in the presence of 10 µg/mL polybrene for 16 hours. After four days, cells were harvested for flow cytometry analysis to determine percent of mTurq2- or mCherry-positive cells. To concentrate lentivirus, lentiviral supernatant was ultracentrifuged at 23,000 RPM at 4 °C for 2 hours in an SW 32 Ti swinging bucket rotor (Beckman Coulter). After centrifugation, media was decanted and pellets were air-dried at room temperature for 15 minutes. Pellets were then resuspended in PBS at room temperature for 30 minutes with gentle trituration. Concentrated lentivirus in PBS was stored for up to one week at 4 °C prior to injection into mice.

**sgRNA validation**
To validate individual sgRNA sequences for the ability to deplete target proteins, AML12 cells stably expressing Cas9-EGFP were transduced with sgAAVS1-mTurq2 or sgRNA-mCherry lentiviruses as described for lentivirus titration. After four days, transduced cells were subjected to fluorescence activated cell sorting to purify cells expressing both EGFP and mTurq2 or EGFP and mCherry and cultured for another seven days. Cells were then harvested to prepare protein lysates and perform immunoblotting. To validate sgRNAs targeting Tap1, cells were treated with 10 ng/mL of IFNγ 20 hours prior to harvest to upregulate Tap1 expression such that it could be detected by immunoblotting.

Immunostaining

Livers were harvested and fixed in 4% paraformaldehyde in PBS at room temperature for 16-24 hours. Tissues were then washed with PBS and frozen in O.C.T. Compound (Tissue-Tek). Tissue sections of 12 to 30 µm thickness were prepared using a cryostat and adhered to Superfrost Plus Slides (Fisher Scientific). Slides were stored at -20 °C until use. To visualize endogenous mCherry and mTurq2 fluorescence, slides were dried at room temperature for 15 minutes, rehydrated in PBS for 5 minutes, permeabilized with 1% Triton X-100 in PBS for 15 minutes, and counterstained with Alexa Fluor 488 Phalloidin (ThermoFisher Scientific) diluted 1:500 in blocking buffer (3% bovine serum albumin and 0.3% Triton X-100 in PBS). To immunostain for endogenous proteins, slides were dried at room temperature for 4-24 hours and rehydrated in PBS for 5 minutes. Antigen retrieval was then performed by pressure cooking slides in sodium citrate buffer (10 mM tri-sodium citrate dihydrate, 0.05% Tween-20, pH 6.0) for
20 minutes in an Instant Pot (Amazon). Slides were rinsed in PBS for 5 minutes, dried briefly, and sections outlined with an ImmEdge hydrophobic pen (Vector Laboratories). Sections were permeabilized with 1% Triton X-100 in PBS for 15 minutes and blocked with blocking buffer for one hour. Sections were then incubated in primary antibodies diluted in blocking buffer at room temperature for 12-24 hours. Sections were washed with blocking buffer three times for 10 minutes each. Sections were then incubated in AlexaFluor secondary antibodies (ThermoFisher Scientific) diluted 1:1,000 in blocking buffer at room temperature for 1-2 hours. In some cases, 5 µg/mL Hoechst 33342 (ThermoFisher Scientific) was added to the secondary antibody solution. Sections were washed with blocking buffer twice for 10 minutes each followed by one wash with PBS for 5 minutes. Slides were then mounted in ProLong Gold Antifade reagent (ThermoFisher Scientific).

The following primary antibodies were used: Cas9 (1:200, clone 7A9-3A3, Abcam ab191468), asialoglycoprotein receptor 1 (ASGR1) (1:500, clone 114, Sino Biological 50083-R114), mCherry (1:500, clone 16D7, ThermoFisher Scientific M11217), monoamine oxidase B (MAO-B) (1:1,000, Novus Biologicals NBP1-87493), lamin B2 (1:1,000, clone EPR9701(B), Abcam ab151735), actin (1:250, clone AC-74, Sigma Aldrich A2228), CD45 (1:500, Abcam ab10558), and Ki-67 (1:200, clone SP6, Abcam ab16667). The Cas9 antibody was directly conjugated to AlexaFluor 647 using the AlexaFluor 647 Antibody Labeling Kit (ThermoFisher Scientific). The actin antibody was directly conjugated to DyLight 405 using the DyLight 405 antibody labeling kit (ThermoFisher scientific).
**Image analysis**

Images were acquired using a CSU-22 spinning disc confocal head (Yokogawa) with Borealis modification (Andor) mounted on an Axiovert 200M microscope (Zeiss) with 10X or 40X objectives (Zeiss), an Orca-ER CCD camera (Hamamatsu), and MetaMorph acquisition software (Molecular Devices).

Images were analyzed using Volocity (Quorum Technologies). To measure MAO-B and lamin B2 intensity, a single Z plane at the center of the cell was identified and the cytoplasm or nucleus was outlined to measure the signal intensity per µm. A similar procedure was done on sections stained only with secondary antibodies to calculate the average background intensity. This average background intensity was subtracted from each MAO-B and lamin B2 intensity measurement and the background-subtracted measurements were then normalized within a given sample (mCherry-positive or -negative hepatocytes within a single liver).

**Immunoblotting**

To prepare protein lysates from cultured cells, 500,000 cells were pelleted and resuspended in 100 µL of 2X Laemmli sample buffer (100 mM Tris pH 6.8, 20% glycerol, 4% SDS, 0.02% bromophenol blue, 5% β-mercaptoethanol). Lysates were homogenized by pipetting followed by trituration through a 31-gauge needle and boiled for 5 minutes. Samples were separated on homemade polyacrylamide gels and transferred to Immobilon-FL membranes (Millipore) via wet transfer. Membranes were
blocked in 5% bovine serum albumin in TBST (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% Tween-20) for 1 hour at room temperature. Membranes were incubated in primary antibody diluted in blocking solution at 4 °C with rocking overnight and washed with TBST for five minutes five times. Membranes were incubated in HRP-conjugated goat anti-rabbit secondary antibody (Abcam ab205718) diluted 1:50,000 in blocking solution at room temperature with rocking for one hour and washed with TBST for five minutes five times. Membranes were incubated in ECL Prime Western Blotting Detection Reagent (GE Healthcare) for five minutes and exposed to CL-Xposure Film (Thermo Fisher Scientific).

The following primary antibodies were used: Cxorf38 (1:1,000, Invitrogen PA5-62139) and Tap1 (1:1,000, Cell Signaling Technology 12341).

**Flow cytometry**

To quantify NK cells in the liver by flow cytometry, livers were harvested, minced into 1 mm pieces, and incubated in digestion solution (1% fetal bovine serum, 40 U/mL DNase, 125 U/mL collagenase type IV in Hank’s balanced salt solution (HBSS) containing 10 mM HEPES) at 37 °C for 30 minutes with shaking. Digested tissue was passaged through a 70 µm strainer, combined with 10 mL of HBSS containing 10 mM HEPES, and pelleted at 500 g for 5 minutes. Pellet was resuspended in 5 mL of 45% Percoll (Sigma) in PBS and gently layered onto 3 mL of 55% Percoll in PBS. Gradient was centrifuged at 900 g at 4 °C for 20 minutes with low acceleration and no braking. The cell layer at the interface of the two Percoll layers was removed and transferred to
10 mL of HBSS containing 10 mM HEPES and pelleted at 500 g for 5 minutes. Cells were diluted to <10^6 cells per 100 µL in flow cytometry buffer (1% fetal bovine serum and 0.05% sodium azide in PBS) and incubated with conjugated primary antibodies on ice in the dark for 20 minutes. Cells were washed twice with flow cytometry buffer, resuspended in flow cytometry buffer, incubated with 500 ng/mL DAPI for 5 minutes, and analyzed using a LSRFortessa flow cytometer with FACSDiva software (BD Biosciences).

The following fluorophore-conjugated primary antibodies were used: CD3-Brilliant Violet 421 (1:20, clone 17A2, BioLegend 100227) and NK1.1-AlexaFluor 488 (1:200, clone PK136, BioLegend 108717).

RNA sequencing

For surgical resection time points, partial hepatectomies were performed on 8 week-old mice as previously described. For toxic injury time points, 8 week-old mice were injected intraperitoneally with 2 µL/gram of 25% carbon tetrachloride diluted in corn oil (Sigma Aldrich). For all time points, livers from three male C57BL/6J mice were harvested, flushed with PBS, immediately immersed in RNAlater (Qiagen), incubated at room temperature for 24 hours, and stored at -20 °C until future use. To isolate RNA, 30 mg of each tissue was removed from RNAlater and homogenized in 700 µL of QIAzol lysis reagent (Qiagen) using the TissueRuptor homogenizer (Qiagen). RNA was purified using the miRNeasy Kit (Qiagen) according to kit instructions and eluted in 30 µL of nuclease-free water. RNA sequencing libraries were prepared using KAPA mRNA
HyperPrep Kit (KAPA Biosystems) according to manufacturer instructions. Briefly, 0.1-1 ug of total RNA was enriched for polyadenylated sequences using oligo-dT magnetic bead capture. The enriched mRNA fraction was then fragmented and first-strand cDNA generated using random primers. Strand specificity was achieved during second-strand cDNA synthesis by replacing dTTP with dUTP to quench the second strand during amplification. The resulting cDNA was A-tailed and ligated with indexed adapters. The library was amplified using a DNA polymerase that cannot incorporate past dUTPs to quench the second strand during PCR. The libraries were quantified using a KAPA qPCR Library Quantification Kit (KAPA Biosystems) as per manufacturer instructions. The samples were sequenced on a HiSeq 2500 (Illumina) based on qPCR concentrations. Base calls were performed by the instrument control software and further processed using the Offline Base Caller version 1.9.4 (Illumina). Samples were mapped with STAR version 2.6.1a to the mouse genome release mm10, using a gtf file from ENSEMBL version GRCh38.91, and setting the maximum intron length ("alignIntronMax") parameter to 50000. We ran featureCounts version 1.6 to assign reads to genes using the same gtf file and setting "-s" parameter to 2. We normalized gene counts with DESeq2 version 1.22.2. FPKMs were calculated using the function fpkm within the DESeq2 package. The FPKMs values for the three replicates were averaged, and protein coding genes were selected based on the annotation in the gtf file.

sgRNA library preparation
Genes with an average FPKM > 0.3 in any of the RNA sequencing time points were chosen to build a liver transcriptome-wide library. sgRNA sequences were designed using the Broad Institute GPP sgRNA Designer\textsuperscript{13,47} using the Azimuth 2.0 rule set. For genes which were not identified by the program, alternative gene names from ENSEMBL versions GRCm38.76-.93 were attempted. A small number of designed sgRNAs targeted multiple genes; the sgRNA names and gene names were manually annotated to indicate all targeted genes for these cases. Non-targeting and control-gene-targeting sgRNAs from Doench et al. 2014\textsuperscript{13} were also included. sgRNA sequences from this control set that were identical to a sequence already in our library were annotated according to the targeted gene; those that did not overlap with sequences in our sgRNA library were annotated as control sgRNAs. The library contains 71,878 sgRNAs targeting 13,189 genes.

For sgRNAs beginning with a nucleotide other than G, a G was prepended. The following adapters were added to all sgRNA sequences:

Upstream: 5’-TATCTTGAGAAAGGACGAAACACC-3’

Downstream: 5’-AAGAGCTATGCTGGAAACAGCATAGC-3’

Multiple rounds of cloning were combined to generate the final plasmid library. The oligonucleotide library (Agilent Technologies) was amplified for 16 cycles using Q5 HotStart Polymerase (New England Biolabs) using a gradient annealing temperature ranging from 50-62 °C across 8, 50 µL reactions. Reactions were pooled and purified by...
DNA Clean and Concentrator 5 (Zymo Research). pLCv2-opti-stuffer-mCherry was digested as described \(^{41}\), and either gel purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research) followed by Ampure XP bead purification (Beckman Coulter) or DNA Clean and Concentrator 5. The library was assembled using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) in 4 x 20 µL reactions at 50 °C for 1 hour using 100 ng of vector per 5-10 ng of PCR amplicon. The reactions were combined and 2.5 µL of the assembly reaction or a control reaction without amplicon were used to transform NEB5-alpha cells (New England Biolabs) to measure background assembly. Subsequently, the assembly reactions were combined, concentrated using Ampure XP beads, resuspended in 8 µL water, and used to electroporate 1-4 tubes of Endura DUO electrocompetent cells (Lucigen) at 1.8 kV distributed over 2 cuvettes (0.1 cm gap width) per tube using a Micropulser Electroporator (Bio-Rad Laboratories). 10-fold serial dilutions of a 10 µL aliquot were plated on LB plates with ampicillin at 100 µg/mL to assess electroporation efficiency, and the remainder of each electroporation (2 cuvettes) was plated on LB agar supplemented with 100 µg/mL ampicillin in 4 x 245 mm square bioassay dishes (Corning). Plates were incubated overnight at 30 °C and colonies were scraped the next morning. DNA was isolated using the ZymoPURE II Plasmid Maxiprep Kit (Zymo Research). Plasmid DNA from multiple rounds of assembly and electroporation were combined according to the measured electroporation efficiency to achieve 25-fold coverage of the library. sgRNA representation was measured by high-throughput sequencing as described below.
To improve coverage of some of the sgRNAs in the library, a second library containing ~7,500 sgRNAs was synthesized and cloned as above, with the following modifications: assembly was performed using NEB Gibson Assembly mix (New England Biolabs) using a ratio of 200 ng vector : 10 ng sgRNA in each 20 µL reaction, and the final combined and concentrated reaction was used to electroporate a single tube of Endura DUO cells.

Subsequent propagation of the plasmid library was performed using 50 ng plasmid library per single tube of Endura DUO cells.

All steps were performed according to manufacturer’s instructions, except where noted.

**Genomic DNA isolation**

Livers were harvested from mice, separated into individual lobes, minced into 15 mg pieces using a razor blade, snap-frozen in liquid nitrogen, and stored at -80 °C until use. Genomic DNA (gDNA) was isolated from livers using the illustra blood genomicPrep Mini Spin Kit (Cytiva) using one column for every 7.5 mg of tissue. The manufacturer’s protocol was used with the following modifications: 20 µL of 10 mg/mL Proteinase K (Millipore-Sigma) solution in water was added per 7.5 mg of tissue. Tissue was disrupted by thoroughly pipetting prior to adding lysis buffer, vortexing, and incubating at 56 °C overnight. Elution was performed using 25 µL of water pre-heated to 70 °C. Samples were combined by lobe and concentration was measured using the Qubit dsDNA HS Assay Kit (Invitrogen).
For the induction, equal amounts of gDNA from each lobe were combined within each mouse, and equal inputs from four mice were combined to prepare a single sequencing library. For the endpoint, gDNA from each lobe within a mouse was combined proportionally to the average lobe mass across mice measured at liver harvest. A sequencing library was prepared for each mouse individually using equal total gDNA input per mouse.

**Sequencing library preparation and DNA sequencing**

All PCR reactions were performed in 50 µL reactions using ExTaq Polymerase (Takara Bio) with the following program:

1 cycle 95 °C 5 min
14 or 28 cycles 95 °C 10 sec
60 °C 15 sec
72 °C 45 sec
1 cycle 72 °C 5 min
1 cycle 4 °C hold

Using the following primers:

Forward: AATGATACGGCGACCACCAGATCTACACCGACTCGGTGCCACTTTT
Reverse:
CAAGCAGAAGACGGCATACGAGATCnnnnnnTTTCTTGGGTAGTTTGCAGTTTT

Where “nnnnnn” denotes the barcode used for multiplexing.

10 ng of plasmid DNA was amplified for 14 cycles in 4 x 50 µL reactions. 1, 3, or 6 µg of gDNA was initially amplified for 28 cycles in 50 µL test PCR reactions. Subsequently, 226 µg of gDNA (induction) was used in 38 reactions, or 75 µg of gDNA (endpoint) was used in 25 reactions per mouse. All reactions were cleaned and concentrated using Ampure XP beads prior to sequencing for 50 cycles on an Illumina Hiseq 2500 using the following primers:

Read 1 sequencing primer:
GTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC

Index sequencing primer:
TTTCAAGTTACGGAACATGATAGTCCATTTTAAAATTTTAAAAACTGCAA
ACTACCCAAGAAA

Base calls were performed by the instrument control software and further processed using the Offline Base Caller (Illumina) v. 1.9.4.

Screen analysis
For initial measurement of sgRNA representation in the plasmid library or induction time point, sequencing reads were mapped to the library, each sgRNA was given a pseudocount of 1, and reads per million (RPM) was calculated as previously described\(^\text{12}\). Raw counts were processed using MAGeCK for downstream analysis\(^\text{14}\).

The plasmid library and induction timepoint were used as control samples and to estimate variance, and each endpoint mouse was processed separately. For mouse 4, sgLmnb2_1 was removed prior to MAGeCK analysis, as the high representation of this sgRNA (an sgRNA used for development of the screening method) was likely due to contamination during sequencing library preparation. Counts data from Shohat et al. (day 18) and Tzelepis et al. (day 14) were processed individually using MAGeCK\(^\text{23,24}\).

The corresponding plasmid libraries were used as control samples. For our screen, Shohat et al., and Tzelepis et al., the size factor for normalization was estimated using the control sgRNA set (Supplementary Table 2). For Tzelepis et al., the three replicate day 14 samples were processed together to generate a single gene score, and those three samples were used to estimate variance. For all screens, the gene test FDR threshold was set to 0.05, the sgRNA p-value was FDR-adjusted, and the gene score was calculated using the median. Twenty-four human hepatocellular carcinoma screens from the CRISPR (Avana) Public 20Q4 release were downloaded from the Broad DepMap portal using “Liver” as a lineage filter and “Hepatocellular Carcinoma” as a lineage subtype filter\(^\text{25,26}\).

All downstream analyses were performed in R version 3.6.0, and all plots were generated in either base R, using the R corrplot package, or in GraphPad Prism Version
For comparisons within our screen, the gene scores from individual mice were not normalized across mice, as each mouse serves as a replicate screen. The gene score for each gene across mice was tested against all gene scores using an unpaired two-sample Wilcoxon test. The p-values from this test were adjusted using the Benjamini-Hochberg (FDR) procedure. The median log2 fold change across mice was used as input for pre-ranked GSEA using the c2.cp.kegg.v7.1.symbols.gmt gene sets. For comparisons between screens from different sources, all screens from all the sources in the specific comparison were quantile normalized to one another using the preprocessCore R package prior to calculating the median log2 fold change within the screens from each source. This normalized median log2 fold change was subtracted from the normalized median log2 fold change of our screens to generate a differential score used as input for pre-ranked GSEA using the c2.cp.kegg.v7.1.symbols.gmt gene sets. For converting mouse gene symbols to human gene symbols, the Mouse_Gene_Symbol_Remapping_to_Human_Orthologs_MSigDB.v7.1.chip was used (note that this excludes genes that have multiple annotations in either human or mouse).

Pearson correlation was used to compare gene effects between mice. Spearman correlation was used to compare gene effects with liver mRNA expression and protein half-life in mouse cells\textsuperscript{15}. The TUSON dataset of predicted tumor suppressor genes was sorted by ascending FDR q-value and the top 50 genes present in the compared datasets were used\textsuperscript{21}. Distribution differences were tested and p-values were calculated using the Kolmogorov-Smirnov test. A one-sided test was used for gene sets for which
a phenotype could be predicted (core essential genes, tumor suppressor genes, and control enriched and depleted genes); all other comparisons used a two-sided test.

For comparison of sex-specific fitness effects, only genes with an average of > 2 sgRNAs detected across mice were considered. For sex-specific enriched genes, a median fold change (log2) > 0.5 across two mice of a given sex and an absolute median fold change (log2) difference of > 0.25 compared to the other sex was required. To identify true tumor-suppressor-like genes, a median fold change (log2) > -0.5 was required in mice of the other sex. For sex-specific depleted genes, a median fold change (log2) of < -0.5 across two mice of a given sex and an absolute median fold change (log2) difference of > 0.75 compared to the other sex was required. To identify true sex-specific essential genes, a median fold change (log2) > -0.5 was required in mice of the other sex.

Software information

STAR version 2.6.1a
Conda version 4.9.2
MAGeCK-RRA version 0.5.9.2
R version 3.6.0

featureCounts version 1.6
DESeq2 version 1.22.2
preprocessCore version 1.48.0
corrplot version 0.84
GSEA version 4.1.0

Mouse_Gene_Symbol_Remapping_to_Human_Orthologs_MSigDB.v7.1.chip

Human_Symbol_with_Remapping_MsigDB.v7.1.chip

c2.cp.kegg.v7.1.symbols.gmt

GraphPad Prism version 7.0d

DATA AVAILABILITY

Sequencing data generated in this manuscript is in the process of being uploaded to publicly accessible online repositories. All other data are included within the manuscript figures or as supplementary information.

METHODS REFERENCES


43. Knouse, K. A., Lopez, K. E., Bachofner, M. & Amon, A. Chromosome Segregation

21 (2013).

program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 

46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and

47. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize

**ACKNOWLEDGMENTS**

We thank Mehreen Khan and Keya Viswanathan of the Whitehead Institute Functional
Genomics Platform for technical assistance, Amanda Chilaka and Sumeet Gupta of the
Whitehead Institute Genome Technology Core for high-throughput sequencing, Inma
Barrasa of the Whitehead Institute Bioinformatics and Research Computing Core for
RNA sequencing analysis, the Whitehead Institute Flow Cytometry Core for cell sorting
and the Whitehead Institute Keck Microscopy Facility for microscopy. We thank Agilent
Technologies for their commitment to supporting academic research. We thank Leah
Schmidt for helpful discussions and Iain Cheeseman, Jonathan Weissman, Peter
Reddien, Raghu Chivukula, Sharon Grossman, Tobiloba Oni, and members of the
Knouse laboratory for their comments on the manuscript. This work was supported by
the National Institutes of Health NIH Director’s Early Independence Award
(DP5OD026369) to K.A.K. K.A.K. is also supported by the Scott Cook and Signe Ostby Fund.

AUTHOR CONTRIBUTIONS
H.R.K. and K.A.K. conceived the project, performed all experiments, analyzed the data, and wrote the manuscript.

COMPETING INTERESTS
H.R.K. and K.A.K. are co-inventors on a patent filed by the Whitehead Institute related to work in this manuscript.

Supplementary information is available for this paper
Correspondence and requests for materials should be addressed to K.A.K. All plasmids and the plasmid library are in the process of being deposited at Addgene.
Fig. 1 | Efficient delivery of sgRNAs and depletion of long-lived proteins in the liver. a, Lentiviral vectors for U6-driven expression of an sgRNA and hepatocyte-specific-expression of a fluorescent reporter (mCherry or mTurq2). b, Images of endogenous mCherry and mTurq2 fluorescence in livers from mice four days after injection of an equal mixture of sgAAVS1-mCherry and sgAAVS1-mTurq2 lentiviruses. Livers were counterstained with phalloidin (green) to label actin. Scale bars, 100 µm. c, Percent mCherry-, mTurq2-, and double-positive hepatocytes in livers from mice four days after injection with an equal mixture of sgAAVS1-mCherry and sgAAVS1-mTurq2 lentiviruses. n = 3 mice per dose and 200 hepatocytes per mouse. Error bars indicate
standard deviation. **d,** Scheme for inducing protein depletion in LSL-Cas9 mice. **e,** Images of livers from LSL-Cas9 mice injected with sgMaob-mCherry followed by PBS or AAV-Cre immunostained for mCherry (magenta), MAO-B (green), and actin (blue). **f,** Cytoplasmic MAO-B intensity per µm in mCherry-positive and mCherry-negative hepatocytes from LSL-Cas9 mice injected with sgMaob-mCherry followed by PBS or AAV-Cre. *n = 1* male and *1* female mouse per condition and *25* cells per mouse. Closed and open circles represent values from male and female mouse, respectively.
**Figure 2**

**a** Protein-coding genes expressed in liver ubiquitously expressed.

**b** Postnatal day 1, Postnatal day 5, Postnatal day 26.

**c** Diagram showing sgRNA representation in livers four days after injection with lentiviral library relative to the sgRNA representation in the plasmid library expressed as reads per million (RPM). n = 2 male and 2 female mice. Pearson correlation $r = 0.97$.

**d** Pairwise comparisons of median fold change (log2) for each gene for each mouse at the endpoint of the screen.

**e** Genes ranked by median gene score across mice (log2) with significantly depleted genes denoted by red points and significantly enriched genes denoted by blue points (FDR < 0.05 by two-sample Wilcoxon test).

**f** Core essential genes (red bars) positioned based on gene rank as in (e) revealing their significant depletion in individual mice and across mice. $p < 2.2 \times 10^{-16}$.

---

**Fig. 2 | A genome-wide screen for hepatocyte fitness in the neonatal mouse liver.**

**a**, Number of protein-coding genes expressed in the liver as determined by RNA sequencing of livers at various time points. **b**, Scheme for performing a genome-wide screen for hepatocyte fitness in neonatal mice. **c**, Representation of sgRNAs in livers four days after injection with lentiviral library relative to the sgRNA representation in the plasmid library expressed as reads per million (RPM). $n = 2$ male and 2 female mice. Pearson correlation $r = 0.97$. **d**, Pairwise comparisons of median fold change (log2) for each gene for each mouse at the endpoint of the screen. **e**, Genes ranked by median gene score across mice (log2) with significantly depleted genes denoted by red points and significantly enriched genes denoted by blue points (FDR < 0.05 by two-sample Wilcoxon test). **f**, Core essential genes (red bars) positioned based on gene rank as in (e) revealing their significant depletion in individual mice and across mice. $p < 2.2 \times 10^{-16}$.
for each mouse and across mice by one-sided Kolmogorov-Smirnov test. Benjamini-Hochberg-adjusted Wilcoxon p-value (-log10) versus the median fold change of gene scores across mice (log2) for each gene in the screen. Highlighted are control gene sets consisting of tumor suppressor genes in hepatocellular carcinoma (expected to enrich, blue) and genes required for hepatocyte viability (expected to deplete, red).

Expected depleted $p = 3.5 \times 10^{-6}$ by one-sided Kolmogorov-Smirnov test, expected enriched $p = 4.7 \times 10^{-4}$ by one-sided Kolmogorov-Smirnov test.
**Fig. 3 | An *in vivo* screen uncovers sex-dependent fitness effects and canonical tumor suppressor genes.**

**a,** Median fold change (log2) across males versus median fold change (log2) across females for each gene. Highlighted are genes uniquely enriched in females (blue), genes uniquely enriched in males (cyan), genes uniquely depleted in females (red), and genes uniquely depleted in males (pink). Point size is proportional to the absolute difference in median log2 fold change between females and males. **b,** Scheme for *in vivo* competition between hepatocytes expressing sgAAVS1-mTurq2 and sgCxorf38-mCherry (top panel). Ratio (log2) of sgCxorf38-mCherry-positive to sgAAVS1-mTurq2-positive hepatocytes in livers from male and female mice 0 or 3 weeks after Cas9 induction (bottom panel). n = one mouse per gender and >200 hepatocytes per mouse for 0 week time point and two mice per gender and >350
hepatocytes per mouse for 3 week time point. \( *p = 0.016 \) (female 0 weeks versus female 3 weeks) by one-tailed Fisher’s exact test. c, Cumulative fraction of tumor suppressor genes (cyan, blue) and other genes (black, grey) based on quantile-normalized median fold change (log2) of their gene scores across screens in mouse embryonic stem (ES) cells and our screen. ES cells \( p > 0.05 \) by one-sided Kolmogorov-Smirnov test, our screen \( p = 6.5 \times 10^{-4} \) by one-sided Kolmogorov-Smirnov test. d, Cumulative fraction of tumor suppressor genes (cyan, blue) and other genes (black, grey) based on quantile-normalized median fold change (log2) of their gene scores across screens in human hepatocellular carcinoma (HCC) cell lines and our screen. HCC cells \( p > 0.05 \) by one-sided Kolmogorov-Smirnov test, our screen \( p = 0.0028 \) by one-sided Kolmogorov-Smirnov test.
Fig. 4 | Class I MHC is essential for hepatocyte fitness.
a, KEGG gene sets exhibiting significant depletion (FDR q-value < 0.05) at the endpoint of the screen ranked by FDR q-value (-log10). Bars extending to the end of the plot indicate an FDR q-value of 0. b, KEGG gene sets exhibiting significant depletion (FDR q-value < 0.05) in our screen relative to screens in either mouse embryonic stem (ES) cells (dark grey bars) or human hepatocellular carcinoma (HCC) cell lines (light grey bars) ranked by FDR q-value (-log10). Bars extending to the end of the plot indicate an FDR q-value of 0. c, Quantile-normalized median fold change (log2) for genes in the KEGG gene set for glycosaminoglycan biosynthesis and heparan sulfate in ES cell screens, HCC cell line screens, and our screen. Genes uniquely depleted in our screen are highlighted in red. The bounds of the box indicate the first and third quartiles, and the whiskers extend to the furthest data point that is within 1.5 times the interquartile range. d, Quantile-normalized median fold change (log2) for genes in the KEGG gene set for antigen processing and presentation in ES cell screens, HCC cell line screens, and our screen. Genes uniquely depleted in our screen are highlighted in red. The bounds of the box indicate the first and third quartiles, and the whiskers extend to the furthest data point that is within 1.5 times the interquartile range. e, Pathway for presentation of peptide antigens on class I MHC complexes. Created with BioRender.com. f, Scheme for in vivo competition between hepatocytes expressing sgAAVS1-mTurq2 and sgTap1-mCherry (top panel). Ratio (log2) of sgTap1-mCherry-positive to sgAAVS1-mTurq2-positive hepatocytes in livers from mice at 0 or 3 weeks after Cas9 induction without or with NK cell depletion (bottom panel). Closed and open circles represent values from male and female mice, respectively. n = one male and one female mouse per time point and >100 hepatocytes per mouse for 0 week time point.
and >200 hepatocytes per mouse for 3 week time point. **p = 0.0031 (0 weeks versus 3 weeks without NK depletion) and *p = 0.026 (3 weeks without NK depletion versus 3 weeks with NK depletion) by one-tailed Fisher’s exact test.
Extended Data Figure 1

a

<table>
<thead>
<tr>
<th>Litter</th>
<th>Sex</th>
<th>Volume (µL)</th>
<th>Mass (mg)</th>
<th>Proportion Hepatocytes</th>
<th>Average Hepatocyte Volume (µL)</th>
<th>Number of Hepatocytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>50</td>
<td>41.5</td>
<td>0.77041117</td>
<td>0.000032937</td>
<td>13,228,483.44</td>
</tr>
<tr>
<td>1</td>
<td>M</td>
<td>75</td>
<td>56</td>
<td>0.71753026</td>
<td>0.000003311</td>
<td>16,253,328.14</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>50</td>
<td>50.1</td>
<td>0.95585104</td>
<td>0.000002877</td>
<td>12,085,246.85</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>50</td>
<td>36.8</td>
<td>0.72751487</td>
<td>0.000002543</td>
<td>14,304,263.60</td>
</tr>
</tbody>
</table>

b

2.5 x 10^7 TU

4 days

14 weeks

mCherry mTurq2 Actin

c

Positive hepatocytes (%)

4 days 14 weeks

Time post-injection

mTurq2 mCherry Both

d

0 GC

2.5x10^7 GC

ASGR1 Cas9 Hoechst Merge

e

Positivity of hepatocytes (%)

Dose (GC)

f

PBS AAV-Cre

mCherry Lamin B2 Hoechst Merge

g

Nuclear lamin B2 intensity per µm

PBS AAV-Cre PBS AAV-Cre

1 week 2 weeks

mCherry negative mCherry positive

h

Unperturbed

sgRNA-mCherry AAV-Cre

CD45 mCherry Hoechst

i

CD45-positive cells per 40X field

Unperturbed sgRNA-mCherry AAV-Cre

j

Unperturbed

sgRNA-mCherry AAV-Cre

Ki67 mCherry Actin Hoechst

k

Ki67-positive hepatocytes (%)

Unperturbed sgRNA-mCherry AAV-Cre

mCherry negative mCherry positive
Extended Data Fig. 1 | Efficient delivery of sgRNAs and depletion of long-lived proteins in the liver. a, Table of values used to estimate number of hepatocytes in postnatal day one livers. Liver volume was measured by volume displacement and percent hepatocytes and hepatocyte volume were measured by immunostaining and microscopy. b, Images of endogenous mCherry and mTurq2 fluorescence in livers from mice four days or 14 weeks after injection of $2.5 \times 10^7$ TU of an equal mixture of sgAAVS1-mCherry and sgAAVS1-mTurq2 lentiviruses. Scale bars, 100 µm. c, Percent mCherry-, mTurq2-, and double-positive hepatocytes in livers from mice four days or 14 weeks after injection of $2.5 \times 10^7$ TU of an equal mixture of sgAAVS1-mCherry and sgAAVS1-mTurq2 lentiviruses. n = 3 mice per time point and 200 hepatocytes per mouse. Error bars indicate standard deviation. d, Images of livers from LSL-Cas9 mice injected with 0 or $2 \times 10^{11}$ GC of AAV-Cre on postnatal day five and harvested four days thereafter immunostained for ASGR1 (hepatocyte marker, magenta) and Cas9 (green) and counterstained with Hoechst (blue). Scale bars, 15 µm. e, Percent Cas9-positive hepatocytes in livers from LSL-Cas9 mice injected with varying doses of AAV-Cre on postnatal day five and harvested four days thereafter as determined by Cas9 immunostaining. n = 1 male and 1 female mouse per dose and 200 hepatocytes per mouse. Error bars indicate standard deviation. f, Images of livers from LSL-Cas9 mice injected with sgLmnb2-mCherry followed by PBS or AAV-Cre immunostained for mCherry (magenta) and lamin B2 (green) and counterstained with Hoechst (blue). Scale bars, 45 µm. g, Nuclear lamin B2 intensity per µm in mCherry-positive and mCherry-negative hepatocytes from LSL-Cas9 mice injected with sgLmnb2-mCherry followed by PBS or AAV-Cre. n = 1 male and 1 female mouse per condition and 25 cells per mouse.
Closed and open circles represent values from male and female mouse, respectively. h, Images of livers from unperturbed postnatal day 12 LSL-Cas9 mice or postnatal day 12 LSL-Cas9 mice injected with $1.25 \times 10^7$ TU of sgMaob-mCherry or sgLmnb2-mCherry lentivirus on postnatal day 1 followed by AAV-Cre on postnatal day 5 immunostained for CD45 (green) and mCherry (magenta) and counterstained for Hoechst (blue). Scale bars, 45 µm. i, Number of CD45-positive cells per 40X field in unperturbed postnatal day 12 LSL-Cas9 mice or postnatal day 12 LSL-Cas9 mice injected with $1.25 \times 10^7$ TU of sgMaob-mCherry or sgLmnb2-mCherry lentivirus on postnatal day 1 followed by AAV-Cre on postnatal day 5 as determined by CD45 immunostaining. n = 2 male and 2 female mice per condition and five fields per mouse. Closed and open circles represent values from male and female mice, respectively. j, Images of livers from unperturbed postnatal day 12 LSL-Cas9 mice or postnatal day 12 LSL-Cas9 mice injected with $1.25 \times 10^7$ TU of sgMaob-mCherry or sgLmnb2-mCherry lentivirus on postnatal day 1 followed by AAV-Cre on postnatal day 5 immunostained for Ki67 (white), mCherry (magenta) and actin (green) and counterstained for Hoechst (blue). Scale bars, 45 µm. k, Percent Ki67-positive hepatocytes in livers from unperturbed postnatal day 12 LSL-Cas9 mice or postnatal day 12 LSL-Cas9 mice injected with $1.25 \times 10^7$ TU of sgMaob-mCherry or sgLmnb2-mCherry lentivirus on postnatal day 1 followed by AAV-Cre on postnatal day 5 as determined by Ki67 immunostaining. n = 2 male and 2 female mice per condition and 50 cells per mouse. Closed and open circles represent values from male and female mice, respectively.
Extended Data Fig. 2 | A genome-wide screen for hepatocyte fitness in the neonatal mouse liver. a, Time points during liver growth, quiescence, and regeneration from which livers were harvested for RNA sequencing. Partial hepatectomy and carbon tetrachloride were used as surgical resection and toxic injury models, respectively. b, Average RNA expression (FPKM) of representative protein-coding genes at different
time points of liver growth, quiescence, and regeneration. Dashed line indicates FPKM cutoff of 0.3. n = 3 male mice per time point c, Number of sgRNAs with a given representation (log10 RPM) for all sgRNAs in the library. d, Pearson correlation (r) for each plot in Fig. 2d. e, Average RNA expression at postnatal day 15 (log10 FPKM) relative to median fold change across mice (log2) for each gene. Spearman $\rho = -0.15$. f, Average protein half-life (log10 hours) relative to median fold change across mice (log2) for each gene. Spearman $\rho = 0.01$. 
Extended Data Fig. 3 | An *in vivo* screen uncovers sex-dependent fitness effects and canonical tumor suppressor genes. Western blot showing Cxorf38 protein levels in Cas9-AML12 cells stably transduced with sgAAVS1, sgTap1, or sgCxorf38. A nonspecific band is shown as a loading control.
Extended Data Fig. 4 | Class I MHC is essential for hepatocyte fitness. a, Quantile-normalized median fold change (log2) of genes in the KEGG gene set for protein export in ES cell screens, HCC cell line screens, and our screen. Genes depleted in more than one screen are highlighted in blue. The bounds of the box indicate the first and third quartiles, and the whiskers extend to the furthest data point that is within 1.5 times the interquartile range. b, Quantile-normalized median fold change (log2) for genes in the KEGG gene set for SNARE interactions in vesicular transport in ES cell screens, HCC
cell line screens, and our screen. Genes depleted in more than one screen are highlighted in blue. The bounds of the box indicate the first and third quartiles, and the whiskers extend to the furthest data point that is within 1.5 times the interquartile range. c, Median fold change (log2) across mice for genes associated with class I MHC or class II MHC within the KEGG gene set for antigen processing and presentation in our screen. The bounds of the box indicate the first and third quartiles, and the whiskers extend to the furthest data point that is within 1.5 times the interquartile range. d, Western blot showing Tap1 protein levels in Cas9-AML12 cells stably transduced with sgAAVS1 or sgTap1. A nonspecific band is shown as a loading control. e, Percent of NK1.1-positive NK cells among CD3-negative non-parenchymal cells in livers of mice without (-) or with (+) intraperitoneal injections of anti-NK1.1 every three days from postnatal day 5 to postnatal day 26.