Fragmentation of Small-cell Lung Cancer Regulatory States in Heterotypic Microenvironments

Dylan L. Schaff[1], Shambhavi Singh[1], Kee-Beom Kim[2], Matthew D. Sutcliffe[1], Kwon-Sik Park[2], and Kevin A. Janes[1,3]

[1]Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia.
[2]Department of Microbiology, Immunology & Cancer Biology, University of Virginia, Charlottesville, Virginia.
[3]Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, Virginia.

D.L. Schaff and S. Singh contributed equally to this article.

Corresponding Author: Kevin A. Janes, University of Virginia, Box 800759 Health System, Charlottesville, VA 22908. Phone: 434-243-2126; Fax: 434-982-3870; E-mail: kjanes@virginia.edu

The authors declare no potential conflicts of interest.

Running title: Small-cell Lung Cancer Heterogeneity In Vitro and In Vivo

Abstract

Small-cell lung cancers derive from pulmonary neuroendocrine cells, which have stem-like properties to reprogram into other cell types upon lung injury. It is difficult to uncouple the plasticity of these transformed cells from heritable changes that evolve in primary tumors or select in metastases to distant organs. Approaches to single-cell profiling are also problematic if the required sample dissociation activates injury-like signaling and reprogramming. Here, we defined cell-state heterogeneities in situ through laser capture microdissection-based 10-cell transcriptomics coupled with stochastic-profiling fluctuation analysis. Using labeled cells from a small-cell lung cancer mouse model initiated by neuroendocrine deletion of p53 and Rb, we profiled cell-to-cell transcriptional-regulatory heterogeneity in spheroid cultures and liver colonies seeded intravenously. Fluctuating transcripts in vitro were partly shared with other epithelial-spheroid models, and candidate heterogeneities increased considerably when cells were delivered to the liver. Colonization of immunocompromised animals drove the fractional appearance of alveolar type II-like markers and poised cells for paracrine stimulation from immune cells and hepatocytes. Immunocompetency further exaggerated the fragmentation of tumor states in the liver, yielding mixed stromal signatures evident in bulk sequencing from autochthonous tumors and metastases. We identified dozens of transcript heterogeneities that recur irrespective of biological context; their mapped orthologs brought together observations of murine and human small-cell lung cancer. Candidate heterogeneities recurrent in the liver also stratified primary human tumors into discrete groups not readily explained by molecular subtype. We conclude that heterotypic interactions in the liver and lung are an accelerant for intratumor heterogeneity in small-cell lung cancer.

Statement of significance

The single-cell regulatory heterogeneity of small-cell lung cancer becomes increasingly elaborate in the liver, a common metastatic site for the disease.

Introduction

The categories, origins, and organization of tumor cell-to-cell heterogeneity are open questions of fundamental importance to cancer biology (1). Within normal tissues, single cells differ by lineage type and regulatory state (2). These distinctions blur, however, when cells lose their proper context because of tissue damage (3,4), transformation (5-7), or metastatic colonization (8,9). The details of such adaptive heterogeneity are expected to depend heavily on the originating cell type, the state of the cell when perturbed, and the local microenvironment where the cell resides.

Within the lung, the pulmonary neuroendocrine cell (PNEC) is a rare-but-important cell type that acts as an airway sensor for damaging stimuli (10). PNECs self-organize into 20–30-cell clusters at airway branch points through dynamic rearrangement of cell-cell contacts and reversible state changes suggesting epithelial-to-mesenchymal transition (EMT) (11). The latest evidence supports that certain PNECs have a reservoir of plasticity to convert into other lung cell types during tissue damage (3). A state of chronic wounding characterizes many tumors–metastases (12), and PNECs are the main cell type of origin for small-cell lung cancer (SCLC) (13-15), a deadly form of lung carcinoma.

Regulatory mechanisms of SCLC plasticity are beginning to be dissected through systems-biology approaches (16,17) and genetically-engineered mouse models (GEMMs) (18). Human SCLC requires loss of *RB1* and *TP53* (19,20)—two tumor suppressors that also developmentally restrict pluripotency (21,22). GEMMs with *Rb1*–*Trp53* deleted by intratracheal delivery of Cre-expressing adenovirus (AdCMV-Cre) give rise to murine SCLCs similar to the classic ASCL1-high subtype of human SCLC (23,24). Deletion of additional tumor suppressors can synergize with *Rb1*–*Trp53* loss (25-27). For example, progression is accelerated by compound deletion of the Rb-family member *p130* (28). These GEMMs (13,14) and others (15) were instrumental in defining PNECs as the cell type of origin for SCLC.

Interestingly, phenotypes of the resulting murine tumors depend on the maturation state of PNECs targeted for *Rb1*–*Trp53* deletion. Restricting adenoviral Cre to PNECs positive for Calca causes far fewer SCLCs to develop compared to when Cre expression is driven by a strong cytomegalovirus (CMV) promoter (29). Both tumor models are metastatic, but only the CMV-driven GEMM upregulates the transcription factor Nfib, which promotes widespread chromatin opening (30) and cell-lineage changes in both primary and metastatic sites (29). Many murine SCLC-derived cell lines are admixtures of cells with neuroendocrine and “non-NE” mesenchymal features (31). Other non-NE SCLC subpopulations are maintained by Notch

signaling (18), which may also become activated in normal PNECs during injury-induced reprogramming (3). There might be other triggers of cell-fate heterogeneity to uncover if SCLC regulatory states could be examined at single-cell resolution without injury-like dissociation of cellular context.

In this work, we examined the in situ transcriptomic regulatory heterogeneities of an established murine SCLC culture derived from an *Rb1*$^{F/F}$*; Trp53*$^{F/F}$ animal administered AdCMV-Cre [KP1 cells (32); **Fig. 1A**]. Using GFP-labeled cells, fluorescence-guided laser capture microdissection (LCM), and 10-cell RNA sequencing (10cRNA-seq) (33), we considered three biological contexts: 1) tumor spheroids cultured in vitro and liver colonies in mice 2) lacking or 3) retaining an intact immune system (**Fig. 1B**). KP1 tumorspheres exhibited cell-to-cell regulatory heterogeneities in cell biology, aging, and metabolism that were shared with spheroid cultures of breast epithelia (33,34). Liver colonization gave rise to pronounced cell-state changes suggesting that paracrine signaling from the lung was partially resurrected in the liver. Liver colonies in immunocompetent animals showed an exacerbated breadth of cell fates, with observed alveolar type II (ATII) markers intermingling with many non-NE stromal markers documented in SCLCs (31) and PNECs (3). Intersecting the three datasets yielded core recurrent heterogeneously expressed genes (RHEGs) and an in vivo RHEG set that was shared by all liver colonies but absent in tumorspheres. Core RHEGs from KP1 cells were broadly shared in bulk human SCLC transcriptomes, yet covariations among cases were not discriminating. By contrast, in vivo RHEGs showed weaker overall correlations but clustered human data into discrete groups that were separate from any 10-cell KP1 transcriptomes. The in vivo RHEGs defined here may reflect a set of injury-like SCLC adaptations that are possible during tumor growth and metastasis at different organ sites.

Materials and Methods

Cell and tissue sources

KP1 cells [generated by K.S.P. (32)] were cultured as self-aggregating spheroids in RPMI medium 1640 (Gibco) with 10% FBS, 1% penicillin-streptomycin, and 1% glutamine. There was no cell-line authentication, and cells were not tested for mycoplasma contamination. KP1-GFP cells were prepared by transducing cells overnight with saturating lentivirus and 8 µg/ml polybrene as previously described (35). GFP-encoding lentivirus was prepared with pLX302 EGFP-V5 cloned by LR recombination of pLX302 (Addgene #25896) and pDONR221_EGFP (Addgene #25899). Stable transductants were selected with 2 µg/ml puromycin until control plates had cleared. Cultured KP1-GFP spheroids were kept to within 10 passages and cryoembedded as described previously (33).

To seed liver colonies, KP1-GFP spheroids were dissociated with 0.05% Trypsin/EDTA (Life Technologies), counted using a hemocytometer, and $2x10^5$ cells were injected via the tail vein of athymic nude (Envigo) or C57/B6 x 129S F_1 hybrid strain of mice (Jackson laboratory). Animals were not randomized. Liver colonies were resected after ~30 days and immediately cryoembedded in NEG-50, frozen in a dry ice-isopentane bath, and stored at –80ºC (33). KP1 spheroids were cryosectioned at –24ºC and liver colonies were cryosectioned at –20ºC, both at 8 µm thickness as previously described (33). All mice were maintained according to practices prescribed by the National Institutes of Health in accordance with the IACUC protocol #9367. All animal procedures were approved by the Animal Care and Use Committee at the University of Virginia, accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Fluorescence-guided LCM

KP1-GFP sections were fixed and dehydrated with ethanol and xylene as described previously for fluorescent cryosections (33). Freshly fixed samples were immediately microdissected on an Arcturus XT LCM instrument (Applied Biosystems) using Capsure HS caps (Arcturus). The smallest spot size on the instrument captured 3–5 SCLC cells per laser shot.

RNA extraction and amplification

RNA extraction and amplification of microdissected samples was performed as described previously to minimize contaminating genomic amplification (33). Briefly, biotinylated

cDNA was synthesized from RNA eluted from captured cells and purified with streptavidin magnetic beads (Pierce) on a 96 S Super Magnet Plate (Alpaqua). Residual RNA was degraded with RNAse H (NEB), and cDNA was poly(A) tailed with terminal transferase (Roche). Poly(A)-cDNA was amplified using AL1 primer (ATTGGATCCAGGCCGCTCTGGACAAAATATGAATTCTTTTTTTTTTTTTTTTTTTTTTTT) and a blend of Taq polymerase (NEB) and Phusion (NEB) for 25 cycles. RNA from bulk KP cell lines was isolated by RNEasy kit (Qiagen).

10-cell sample selection by quantitative PCR (qPCR)

Detection of transcripts by qPCR was performed on a CFX96 real-time PCR instrument (Bio-Rad) as described previously (36). 0.1 µl of preamplification material was used in the qPCR reaction. For each sample, we quantified the expression of *Gapdh* and *Rpl30* as loading controls. Samples were retained if geometric mean quantification cycle of *Gapdh–Rpl30* was within 3.5x interquartile range of the median; samples outside that range were excluded because of over- or under-capture during LCM. For liver colonies, we also excluded samples with detectable quantification cycles of three high-abundance hepatocyte markers: *Alb*, *Fgb*, and *Cyp3a11*.

Library preparation

Ten-cell sequencing libraries were prepared by reamplification, purification, and tagmentation as described previously (33). Briefly, each poly(A) PCR cDNA sample was re-amplified by PCR within its exponential phase (typically 10 to 20 cycles). Re-amplified cDNA was then twice purified with Ampure Agencourt XP SPRI beads, and samples were quantified on a CFX96 real-time PCR instrument (Bio-Rad) using a Qubit BR Assay Kit (Thermo Fisher). Samples were diluted to 0.2 ng/µl and tagmented with the Nextera XT DNA Library Preparation Kit (Illumina). Bulk KP libraries were prepared from 500 ng of total RNA by the Genome Analysis and Technology Core at the University of Virginia using mRNA oligo dT-purified with the NEB Next Ultra RNA library preparation kit (NEB).

RNA sequencing

10cRNA-seq data were sequenced and aligned as previously described (33). Ten-cell samples were multiplexed at an equimolar ratio, and 1.3 pM of the multiplexed pool was sequenced on a NextSeq 500 instrument with NextSeq 500/550 Mid/high Output v1/v2/v2.5 kits (Illumina) to obtain 75-bp paired-end reads. Bulk KP RNA samples were sequenced on a

NextSeq 500 to obtain 50-bp single-end reads. Adapters were trimmed using fastq-mcf in the EAutils package (version ea-utils.1.1.2-779) with the following options: -q 10 -t 0.01 -k 0 (quality threshold 10, 0.01% occurrence frequency, no nucleotide skew causing cycle removal). Quality checks were performed with FastQC (version 0.11.8) and multiqc (version 1.7). Mouse datasets were aligned to the mouse transcriptome (GRCm38.82), reference sequences for ERCC spike-ins, and pLX302-EGFP by using RSEM (version 1.3.0) and Bowtie 2 (version 2.3.4.3). RSEM processing of the 10cRNA-seq data also included the following options: --single-cell-prior --paired-end. Counts from RSEM processing were converted to transcripts per million (TPM) by dividing each value by the total read count for each sample and multiplying by 10^6. Total read count for TPM normalization did not include mitochondrial genes or ERCC spike-ins.

RNA FISH

A 150-bp fragment of human *SOX4* was cloned into pcDNA3, used as a template for in vitro transcription of a digoxigenin-labeled riboprobe for RNA FISH, and imaged as previously described (34). Loading-control riboprobes for *GAPDH*, *HINT1*, and *PRDX6* were previously reported (35).

Immunohistochemistry

Staining was performed by the Biorepository and Tissue Research Facility at the University of Virginia with 4 µm paraffin sections. For F4/80, antigen retrieval and deparaffinization were performed in PT Link (Dako) using low pH EnVision FLEX Target Retrieval Solution (Dako) for 20 minutes at 97°C. Staining was performed on a robotic platform (Autostainer, Dako). Endogenous peroxidases were blocked with peroxidase and alkaline phosphatase blocking reagent (Dako) before incubating the sections with F4/80 antibody (AbD Serotech, #MCA497R) at 1:200 dilution for 60 minutes at room temperature. Antigen–antibody complex was detected by using rabbit anti-rat biotin and streptavidin-HRP (Vector Laboratories) followed by incubation with 3,3'-diaminobenzidine tetrahydrochloride (DAB+) chromogen (Dako). For Cd3, sections were deparaffinized using EZ Prep solution (Ventana), and staining was performed on a robotic platform (Ventana Discover Ultra Staining Module). A heat-induced antigen retrieval protocol set for 64 min was carried out using Cell Conditioner 1 (Ventana). Endogenous peroxidases were blocked with peroxidase inhibitor (CM1) for 8 minutes before incubating the section with CD3 antibody (Dako, #A0452) at 1:300 dilution for 60 minutes at room temperature. Antigen-antibody complex was detected using DISCOVERY OmniMap anti-rabbit multimer RUO detection system and DISCOVERY ChromoMap DAB Kit (Ventana). All

slides were counterstained with hematoxylin, dehydrated, cleared, and mounted for assessment. For both F4/80 and Cd3 stains, cells were counted visually and reported as the average of multiple 10x-field images surrounding individual KP1 cell colonies in liver sections obtained from athymic nude and C57/B6 x 129S F_1 hybrid mice.

Immunoblot analysis

Quantitative immunoblotting was performed as previously described (37). Primary antibodies recognizing the following proteins or epitopes were used: Notch2 (Cell Signaling #5732, 1:1000), vinculin (Millipore #05-386, 1:10,000), GAPDH (Ambion #AM4300, 1:20,000), tubulin (Abcam #ab89984, 1:20,000).

Mouse-to-human ortholog mapping

Human orthologs for mouse genes were obtained from the Ensembl biomart in R using the getAttributes function. For genes with multiple human-ortholog mappings, we used expression characteristics of the human datasets considered [MCF10A-5E (33) and human SCLC (38)] to favor more-reliable clustering afterwards. For mouse genes with two human mappings, the human ortholog with higher expression variance in the corresponding human dataset was retained. For mouse genes with greater than two human mappings, two orthologs with the highest expression correlation were identified. From these, the ortholog with the higher expression variance was retained, as in the two-mapping case. Any remaining mouse gene names were capitalized in accordance with human gene symbol conventions.

Overdispersion-based stochastic profiling

Stochastic profiling with 10cRNA-seq data was performed exactly as described in the accompanying contribution (39).

Robust identification of transcriptional heterogeneities through subsampling

To minimize the contribution of outliers to the overdispersion analysis in samples collected from liver colonies, we generated 100 subsampled simulations for overdispersion-based stochastic profiling. After sample selection (see above), there were 33 10-cell samples plus 35 pool-and-split controls for nude liver colonies and 31 10-cell samples plus 24 pool-and-split controls for C57/B6 x 129S F_1 hybrid liver colonies. For each dataset, overdispersion-based stochastic profiling was performed 100 times with random downsampling to 28 10-cell

samples and 20 pool-and-split controls (39). Only genes that recurred as candidates in >75% of simulations were evaluated further as candidate heterogeneously expressed genes.

Filtering out hepatocyte contamination in heterogeneous expressed genes

Among overdisperse transcripts in liver colonies, we further excluded genes that might vary because of residual hepatocyte capture during LCM. For each 10-cell sample, we calculated the geometric mean abundance of 11 liver-specific markers (*Alb*, *Fgb*, *Cyp3a11*, *Ambp*, *Apoh*, *Hamp*, *Ass1*, *Cyp2f2*, *Glul*, *Hal*, and *Pck1)* from published studies (40-46). Candidates that were significantly correlated with the mean liver signature ($p < 0.05$ by Fisher Z-transformed Spearman ρ correlation) were removed from further consideration for the in vivo study.

Continuous overdispersion analysis

Overdispersion values from the 2007 transcripts identified as candidate heterogeneities in either the KP1 spheres or in vivo conditions were recorded for 100 subsampling iterations. For the KP1 in vitro spheroids, 100 iterations of leave-one-out crossvalidation were performed as detailed in the accompanying contribution (47). Transcripts were retained if the 5th percentile of overdispersion in the C57/B6 x 129S F_1 hybrid condition was greater than the 95th percentiles of the other two conditions. If a gene was not expressed in a condition, the 5th and 95th percentiles were set to zero, and the gene was assigned to the overall median overdispersion during clustering.

Statistics

Sample sizes for stochastic profiling were determined by Monte Carlo simulation (48). Significance of overlap between candidate genes in KP1 spheroids and MCF10A-5E spheroids was evaluated using the hypergeometric test using the “phyper” function in R and a background of 20,000 genes. Pearson correlation between pairs of transcripts detected in both KP1 and MCF10A-5E spheroids were assessed using the “cor.test” function. Significant increases in number of candidate genes between different conditions were assessed by the binomial test using the “binom.test” function in R. Spearman ρ correlation between overdisperse transcripts and liver markers was calculated using the “cor.test” function. Spearman ρ correlations were Fisher Z-transformed using the “FisherZ” function from the R package “DescTools” (version 0.99.31). Co-occurrence of transcript fluctuations was evaluated by hypergeometric test after binning 10cRNA-seq above or below the geometric mean of the two transcripts compared.

Differences in cell number by immunohistochemistry were assessed by the Wilcoxon rank sum test using “wilcox.test”. Significance of overlaps between candidate genes identified in spheroids, nude mice, and C57/B6 x 129S F_1 hybrid mice were assessed by Monte Carlo simulations and corrected for multiple hypothesis testing as described (39). Significance of differences in protein abundance by immunoblotting were assessed by one-way ANOVA with Tukey HSD post-hoc test. Hierarchical clustering was performed using “pheatmap” with standardized values, Euclidean distance, and “ward.D2” linkage or non-standardized values, Pearson distance, and “ward.D2” linkage. Gene set enrichment analyses were performed through the Molecular Signatures Database (49). Overlaps between gene lists and hallmark gene sets were computed using a hypergeometric test with false-discovery rate correction for multiple comparisons.

Data availability

Bulk and 10cRNA-seq data from this study is available through the NCBI Gene Expression Omnibus (GSE147358, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147358 Reviewer token: ufczqoiwzbwjnet). Other RNA-seq datasets were obtained from the Gene Expression Omnibus (GEO): MCF10A-5E 10cRNA-seq (GSE120261), AdCMV-Cre and AdCalca-Cre GEMM (GSE116977), and human SCLC tumor (GSE60052).

Results

Study design and rationale

We sought to define how SCLC regulatory heterogeneity was compiled in different microenvironments. To avoid confounding variation in GEMM tumors that arise autochthonously, we used KP1 cells, a polyclonal *Trp53*$^{\Delta/\Delta}$*Rb1*$^{\Delta/\Delta}$ line derived from a tumor initiated by intratracheal administration of AdCMV-Cre. We sequenced the bulk transcriptome of KP1 cells and found that they were very similar to three other *Trp53*$^{\Delta/\Delta}$*Rb1*$^{\Delta/\Delta}$ lines prepared in similar genetic backgrounds (GSE147358). By contrast, autochthonous SCLC tumors from related GEMMs (29) were different and also more variable among primary tumors, as expected (Supplementary Fig. 1). Before starting, we genetically labeled KP1 cells with EGFP for unambiguous isolation of cells administered in vivo (**Fig. 1A**).

SCLCs frequently metastasize to the liver (50). We mimicked the terminal steps of metastatic colonization and outgrowth by tail-vein injection of EGFP-labeled KP1 cells, which readily establish lesions in the livers of athymic nude mice (**Fig. 1B**). Although KP1 cells have a mixed genetic background, we discovered that subcutaneous xenografts and liver colonies were 100% successful in first-generation crosses (F_1) of C57/B6 and 129S inbred strains. Inoculating C57/B6 x 129S F_1 hybrid animals thus afforded a third setting in which liver colonization and expansion could occur in the presence of a cell-mediated immune response.

KP1 tumorspheres share adaptive transcriptional regulatory heterogeneities with breast-epithelial spheroids

Cultured KP1 cells grow as spheroidal aggregates that can be readily dissociated enzymatically, but juxtacrine cell-cell interactions may contribute to the overall heterogeneity of the population (18). Therefore, we processed KP1 spheroids exactly as if they were tissue, cryoembedding within seconds and sectioning–staining as described (33). Cells were microdissected from the outermost periphery of each spheroid to ensure that all cells profiled had equal availability of nutrients. We gathered 10-cell pools across multiple spheroids to average out subclonal differences within the line and highlight pervasive heterogeneities that characterize spheroid culture. Using 10cRNA-seq (33), we measured the transcriptomes of 28 separate 10-cell groups of KP1 cells along with 20 pool-and-split controls as 10-cell equivalents obtained by LCM. The data were analyzed for candidate regulatory heterogeneities by stochastic profiling (34) implemented with an overdispersion metric optimized for RNA-seq

(39,51,52). The analysis yielded 405 candidate genes that were much more variable in the 10-cell samples than expected given their average abundance and technical reproducibility (**Fig. 2A**).

Samples were collected across multiple days to assess whether batch effects dominated the fluctuation analysis. We clustered gene candidates hierarchically and asked whether the fluctuation signatures clustered according to when the 10-cell samples were collected (**Fig. 2B**). Each grouping was comprised of 10-cell profiles from all batches, supporting that the analytical strategy was robust amidst day-to-day variations in LCM, RNA extraction, and sample amplification. Standard gene set enrichment analysis indicated hallmarks for cell-cycle transitions, Myc–mTORC1 signaling, and metabolism (Supplementary File S1), consistent with the variable growth of spheres in the culture.

Previously, our group used 10cRNA-seq to revisit an earlier analysis of transcriptional regulatory heterogeneity in 3D cultured MCF10A-5E breast-epithelial spheroids (33,34). With an analytical pipeline for stochastic profiling by 10cRNA-seq now in hand (39), we quantified the gene-by-gene overdispersion and identified 1129 candidate heterogeneities (Supplementary Fig. 2A and 2B). The list included multiple transcripts that were independently validated to be heterogeneous by RNA fluorescence in situ hybridization (34,53), including one transcript (*SOX4*) that we validated here (Supplementary Fig. 2B and 2C). The analysis provided a second context for regulatory heterogeneity that exists during spheroidal growth.

Murine SCLC cells and human breast epithelial cells are undoubtedly very different, but normal PNECs derive from an epithelial lineage (11) and often adopt a columnar morphology similar to that seen in the breast. The KP1 study detected significantly fewer genes as regulated heterogeneously compared to MCF10A-5E ($p < 10^{-15}$ by binomial test), corroborating the differences in spheroid culture format. MCF10A-5E cells were 3D cultured in reconstituted basement membrane, which traps secreted factors locally around the spheroids (37), whereas KP1 spheroids develop freely in suspension. Despite differences in the overall number of candidates, we found significant overlap in shared genes after mapping mouse and human orthologs (see Materials and Methods; **Fig. 2C**). Intersecting the two gene groups only marginally enriched for cell-cycling transcripts (54) (six of 57 genes, $p = 0.04$ by hypergeometric test), suggesting other biological processes in addition to proliferation. The intersection raised the possibility that cell growth–competition within epithelial spheroids elicits a set of RHEGs, which generalize beyond a specific culture format.

We next asked whether there might be any common heterogeneities in regulation between the two contexts after correcting for transcript abundance. When the standardized

fluctuations of KP1 and MCF10A-5E spheroids were coclustered by gene ortholog, there were multiple close pairings consistent with shared biology or biological category (**Fig. 2D**). For instance, the importin *KPNA2* covaried with its exportin, *CSE1L* (**Fig. 2E**) (55). We also observed cross-species correlations in genes functioning at the interface of the plasma membrane and endoplasmic reticulum: *ESYT1* and *SPTAN1* (**Fig. 2F**) (56). Although near the detection limit for both cell types, we noted cofluctuations in *SIRT3* and *CTC1*, two factors implicated in cellular longevity (**Fig. 2G**) (57,58). Together, these gene pairings provide a basis for hypotheses about single-cell regulatory pathways that become co-activated when epithelia proliferate outside of their normal polarized context.

Elsewhere among the spheroid RHEGs, we found instances of mutually exclusive transcript heterogeneities, such as with *HUWE1* and *TRIP12* (**Fig. 2H**). These E3 ubiquitin ligases have been reported to operate independently in triggering ubiquitin fusion degradation (59), an unusual proteasomal pathway not studied in cancer. Separately, we recognized a preponderance of metabolic enzymes related to lipids and clustered the fatty acid elongase *ELOVL1*, the β-oxidation dehydrogenase *ACADVL*, and the α-oxidation hydroxylase *PHYH* (**Fig. 2I**). Even with 10-cell pooling, we rarely observed these enzymes abundantly expressed in the same sample, suggesting independent states of lipid synthesis and degradation that could be mined deeply in the future for covariates. Expanding candidate lists around positive and negative covariates has proved powerful in mechanistic follow-on work (37,60).

SCLC reprogramming and paracrine signaling are initiated by colonization of KP1 cells to the liver

To begin examining how heterotypic interactions augment SCLC regulatory heterogeneity, we dissociated KP1 spheroids and colonized the liver of athymic nude mice (**Fig. 1B**). Upon entering the liver circulation, cancer cells extravasate from sinusoids and proliferate amidst hepatocytes. We ensured that SCLC-hepatocyte communication was reflected in the 10cRNA-seq data by sampling the margins of separate GFP+ KP1 liver colonies, analogous to the spheroid margins sampled in vitro (**Fig. 3A**). Focusing on the KP1–hepatocyte interface implied that some level of cell contamination would be introduced by collateral pickup during the LCM step. We rigorously controlled for hepatocyte contamination through a two-step negative-selection procedure. Samples were excluded if hepatocyte markers were abundant by qPCR (Supplementary File S2), and transcripts were removed post-analysis if they covaried with the residual hepatocyte content in the sequenced sample (**Fig. 3B**). Additionally, we oversampled the in vivo samples, collecting 33 10-cell pools that were subsampled 100 times as random

groups of 28 for the dispersion analysis (see Materials and Methods; **Fig. 3B** and Supplementary Fig. S3). The pipeline collectively identified 898 robust candidates fluctuating independently of residual liver markers and appearing in ≥75% of subsampling runs (**Fig. 3C**).

Enriched gene sets were virtually identical to spheroid cultures, except for the addition of STAT5 and interferon γ hallmarks likely resulting from innate immune responses (Supplementary File S1). Beyond the significant increase in candidates ($p < 10^{-15}$ by binomial test), we noted that sample-to-sample fluctuations were qualitatively more dramatic on the margin of liver colonies when compared to KP1 spheroids (**Fig. 2B** and **3C**). Multiple, smaller subsets of candidates were especially interesting. For example, among the robust candidates were the alveolar type II (ATII) markers *Cd74* (61) and *Lyz2* (62), as well as the Cd74 ligand, *Mif*. Surprisingly, when the sample-by-sample fluctuations of these three genes were clustered, we did not detect any significant co-occurrence that would have suggested full transdifferentiation to an ATII phenotype (see Materials and Methods; **Fig. 3D**). The results agree with scRNA-seq data obtained in deprogrammed PNECs (3), where *Cd74* and *Lyz2* markers are anti-correlated among cells with a non-NE phenotype (Supplemental Fig. S4). The patterns detected by stochastic profiling suggest that a subset of KP1 cells reprogram into partial ATII-like states, only one of which senses Mif produced locally (**Fig. 3E)**.

Other groups of transcripts required inputs from non-KP1-derived cell types in the liver to rationalize. Two robust candidates were the NF-κB subunit *Rela* and an NF-κB target gene, *Sod2* (63), which co-occurred strongly ($p < 0.1$) when considering that NF-κB is mostly regulated posttranslationally (**Fig. 3F**). Within the candidate heterogeneities, we also identified the NF-κB-inducing receptor *Ltbr* (64), which varied separately from *Rela–Sod2*. However, the Ltbr ligand (*Ltb*) was effectively absent in KP1 cells (less than 1.5 TPM in bulk samples and pool-and-split controls from liver colonies). We searched Tabula Muris (40) and found that *Ltb* is abundantly expressed in hepatic natural killer (NK) cells, the most-prevalent lymphocyte population in the liver (65). Given that NK cell activity is retained or enhanced in athymic nude mice (66), their paracrine communication with KP1 cells is a plausible mechanism for heterogeneous NF-κB pathway activation in the liver.

We also found evidence for variable regulation in signal transducers of interleukin 1-family cytokines. The inhibitory adaptor *Tollip* (67), the mitogen-activated protein kinase (MAPK) kinase kinase *Map3k7* and its activator *Tab1* (68), the downstream stress-activated MAPKs *Mapk8* (or *Jnk1*) and *Mapk14* (or *p38α*), and the MAPK phosphatase *Dusp8* were all robust transcript heterogeneities in KP1 liver colonies (**Fig. 3C**). Clustering the 10-cell fluctuations of these genes indicated that elevated *Mapk14* levels co-occurred with reduced

abundance of *Mapk8* and *Dusp8* ($1 - p < 0.1$; **Fig. 3G**). Signaling along these parallel MAPK effector pathways may be weighted differently among SCLC cells in the liver colony. Although no relevant receptors were detectably overdisperse in KP1 cells, we consistently detected *Il1rl1*, which is the receptor for Il33 of the interleukin 1 family. PNECs normally receive Il33 stimulation as an alarmin from ATII cells during lung injury or infection (69), but Il33 is also highly expressed in hepatocytes and liver sinusoids (40,70). The widespread single-cell adaptations downstream of Il33 support the hypothesis that SCLC cells redeploy native damage-response pathways in the liver microenvironment.

Immunocompetency exacerbates stromal non-NE phenotypes in SCLC liver colonies

We built upon the results in athymic nude mice by repeating the liver colonization experiments in C57/B6 x 129S F_1 hybrid mice. Compared to KP1 colonies in athymic mice, the C57/B6 x 129S F_1 hybrid colonies had a higher proportion of Cd3+ T cells and a reduced proportion of F4/80+ macrophages along the colony margin (**Fig. 4A–D**), indicating different microenvironments. Stochastic profiling of the KP1 colony margins was performed in C57/B6 x 129S F_1 hybrid livers exactly as for athymic nude animals (**Fig. 3B**). Abundance of liver markers in the C57/B6 x 129S F_1 hybrid samples were as low and uncorrelated as in the nude samples (Supplementary Fig. S5). From 31 10-cell transcriptomic profiles, we robustly identified 1025 regulatory heterogeneities within KP1 cells colonized to an immunocompetent liver (**Fig. 4E**).

Gene set enrichment of the C57/B6 x 129S F_1 hybrid candidates reconstituted most of the hallmarks identified previously along with a moderate signature for hypoxia (Supplementary File S1). In search of shared themes, we compared the KP1 candidate genes from the three biological contexts and found that all two- and three-way intersections were significant ($p < 0.001$ by Monte-Carlo simulation; **Fig. 4F**). This suggested that biological meaning might be embedded in the heterogeneity trends between groups. In lieu of hard overdispersion thresholds (as in **Fig. 2A**), we next analyzed the adjusted variance as a continuous measure of predicted heterogeneity. Beginning with the 2007 transcripts predicted to be heterogeneously regulated in at least one context (**Fig. 4F**), we searched for genes with significant overdispersion increases in the immunocompetent setting (see Materials and Methods). We identified 202 transcripts meeting these criteria, which included multiple neuroendocrine markers (*Rtn2*, *Pcsk1*), *Cd74*, and a new group of stromal transcripts (*Bgn*, *Sparc*, *Mgp*, *Cep19*; **Fig. 4G**) (3). Mesenchymal transitions of SCLC cells can be driven by activated Kras (31), and we noticed that the dispersion of wildtype *Hras* increased alongside the stromal transcripts.

However, when 10-cell fluctuations were clustered, we found that the co-occurrence of *Cd74–Bgn–Sparc* associated with a lack of elevated *Hras* abundance (**Fig. 4H**), excluding a straightforward EMT-like state change. The stromal markers *Mgp* and *Cep19* were also uncoupled from *Cd74–Bgn–Sparc*. We conclude that immunocompetency drives a further diversification of SCLC toward stromal phenotypes in the setting of liver colonization.

Marker gene aberrations are partly retained in autochthonous SCLC tumors and metastases

The non-NE markers identified by stochastic profiling prompted a more-systematic evaluation of marker-gene status in 10-cell and bulk samples. For comparison, we used RNA-seq data from autochthonous tumors and metastases of *Rb1*$^{F/F}$*Trp53*$^{F/F}$*Rbl2*$^{F/F}$ mice administered AdCMV-Cre or adenoviral Cre driven the Calca promoter (AdCalca-Cre) (29). Curiously, for the ATII markers *Cd74* and *Lyz2*, the autochthonous samples indicated that abundance was higher in the primary tumor and reduced in the metastasis (**Fig. 5A** and **5B**, black squares vs. brown filled triangles). Similar results were obtained with the stromal markers, *Bgn* and *Sparc*, although the tumor-metastasis differences were less dramatic (**Fig. 5C** and **5D**). These observations are reconcilable with the 10-cell data if the spheroid observations are not taken as a proxy for the primary tumor. Rather, in vitro cultures reflect the SCLC states achievable from purely homotypic cell-cell interactions. Paracrine inputs from non-NE cells of the lung could just as feasibly drive SCLC reprogramming as non-NE cells of the liver.

Interestingly, abundance of the stromal marker *Mgp* was quite different between the two autochthonous GEMMs (**Fig. 5E**). KP1 cells were isolated from an animal infected with AdCMV-Cre (32). The sporadic increases in *Mgp* abundance observed upon liver colonization were consistent with the other stromal markers found to be high in AdCMV-Cre tumors and metastases. By contrast, *Mgp* abundance in AdCalca-Cre-derived samples was uniformly low. AdCalca-Cre has been speculated to target a more-differentiated subset of PNECs compared to AdCMV-Cre (29). In support of this claim, we found that the KP1 gains in *Mgp* expression in vivo coincided with loss of endogenous *Calca* itself (**Fig. 5E** and **5F**). *Mgp* is an inhibitory morphogen for lung development (71) and its inducibility may mark the PNEC progenitor pool targeted by AdCMV-Cre.

In addition to *Calca*, other neuroendocrine markers (*Ascl1*, *Pcsk1*) declined substantially when KP1 cells were engrafted to the liver (**Fig. 5G** and **5H**). Yet, deprogramming appeared incomplete, as multiple neuroendocrine markers (*Uchl1*, *Resp18*) remained largely unchanged and at comparable abundance to autochthonous models (**Fig. 5I** and **5J**). Among the markers

identified by a gradient of 10-cell dispersion (**Fig. 4G**), several showed no discernible change in median abundance (**Fig. 5K** and **5L**) and thus would be impossible to identify in bulk samples. One of the transcripts correlating strongly with non-NE markers in PNECs (*Ldhb*) (3) recurred as a candidate heterogeneity in all three KP1 settings (**Fig. 5M**). Lastly, we identified a characteristic non-NE marker (*Igfbp7*) (31) where both median abundance and dispersion increased specifically in immunocompetent livers (**Fig. 5N**). Such miscoordination of markers could occur if SCLC cells fragmented their regulatory states upon encountering progressively more-diverse cellular microenvironments.

Mature Notch2 protein abundance is rapidly altered during KP1 cell dissociation

KP1 cells expressed multiple non-NE transcripts in the liver of C57/B6 x 129S F_1 hybrid mice (**Fig. 4G** and **4H**), and single-cell transcriptomics has associated non-NE changes with activation of the Notch pathway (3,18). Unexpectedly, despite measurable expression of *Notch2* by 10cRNA-seq (3.4 ± 9.4 TPM), we almost never detected the Notch target gene *Hes1* in vivo (0 ± 0.1 TPM). For normal PNECs, dedifferentiation to a non-NE state occurs during tissue damage, which may be mimicked by the cell-dissociation steps required for conventional single-cell expression profiling (3). Notch-pathway activation of cell lines also reportedly occurs during routine passaging (72), prompting us to ask whether such artifacts could arise in KP1 cells. Notch1 is nearly absent in the line (less than 0.5 TPM for *Notch1* vs. 29 TPM for *Notch2* in bulk; GSE147358), and reliable activation-specific antibodies for Notch2 are not available. Therefore, we used an antibody recognizing an intracellular epitope of full-length Notch2 and its processed transmembrane (NTM) subunit, which is the precursor for pathway activation (73). Within five minutes of KP1 dissociation using either trypsin or accutase, we noted considerable decreases in total Notch2 protein (full-length + NTM; **Fig. 6A–C**). Furthermore, trypsin significantly increased the ratio of NTM-processed to full-length Notch2 ($p < 0.01$ by ANOVA; **Fig. 6D**), suggesting that trypsinized cells may be more primed to activate Notch signaling. Our results support earlier speculation (3) that Notch activation in PNEC-like cells may be an artifact of the sample processing that precedes scRNA-seq but is avoided by 10cRNA-seq (33).

Human SCLCs are merged or stratified by different classes of KP1 RHEGs

We returned to two statistically significant overlaps from the three studies in KP1 cells (**Fig. 4F**). The three-way intersection of 26 transcripts defined a core group of RHEGs, which we viewed as a set of cell-autonomous heterogeneities intrinsic to KP1 cells and perhaps

SCLCs more generally. We tested this concept by identifying the human orthologs of the KP1 core RHEG set and clustering our data alongside bulk RNA-seq profiles from 79 cases of SCLC in humans (38). The standardized fluctuations of the core RHEGs in human samples were largely indistinguishable from the KP1 observations, with most sample co-clusters containing mouse and human data (**Fig. 7A**). Moreover, when the pairwise correlations of core RHEGs were organized hierarchically, it was difficult to discern any strongly linked groups of observations (**Fig. 7B**). This would be expected if core RHEGs were broadly but independently "active" (induced heterogeneously). Accordingly, we found very little evidence of coordination outside a small row cluster of genes involved in biological processes that were largely unrelated—cell cycle-dependent ubiquitination (*CCNF*), carbonyl stress (*HAGH*), splicing (*SNRNP200*), calcium homeostasis (*CHERP*), and DNA methylation (*MBD1*) (**Fig. 7A**). Although the existence of core RHEGs in mammalian SCLCs awaits direct testing in human samples, the analysis here provides a GEMM-informed set of targets worth examining further.

The second overlap of interest was the two-way intersection of 149 genes that emerged as candidate heterogeneities in both settings of liver colonization (**Fig. 4F**). We defined these in vivo RHEGs as reflecting the SCLC regulatory heterogeneity triggered by heterotypic cell-cell interactions in the microenvironment. In contrast to core RHEGs, we expected different activation patterns of in vivo RHEGs in the liver versus the lung, and even among different SCLC subtypes or primary-tumor sites in the lung. We extracted human orthologs of the in vivo RHEGs and clustered the KP1 observations together with the human SCLCs (**Fig. 7C**). There was far less intermixing between human and KP1 samples, consistent with the different heterotypic interactions anticipated between primary and metastatic sites. The pairwise correlation structure of in vivo RHEGs was also qualitatively distinct, with multiple groups of covariates comprised entirely of human SCLCs (**Fig. 7D**). Importantly, these clusters each contained mixtures of various SCLC subtypes based on the relative abundance of key transcription factors (24) (Supplementary Fig. S6). The KP1 observations in the liver reflect one SCLC subtype and do not precisely capture human variation in the lung. However, the in vivo RHEG set derived from those observations may stratify clinical cases by differences in tumor ecosystem.

Discussion

PNECs are a particularly versatile cell type (3), and it is perhaps unsurprising that derivative SCLC cells show the deranged plasticity reported here. Less obvious is whether dispersed SCLC states are engaged hierarchically or chaotically—our work with a representative GEMM-derived SCLC line argues for the former. Cell-autonomous regulatory heterogeneities expand qualitatively in vivo through heterotypic cell-cell interactions absent from in vitro culture. The documented cell-state changes upon liver colonization could simply reflect the injury-like state of tumors and metastases (12). Alternatively, the reprogramming events could provide trophic support to the cellular ecosystem (7,18). The candidate heterogeneities identified by stochastic profiling and 10cRNA-seq create a resource to guide future functional studies that perturb specific emergent heterogeneities in vivo.

The KP1 results with Notch2 reinforce that SCLC cells are very sensitive to juxtacrine inputs (18). SCLC tumorsphere growth in vitro elicits its own cell-to-cell heterogeneities, which have some commonalities with spheroids of MCF10A-5E basal-like breast cells, a distant epithelial cell type. Intrinsic to spheroid culture are subclonal reorganization and competition, two processes important for primary tumor initiation and the end stages of metastatic colonization. Cell crowding and sequestration alter lipid metabolism (74,75), which could explain the catabolic and anabolic lipid enzymes identified within the spheroid RHEG set. The notion of spheroid RHEGs may generalize to clonogenic soft-agar assays of anchorage-independent growth, which remain widely used as surrogates for tumorigenicity (76).

The candidate regulatory heterogeneities identified in KP1 liver colonies reflect several of the deprogramming and reprogramming events recently described in PNECs (3). In addition, they suggest routes of paracrine communication that are equally realistic for the lung as for the liver. From this perspective, the stratification of primary human SCLCs by in vivo RHEGs is intriguing. SCLCs usually initiate in the bronchi, but there are differences in cell composition at different depths of the lung (77) as well as lobular biases in the primary sites typical for SCLC (78). Different SCLC subtypes (24) might arise in similar microenvironments, yielding the mixed-subtype clusters identified here. The stromal heterogeneities induced by the immunocompetent setting may also relate to fibrotic lung diseases, where PNECs hyperplasia is known to occur (79). The genome of SCLCs is known to be highly mutated (19,20), but our study indicates that cell-fate variability arises on a much faster time scale in vivo.

Figures

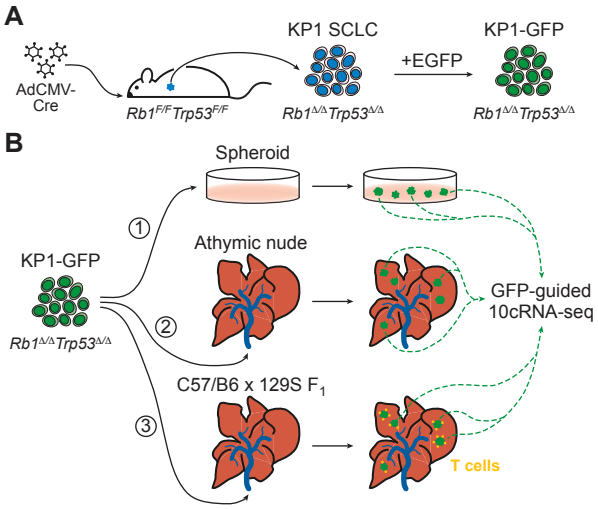


Figure 1. Stochastic profiling of transcriptional regulatory heterogeneity in three isogenic SCLC contexts. **A,** Derivation of KP1 small-cell lung cancer (SCLC) cells by intratracheal administration of adenovirus delivering cytomegalovirus promoter-driven Cre recombinase (AdCMV-Cre) to *Rb1*$^{F/F}$*Trp53*$^{F/F}$ animals. The KP1 SCLC line was engineered to express ectopic enhanced green fluorescent protein (EGFP) for fluorescence-guided microdissection. **B,** The KP1-GFP derivative line was 1) cultured as three-dimensional spheroids in vitro or colonized to the liver of 2) athymic nude mice or 3) C57/B6 x 129S F_1 hybrid mice harboring an intact immune system. GFP-positive cells from multiple spheroids and liver colonies were randomly captured and measured by 10-cell RNA sequencing (10cRNA-seq) for stochastic profiling.

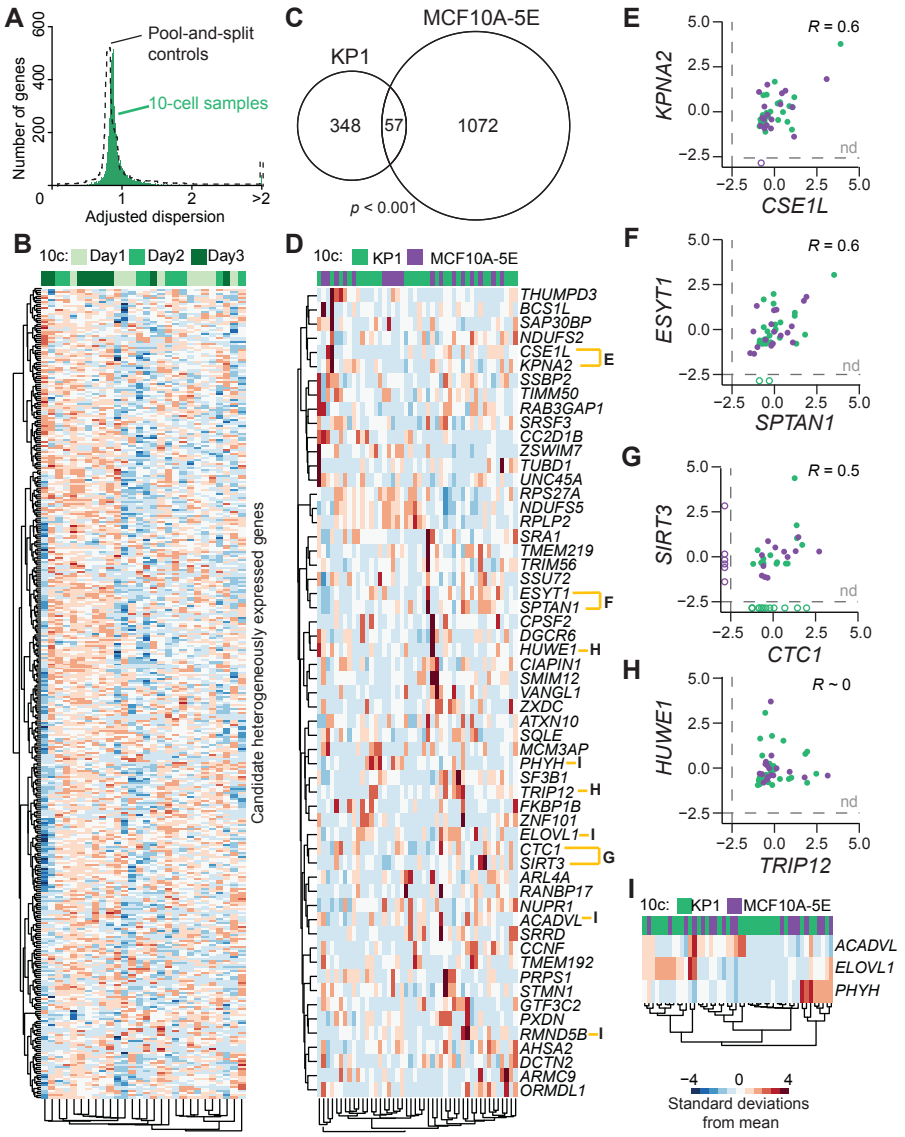


Figure 2. Shared transcriptional regulatory heterogeneities between KP1 spheroids and MCF10A-5E breast-epithelial spheroids. **A,** Overdispersion plot showing 10-cell sample distribution for KP1 spheroids (green) overlaid on pool-and-split controls of 10-cell equivalents (dashed). **B,** Clustergram of the 405 transcripts identified as candidate heterogeneities within KP1 spheroids. Sample acquisition days are annotated. Data were log transformed before standardization. **C,** Venn diagram of orthologous candidates between KP1 spheroids and MCF10A-5E spheroids analyzed in Supplementary Fig. S2. Significance of the intersection was assessed by hypergeometric test with 12,612 total detectable transcripts in KP1 cells and 12,927 total detectable transcripts in MCF10A-5E cells. **D,** Clustergram of the spheroid RHEGs annotated by human ortholog. **E–H,** Pairwise Pearson correlations between the indicated gene pairs in **C** among samples where both genes were detected (filled). nd, not detected. **I,** Clustergram of three transcripts encoding enzymes for lipid metabolism. In **D–I**, 10-cell samples of KP1 cells (green) and MCF10A-5E cells (purple) were standardized separately by z-score before clustering or correlation.

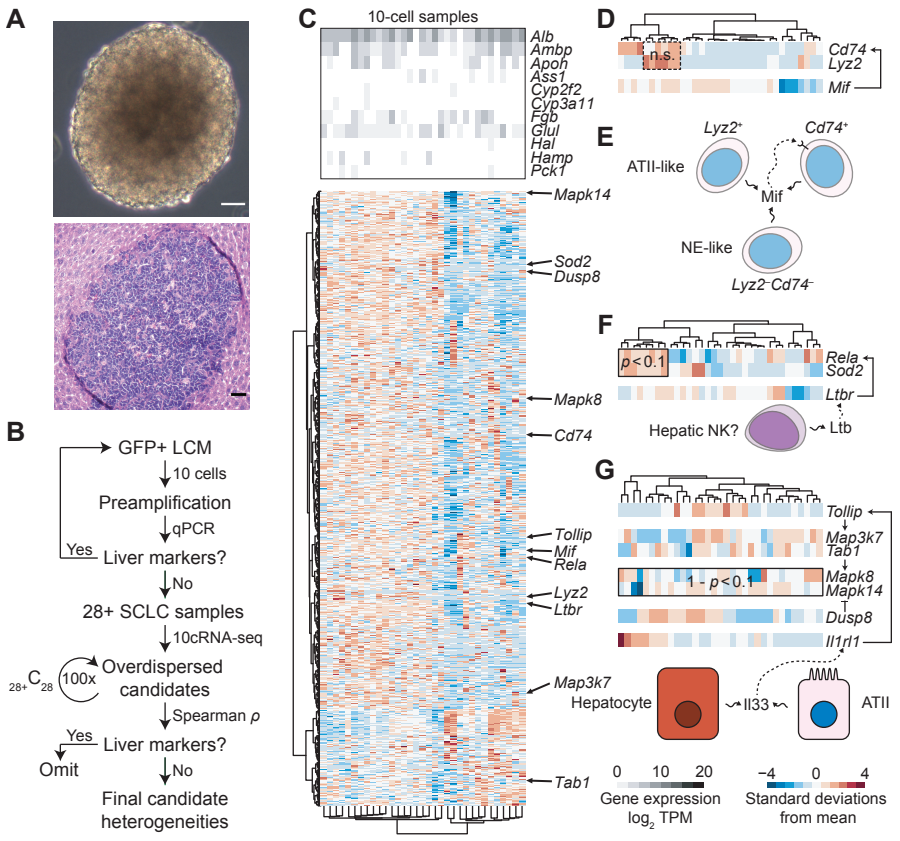


Figure 3. KP1 liver colonization in athymic nude mice causes partial reprogramming and engages heterotypic paracrine-signaling networks. **A,** Phase-contrast image of cultured KP1 spheroids (upper) compared to a brightfield hematoxylin-eosin stain of a KP1 liver colony in an athymic nude mouse (lower). Scale bar is 80 µm. **B,** Flowchart illustrating the experimental and analytical strategy controlling for liver contamination in 10cRNA-seq data and in candidate heterogeneities identified by stochastic profiling. Subsampling results from the 100 dispersion analyses of 28 10-cell samples are shown in Supplementary Fig. S3. **C,** Relative abundance of liver markers (upper) and log-standardized 10-cell sampling fluctuations of robust candidate heterogeneities identified by stochastic profiling (lower). **D,** Log-standardized sampling fluctuations of the alveolar type II (ATII) markers *Cd74* and *Lyz2* together with the Cd74 ligand, *Mif*. **E,** Schematic illustrating the hypothesized relationship between neuroendocrine (NE)- and ATII-like states and Mif signaling. **F,** Log-standardized sampling fluctuations of the *Rela* transcription factor, *Sod2* (a Rela target gene), and *Ltbr* (a Rela-inducing receptor). The ligand for Ltbr is produced by liver-resident NK cells (40). **G,** Log-standardized sampling fluctuations of the indicated signaling transcripts and their pathway relationships downstream of the Il33 receptor, Il1rl1, which is present in KP1 cells but not heterogeneously regulated. Il33 is produced by hepatocytes in the liver and ATII cells in the lung. For **D, F,** and **G,** enriched or unenriched coexpression was evaluated by hypergeometric test of 10-cell observations above their respective log means. n.s., $p > 0.1$ and $1 - p > 0.1$.

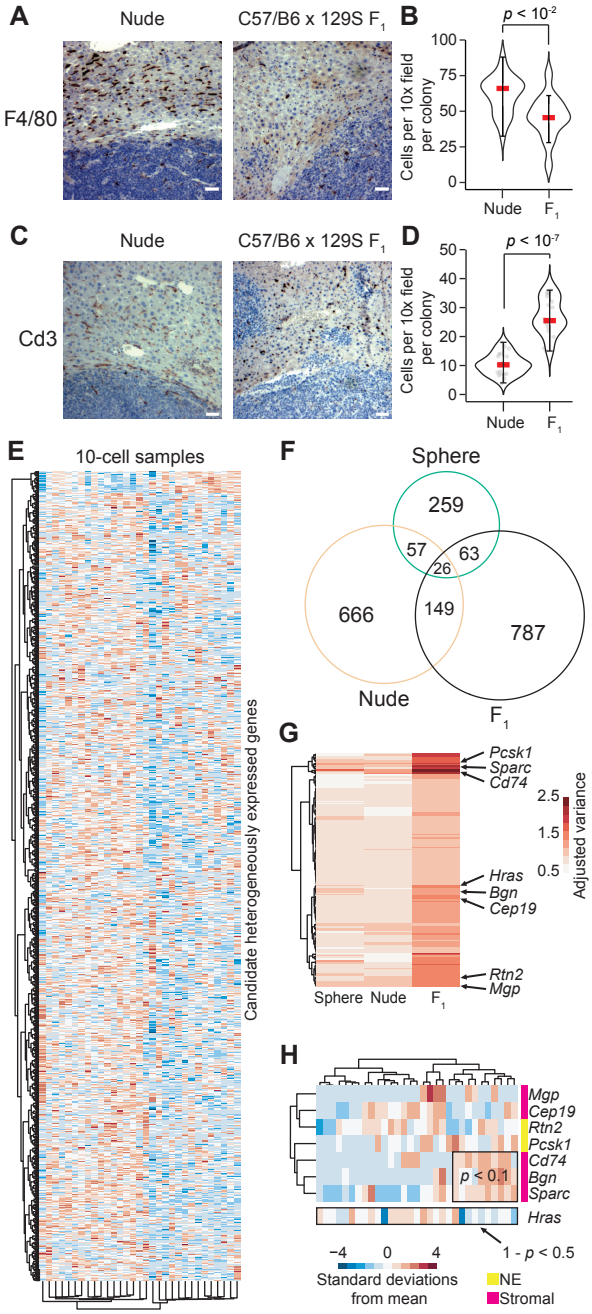

A
Nude
C57/B6 x 129S F1
F4/80
B
p < 10-2
Cells per 10x field per colony
100
75
50
25
0
Nude
F1
C
Nude
C57/B6 x 129S F1
Cd3
D
p < 10-7
Cells per 10x field per colony
50
40
30
20
10
0
Nude
F1
E
10-cell samples
Candidate heterogeneously expressed genes
F
Sphere
259
57
63
26
666
149
787
Nude
F1
G
Pcsk1
Sparc
Cd74
Hras
Bgn
Cep19
Rtn2
Mgp
2.5
1.5
0.5
Adjusted variance
Sphere
Nude
F1
H
Mgp
Cep19
Rtn2
Pcsk1
Cd74
Bgn
Sparc
Hras
p < 0.1
1 - p < 0.5
−4
0
4
Standard deviations from mean
NE
Stromal

Figure 4. Stromal markers emerge heterogeneously when KP1 cells colonize immunocompetent liver. **A,** Immunohistochemistry of athymic nude (left) and C57/B6 x 129S F_1 hybrid (right) livers stained for the macrophage marker F4/80. **B,** Quantification of F4/80+ cells per 10x field is shown as the median of n = 23 nude colonies and n = 14 C57/B6 x 129S F_1 hybrid colonies. **C,** Immunohistochemistry of athymic nude (left) and C57/B6 x 129S F_1 hybrid (right) livers stained for the T cell marker, Cd3. **D,** Quantification of Cd3+ cells per 10x field is shown (right) as the median of n = 24 nude colonies and n = 23 C57/B6 x 129S F_1 hybrid colonies. **E,** Log-standardized 10-cell sampling fluctuations of robust candidate heterogeneities identified by stochastic profiling. **F,** Venn diagram comparing the heterogeneous transcripts identified in **Fig. 2B**, **3C**, and **E**. All two- and three-way intersections were significant ($p < 0.001$ by Monte-Carlo simulation). **G,** Regulatory heterogeneities with abrupt increases in abundance-adjusted variance in C57/B6 x 129S F_1 hybrid liver colonies. Stromal and neuroendocrine (NE) markers are highlighted. **H,** Log-standardized sampling fluctuations for the markers highlighted in **G.** Enriched or unenriched coexpression was evaluated by hypergeometric test of 10-cell observations above their respective log means. For **B** and **D**, differences were assessed by rank-sum test. Scale bar in **A** and **C** is 80 µm.

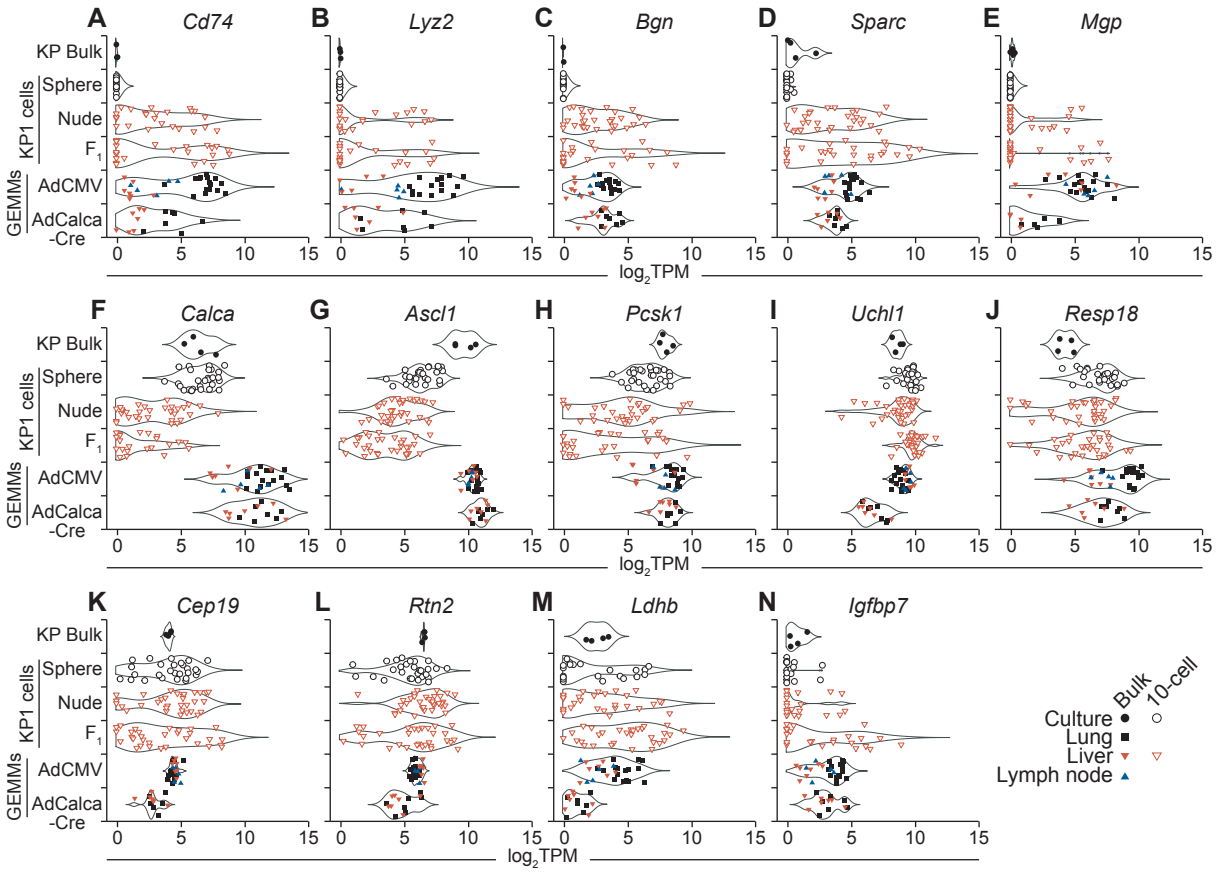


Figure 5. Comparison of 10cRNA-seq observations to bulk RNA-seq from primary SCLC tumors and metastases of SCLC GEMMs. **A** and **B,** Sporadic expression of the ATII markers *Cd74* (**A**) and *Lyz2* (**B**) upon liver colonization of KP1 cells compared to SCLC GEMM tumors and metastases. **C–E,** Abundance changes in the stromal markers *Bgn* (**C**), *Sparc* (**D**), and *Mgp* (**E**) in vivo. **F–J,** Reduced in vivo abundance of the neuroendocrine markers *Calca* (**F**), *Ascl1* (**G**), and *Pcsk1* (**H**), but not *Uchl1* (**I**) or *Resp18* (**J**). **K** and **L,** Context-dependent dispersion changes without abundance changes for the stromal marker *Cep19* (**K**) and the neuroendocrine marker *Rtn2* (**L**). **M,** The non-neuroendocrine marker *Ldhb* is a RHEG in KP1 cells. **N,** Heterogeneous regulation of the non-neuroendocrine marker *Igfbp7* in KP1 cells colonized to the liver of immunocompetent animals.

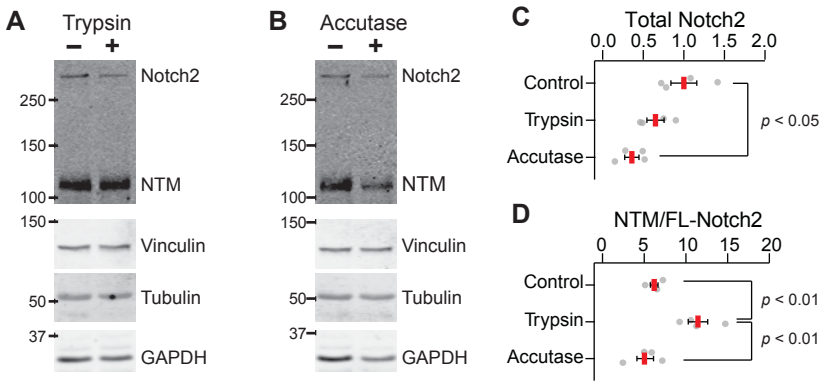


Figure 6. Cell-dissociation enzymes rapidly disrupt intracellular precursors of Notch2 signaling in KP1 cells. **A** and **B,** Immunoblots of full-length Notch2 and the processed Notch transmembrane (NTM) subunit in KP1 cells after treatment of 0.05% trypsin or 1x accutase for five minutes. Vinculin, tubulin, and GAPDH were used as loading controls. **C,** Relative abundance of total Notch2 (full-length + NTM) for the indicated conditions. Data are normalized to control KP1 cells lysed without dissociation. **D,** Ratiometric abundance of NTM / full-length (FL) Notch2 for the indicated conditions. For **C** and **D**, data are shown as the mean ± s.e.m. from n = 4 independent biological samples. Differences in means were assessed by ANOVA with Tukey HSD post-hoc test.

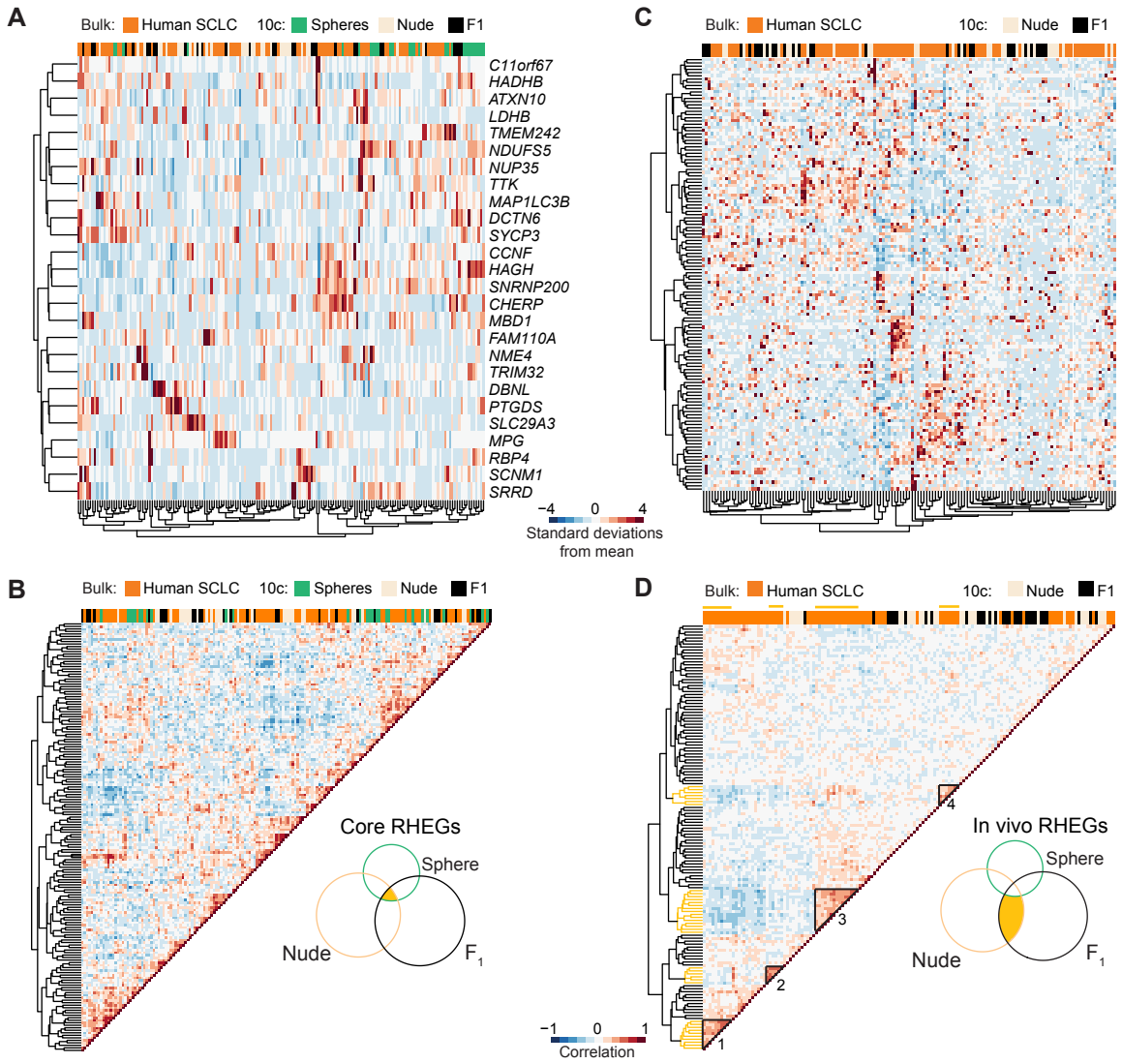
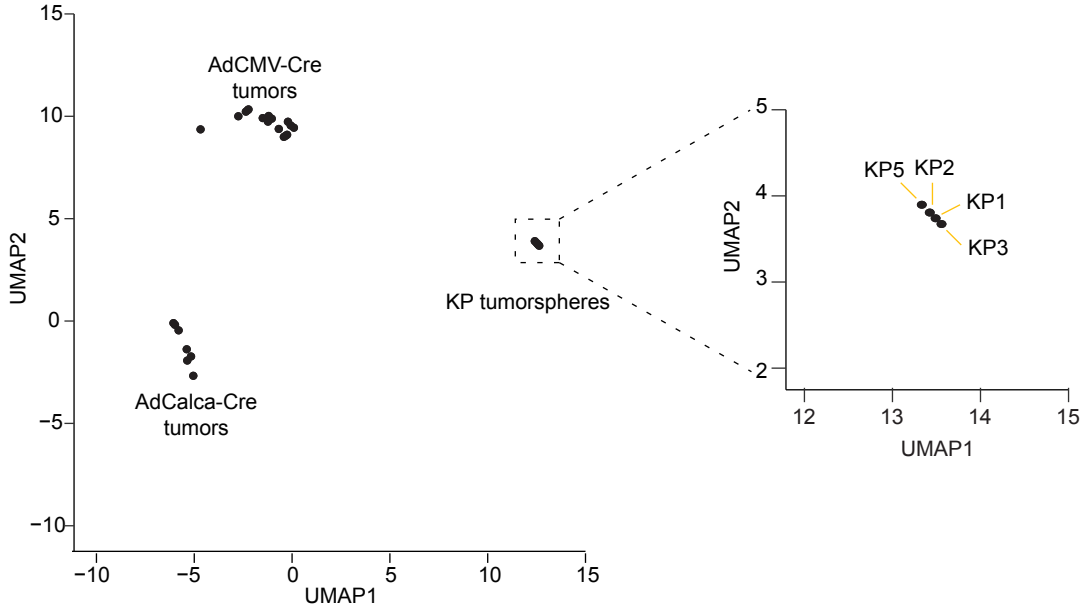
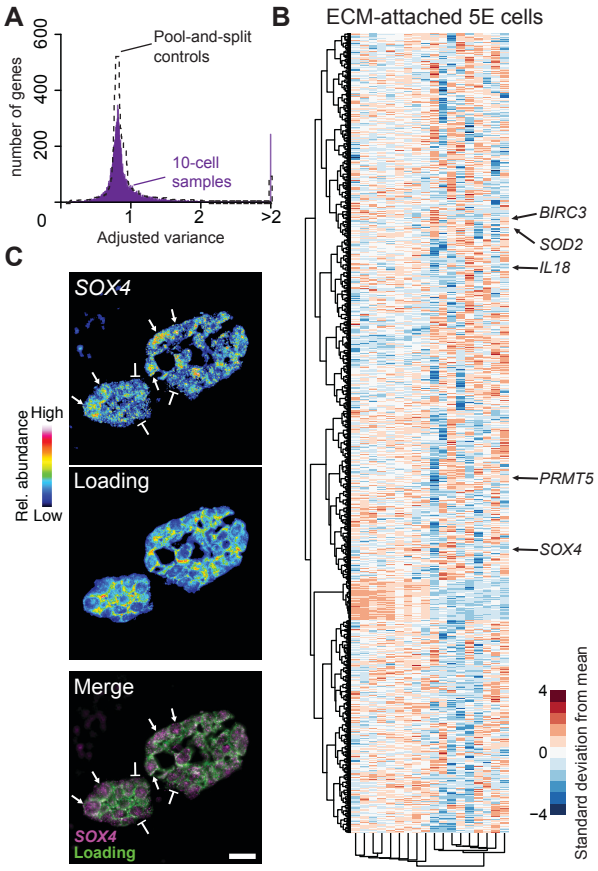


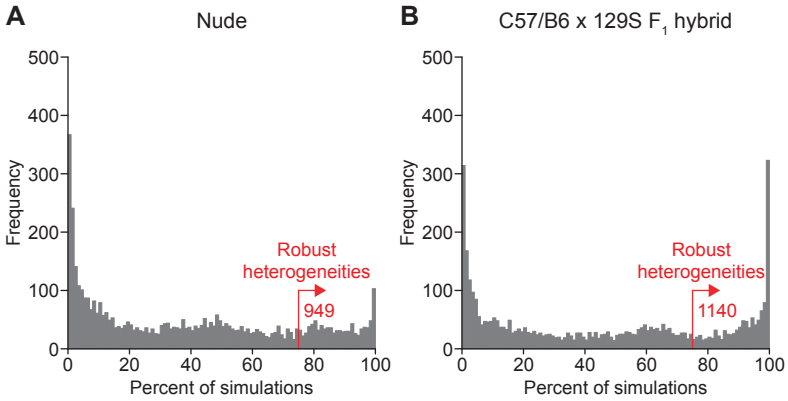
Figure 7. Orthologous RHEG fluctuations in primary human SCLCs. **A,** Core RHEGs and their orthologs intermix KP1 10-cell observations and bulk RNA-seq data from human cases of SCLC (38). **B,** Pearson correlation matrix for core RHEGs clustered hierarchically. The Venn diagram intersection for core RHEGs is highlighted from **Fig. 4F**. **C,** In vivo RHEGs and their orthologs do not merge KP1 and human observations but identify subgroups of clinical SCLCs. **D,** Pearson correlation matrix for in vivo RHEGs clustered hierarchically. Groups of covarying human SCLC cases are indicated in black triangles and yellow margins and numbered as in Supplementary Fig. S6. The Venn diagram intersection for core RHEGs is highlighted from **Fig. 4F**. Murine data and human data were standardized separately by z-score before clustering or correlation.



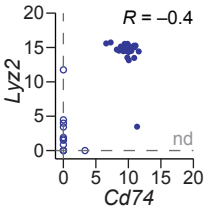
Supplementary Figure S1. KP1 cells are representative of SCLC lines derived from *Rb1*$^{F/F}$*Trp53*$^{F/F}$ mice administered CMV-driven adenoviral Cre (AdCMV-Cre). Autochthonous AdCMV-Cre-initiated and AdCalca-Cre-initiated tumors from *Rb1*$^{F/F}$*Trp53*$^{F/F}$*Rbl2*$^{F/F}$ mice (29) are shown for comparison. Genotypes for the KP SCLC lines are: KP1, *Rb1*$^{\Delta/\Delta}$*Trp53*$^{\Delta/\Delta}$ (32); KP2, *Rb1*$^{\Delta/\Delta}$*Trp53*$^{\Delta/\Delta}$*Ptch1*$^{+/Lacz}$ (32); KP3, *Rb1*$^{\Delta/\Delta}$*Trp53*$^{\Delta/\Delta}$*Axin2*$^{+/Lacz}$ (80); KP5, *Rb1*$^{\Delta/\Delta}$*Trp53*$^{\Delta/\Delta}$*Tg*$^{BAT\text{-}lacZ}$ (81).



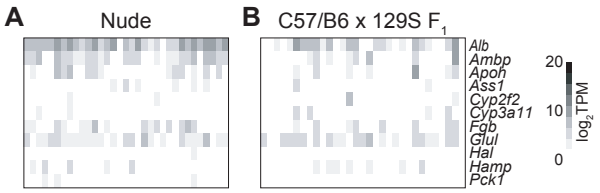
Supplementary Figure S2. Stochastic profiling of transcriptional regulatory heterogeneities in MCF10A-5E spheroids by 10cRNA-seq. **A,** Overdispersion plot showing 10-cell sample distribution for MCF10A-5E spheroids (purple) overlaid on pool-and-split controls of 10-cell equivalents of KP1 spheroids (dashed). **B,** Clustergram of the 1129 transcripts identified as candidate heterogeneities within MCF10A-5E spheroids. Data were log transformed before standardization. Indicated transcripts were independently validated as heterogeneous by RNA FISH (34,53). **C,** RNA FISH validation of *SOX4*, a gene candidate identified by 10cRNA-seq stochastic profiling. Pseudocolor image for *SOX4* is shown above a loading control hybridization comprised of *GAPDH*, *HINT1*, and *PRDX6* (35). Scale bar is 20 µm.



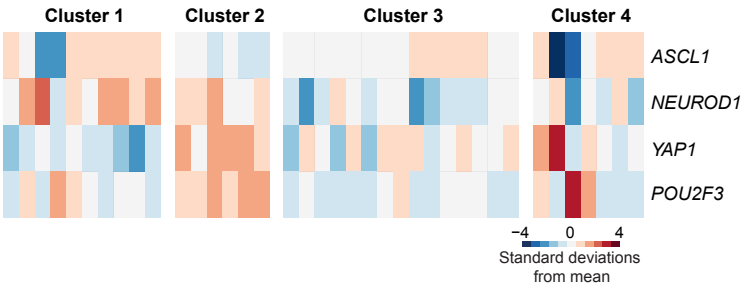
Supplementary Figure S3. Subsampling identifies robust transcriptional regulatory heterogeneities within KP1 liver colonies. **A,** Subsampled dispersion analysis of 33 10-cell observations of KP1 cells colonized to the liver of athymic nude mice. **B,** Subsampled dispersion analysis of 31 10-cell observations of KP1 cells colonized to the liver of immunocompetent C57/B6 x 129S F_1 hybrid mice. Datasets were randomly downsampled to 28 10-cell observations and analyzed for overdispersion as described (39), and candidates appearing in >75% of subsampling runs were considered robust heterogeneities.



Supplementary Figure S4. *Cd74* and *Lyz2* are anti-correlated in single PNEC-derived non-NE cells. scRNA-seq reads of *Lyz2* and *Cd74* are shown for non-NE cells (3). Significance of the Pearson correlation (*R*) was tested after Fisher Z transformation (one-sided $p < 0.05$).



Supplementary Figure S5. Liver contamination in KP1 samples from C57/B6 x 129S F_1 hybrid liver colonies is low and uncorrelated as in athymic nude liver colonies. **A,** Relative abundance of liver markers in 10-cell samples from nude liver colonies, reprinted from **Fig. 3C** for comparison. **B,** Relative abundance of liver markers in 10-cell samples from C57/B6 x 129S F_1 hybrid liver colonies, column clustered as in **Fig. 4E**.



Supplementary Figure S6. In vivo RHEG clusters of human SCLC are not entirely explained by known SCLC subtypes. The SCLC-A, SCLC-N, SCLC-Y, and SCLC-P subtypes are based on the above transcription factor abundances (24). Cluster numbers are as in **Fig. 7D**.

Acknowledgments

We thank Emily Farber and Suna Onengut-Gumuscu at the UVA Center for Public Health Genomics for RNA-seq library preparation and sequencing, Craig Rumpel and Patcharin Pramoonjago at the UVA Biorepository and Tissue Research Facility for LCM maintenance and histology services, UVA Research Computing for high-performance computing access and consulting, Henry Pritchard for assistance with NCBI GEO deposition, and Cheryl Borgman for critical evaluation of the manuscript. This work was supported by the National Institutes of Health #R01-CA194470 (K.A.J.), #U01-CA215794 (K.A.J.), #R01-CA194461 (K.S.P.), and #U01-CA224293 (K.S.P.), the David & Lucile Packard Foundation #2009-34710 (K.A.J.), a UVA Cancer Center support grant #P30-CA044579, the UVA Medical Scientist Training Program (S.S.), a Wagner Fellowship (S.S.), and a Harrison Undergraduate Research Award (D.L.S.).

References

1. Suva ML, Tirosh I. Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. Mol Cell **2019**;75:7-12
2. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E*, et al.* The Human Cell Atlas. Elife **2017**;6
3. Ouadah Y, Rojas ER, Riordan DP, Capostagno S, Kuo CS, Krasnow MA. Rare Pulmonary Neuroendocrine Cells Are Stem Cells Regulated by Rb, p53, and Notch. Cell **2019**;179:403-16 e23
4. van Es JH, Sato T, van de Wetering M, Lyubimova A, Yee Nee AN, Gregorieff A*, et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. Nat Cell Biol **2012**;14:1099-104
5. McEvoy J, Flores-Otero J, Zhang J, Nemeth K, Brennan R, Bradley C*, et al.* Coexpression of normally incompatible developmental pathways in retinoblastoma genesis. Cancer Cell **2011**;20:260-75
6. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K*, et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature **2016**;539:309-13
7. Yao M, Ventura PB, Jiang Y, Rodriguez FJ, Wang L, Perry JSA*, et al.* Astrocytic trans-Differentiation Completes a Multicellular Paracrine Feedback Loop Required for Medulloblastoma Tumor Growth. Cell **2020**;180:502-20 e19
8. Nieto MA, Huang RY, Jackson RA, Thiery JP. Emt: 2016. Cell **2016**;166:21-45
9. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J*, et al.* Integrative clinical genomics of metastatic cancer. Nature **2017**;548:297-303
10. Garg A, Sui P, Verheyden JM, Young LR, Sun X. Consider the lung as a sensory organ: A tip from pulmonary neuroendocrine cells. Curr Top Dev Biol **2019**;132:67-89
11. Kuo CS, Krasnow MA. Formation of a Neurosensory Organ by Epithelial Cell Slithering. Cell **2015**;163:394-405
12. Dvorak HF. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N Engl J Med **1986**;315:1650-9
13. Sutherland KD, Proost N, Brouns I, Adriaensen D, Song JY, Berns A. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. Cancer Cell **2011**;19:754-64
14. Park KS, Liang MC, Raiser DM, Zamponi R, Roach RR, Curtis SJ*, et al.* Characterization of the cell of origin for small cell lung cancer. Cell Cycle **2011**;10:2806-15
15. Song H, Yao E, Lin C, Gacayan R, Chen MH, Chuang PT. Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis. Proc Natl Acad Sci U S A **2012**;109:17531-6
16. Udyavar AR, Wooten DJ, Hoeksema M, Bansal M, Califano A, Estrada L*, et al.* Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity. Cancer Res **2017**;77:1063-74
17. Wooten DJ, Groves SM, Tyson DR, Liu Q, Lim JS, Albert R*, et al.* Systems-level network modeling of Small Cell Lung Cancer subtypes identifies master regulators and destabilizers. PLoS Comput Biol **2019**;15:e1007343
18. Lim JS, Ibaseta A, Fischer MM, Cancilla B, O'Young G, Cristea S*, et al.* Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer. Nature **2017**;545:360-4

19. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH*, et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet **2012**;44:1104-10
20. George J, Lim JS, Jang SJ, Cun Y, Ozretic L, Kong G*, et al.* Comprehensive genomic profiles of small cell lung cancer. Nature **2015**;524:47-53
21. Jain AK, Allton K, Iacovino M, Mahen E, Milczarek RJ, Zwaka TP*, et al.* p53 regulates cell cycle and microRNAs to promote differentiation of human embryonic stem cells. PLoS Biol **2012**;10:e1001268
22. Kareta MS, Gorges LL, Hafeez S, Benayoun BA, Marro S, Zmoos AF*, et al.* Inhibition of pluripotency networks by the Rb tumor suppressor restricts reprogramming and tumorigenesis. Cell Stem Cell **2015**;16:39-50
23. Meuwissen R, Linn SC, Linnoila RI, Zevenhoven J, Mooi WJ, Berns A. Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. Cancer Cell **2003**;4:181-9
24. Rudin CM, Poirier JT, Byers LA, Dive C, Dowlati A, George J*, et al.* Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer **2019**;19:289-97
25. Cui M, Augert A, Rongione M, Conkrite K, Parazzoli S, Nikitin AY*, et al.* PTEN is a potent suppressor of small cell lung cancer. Mol Cancer Res **2014**;12:654-9
26. McFadden DG, Papagiannakopoulos T, Taylor-Weiner A, Stewart C, Carter SL, Cibulskis K*, et al.* Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell **2014**;156:1298-311
27. Jia D, Augert A, Kim DW, Eastwood E, Wu N, Ibrahim AH*, et al.* Crebbp Loss Drives Small Cell Lung Cancer and Increases Sensitivity to HDAC Inhibition. Cancer Discov **2018**;8:1422-37
28. Schaffer BE, Park KS, Yiu G, Conklin JF, Lin C, Burkhart DL*, et al.* Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. Cancer Res **2010**;70:3877-83
29. Yang D, Denny SK, Greenside PG, Chaikovsky AC, Brady JJ, Ouadah Y*, et al.* Intertumoral Heterogeneity in SCLC Is Influenced by the Cell Type of Origin. Cancer Discov **2018**;8:1316-31
30. Denny SK, Yang D, Chuang CH, Brady JJ, Lim JS, Gruner BM*, et al.* Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. Cell **2016**;166:328-42
31. Calbo J, van Montfort E, Proost N, van Drunen E, Beverloo HB, Meuwissen R*, et al.* A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. Cancer Cell **2011**;19:244-56
32. Park KS, Martelotto LG, Peifer M, Sos ML, Karnezis AN, Mahjoub MR*, et al.* A crucial requirement for Hedgehog signaling in small cell lung cancer. Nat Med **2011**;17:1504-8
33. Singh S, Wang L, Schaff DL, Sutcliffe MD, Koeppel AF, Kim J*, et al.* In situ 10-cell RNA sequencing in tissue and tumor biopsy samples. Sci Rep **2019**;9:4836
34. Janes KA, Wang CC, Holmberg KJ, Cabral K, Brugge JS. Identifying single-cell molecular programs by stochastic profiling. Nat Methods **2010**;7:311-7
35. Wang L, Brugge JS, Janes KA. Intersection of FOXO- and RUNX1-mediated gene expression programs in single breast epithelial cells during morphogenesis and tumor progression. Proc Natl Acad Sci U S A **2011**;108:E803-12
36. Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA. Common effector processing mediates cell-specific responses to stimuli. Nature **2007**;448:604-8
37. Bajikar SS, Wang CC, Borten MA, Pereira EJ, Atkins KA, Janes KA. Tumor-Suppressor Inactivation of GDF11 Occurs by Precursor Sequestration in Triple-Negative Breast Cancer. Dev Cell **2017**;43:418-35 e13

38. Jiang L, Huang J, Higgs BW, Hu Z, Xiao Z, Yao X*, et al.* Genomic Landscape Survey Identifies SRSF1 as a Key Oncodriver in Small Cell Lung Cancer. PLoS Genet **2016**;12:e1005895
39. Singh S, Sutcliffe MD, Repich K, Atkins KA, Harvey J, Janes KA. Pan-cancer drivers are recurrent transcriptional regulatory heterogeneities in early-stage luminal breast cancer. Cancer Res **2020**:co-submitted
40. Tabula Muris C, Overall c, Logistical c, Organ c, processing, Library p*, et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature **2018**;562:367-72
41. Zhang F, Xu X, Zhou B, He Z, Zhai Q. Gene expression profile change and associated physiological and pathological effects in mouse liver induced by fasting and refeeding. PLoS One **2011**;6:e27553
42. Fish RJ, Neerman-Arbez M. Fibrinogen gene regulation. Thromb Haemost **2012**;108:419-26
43. Yanagimoto T, Itoh S, Sawada M, Kamataki T. Mouse cytochrome P450 (Cyp3a11): predominant expression in liver and capacity to activate aflatoxin B1. Arch Biochem Biophys **1997**;340:215-8
44. Salier JP, Chan P, Raguenez G, Zwingman T, Erickson RP. Developmentally regulated transcription of the four liver-specific genes for inter-alpha-inhibitor family in mouse. Biochem J **1993**;296 (Pt 1):85-91
45. Leduc MS, Shimmin LC, Klos KL, Hanis C, Boerwinkle E, Hixson JE. Comprehensive evaluation of apolipoprotein H gene (APOH) variation identifies novel associations with measures of lipid metabolism in GENOA. J Lipid Res **2008**;49:2648-56
46. Franzen O, Gan LM, Bjorkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database (Oxford) **2019**;2019
47. Sutcliffe MD, Galvao RP, Wang L, Kim J, Singh S, Zong H*, et al.* Single-cell bottlenecks and dead-ends during glioma premalignancy. Cancer Res **2020**:co-submitted
48. Wang L, Janes KA. Stochastic profiling of transcriptional regulatory heterogeneities in tissues, tumors and cultured cells. Nat Protoc **2013**;8:282-301
49. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst **2015**;1:417-25
50. Kagohashi K, Satoh H, Ishikawa H, Ohtsuka M, Sekizawa K. Liver metastasis at the time of initial diagnosis of lung cancer. Med Oncol **2003**;20:25-8
51. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nat Methods **2014**;11:740-2
52. Fan J, Salathia N, Liu R, Kaeser GE, Yung YC, Herman JL*, et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat Methods **2016**;13:241-4
53. Bajikar SS, Fuchs C, Roller A, Theis FJ, Janes KA. Parameterizing cell-to-cell regulatory heterogeneities via stochastic transcriptional profiles. Proc Natl Acad Sci U S A **2014**;111:E626-35
54. Santos A, Wernersson R, Jensen LJ. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. Nucleic Acids Res **2015**;43:D1140-4
55. Kutay U, Bischoff FR, Kostka S, Kraft R, Gorlich D. Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. Cell **1997**;90:1061-71
56. Giordano F, Saheki Y, Idevall-Hagren O, Colombo SF, Pirruccello M, Milosevic I*, et al.* PI(4,5)P(2)-dependent and Ca(2+)-regulated ER-PM interactions mediated by the extended synaptotagmins. Cell **2013**;153:1494-509
57. Chen LY, Redon S, Lingner J. The human CST complex is a terminator of telomerase activity. Nature **2012**;488:540-4

58. Haigis MC, Guarente LP. Mammalian sirtuins--emerging roles in physiology, aging, and calorie restriction. Genes Dev **2006**;20:2913-21
59. Poulsen EG, Steinhauer C, Lees M, Lauridsen AM, Ellgaard L, Hartmann-Petersen R. HUWE1 and TRIP12 collaborate in degradation of ubiquitin-fusion proteins and misframed ubiquitin. PLoS One **2012**;7:e50548
60. Pereira EJ, Burns JS, Lee CY, Marohl T, Calderon D, Wang L*, et al.* Sporadic activation of an oxidative stress-dependent NRF2–p53 signaling network in breast epithelial spheroids and premalignancies. bioRxiv **2019**:862474
61. Marsh LM, Cakarova L, Kwapiszewska G, von Wulffen W, Herold S, Seeger W*, et al.* Surface expression of CD74 by type II alveolar epithelial cells: a potential mechanism for macrophage migration inhibitory factor-induced epithelial repair. Am J Physiol Lung Cell Mol Physiol **2009**;296:L442-52
62. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature **2014**;507:190-4
63. Das KC, Lewis-Molock Y, White CW. Thiol modulation of TNF alpha and IL-1 induced MnSOD gene expression and activation of NF-kappa B. Mol Cell Biochem **1995**;148:45-57
64. Mackay F, Majeau GR, Hochman PS, Browning JL. Lymphotoxin beta receptor triggering induces activation of the nuclear factor kappaB transcription factor in some cell types. J Biol Chem **1996**;271:24934-8
65. Mikulak J, Bruni E, Oriolo F, Di Vito C, Mavilio D. Hepatic Natural Killer Cells: Organ-Specific Sentinels of Liver Immune Homeostasis and Physiopathology. Front Immunol **2019**;10:946
66. Budzynski W, Radzikowski C. Cytotoxic cells in immunodeficient athymic mice. Immunopharmacol Immunotoxicol **1994**;16:319-46
67. Burns K, Clatworthy J, Martin L, Martinon F, Plumpton C, Maschera B*, et al.* Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. Nat Cell Biol **2000**;2:346-51
68. Shibuya H, Yamaguchi K, Shirakabe K, Tonegawa A, Gotoh Y, Ueno N*, et al.* TAB1: an activator of the TAK1 MAPKKK in TGF-beta signal transduction. Science **1996**;272:1179-82
69. Yasuda K, Muto T, Kawagoe T, Matsumoto M, Sasaki Y, Matsushita K*, et al.* Contribution of IL-33-activated type II innate lymphoid cells to pulmonary eosinophilia in intestinal nematode-infected mice. Proc Natl Acad Sci U S A **2012**;109:3451-6
70. Pichery M, Mirey E, Mercier P, Lefrancais E, Dujardin A, Ortega N*, et al.* Endogenous IL-33 is highly expressed in mouse epithelial barrier tissues, lymphoid organs, brain, embryos, and inflamed tissues: in situ analysis using a novel Il-33-LacZ gene trap reporter strain. J Immunol **2012**;188:3488-95
71. Yao Y, Nowak S, Yochelis A, Garfinkel A, Bostrom KI. Matrix GLA protein, an inhibitory morphogen in pulmonary vascular development. J Biol Chem **2007**;282:30131-42
72. Liu W, Morgan KM, Pine SR. Activation of the Notch1 Stem Cell Signaling Pathway during Routine Cell Line Subculture. Front Oncol **2014**;4:211
73. Bray SJ. Notch signalling in context. Nat Rev Mol Cell Biol **2016**;17:722-35
74. Frechin M, Stoeger T, Daetwyler S, Gehin C, Battich N, Damm EM*, et al.* Cell-intrinsic adaptation of lipid composition to local crowding drives social behaviour. Nature **2015**;523:88-91
75. Schafer ZT, Grassian AR, Song L, Jiang Z, Gerhart-Hines Z, Irie HY*, et al.* Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. Nature **2009**;461:109-13

76. Shin SI, Freedman VH, Risser R, Pollack R. Tumorigenicity of virus-transformed cells in nude mice is correlated specifically with anchorage independent growth in vitro. Proc Natl Acad Sci U S A **1975**;72:4435-9
77. Rackley CR, Stripp BR. Building and maintaining the epithelium of the lung. J Clin Invest **2012**;122:2724-30
78. Combs SE, Hancock JG, Boffa DJ, Decker RH, Detterbeck FC, Kim AW. Bolstering the case for lobectomy in stages I, II, and IIIA small-cell lung cancer using the National Cancer Data Base. J Thorac Oncol **2015**;10:316-23
79. Benson RE, Rosado-de-Christenson ML, Martinez-Jimenez S, Kunin JR, Pettavel PP. Spectrum of pulmonary neuroendocrine proliferations and neoplasms. Radiographics **2013**;33:1631-49
80. Ponath P, Menezes D, Pan C, Chen B, Oyasu M, Strachan D*, et al.* A Novel, Fully Human Anti-fucosyl-GM1 Antibody Demonstrates Potent In Vitro and In Vivo Antitumor Activity in Preclinical Models of Small Cell Lung Cancer. Clin Cancer Res **2018**;24:5178-89
81. Maretto S, Cordenonsi M, Dupont S, Braghetta P, Broccoli V, Hassan AB*, et al.* Mapping Wnt/beta-catenin signaling during mouse development and in colorectal tumors. Proc Natl Acad Sci U S A **2003**;100:3299-304