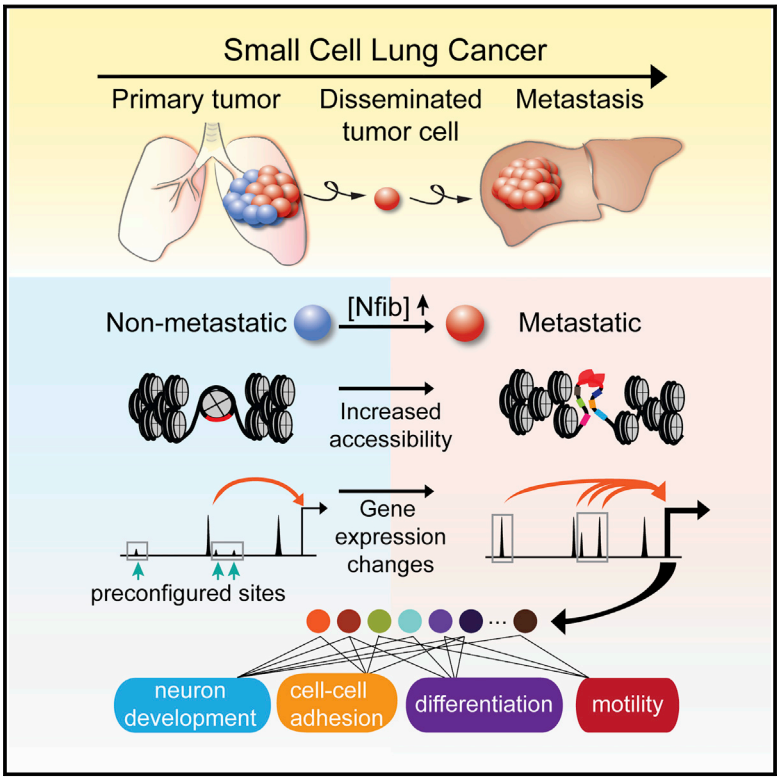
Article

Cell

Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility

Graphical Abstract



Authors

Sarah K. Denny, Dian Yang, Chen-Hua Chuang, ..., Julien Sage, William J. Greenleaf, Monte M. Winslow

Correspondence

wjg@stanford.edu (W.J.G.), mwinslow@stanford.edu (M.M.W.)

In Brief

Global chromatin alterations driven by a key transcriptional regulator promotes metastasis of small cell lung cancer.

Highlights

- SCLC metastases have profound chromatin changes compared to primary tumors
- Nfib drives an increase of chromatin accessibility at distal regulatory elements
- Nfib is necessary and sufficient for metastatic ability in vivo
- Nfib co-opts pro-metastatic neuronal gene expression programs

Accession Numbers

GSE81258

Denny et al., 2016, Cell *166*, 1–15
July 14, 2016
http://dx.doi.org/10.1016/j.cell.2016.05.052

CellPress

Please cite this article in press as: Denny et al., Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility, Cell (2016), http://dx.doi.org/10.1016/j.cell.2016.05.052

Article

Cell

Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility

Sarah K. Denny,[1,10] Dian Yang,[2,10] Chen-Hua Chuang,[3] Jennifer J. Brady,[3] Jing Shan Lim,[2] Barbara M. Grüner,[3] Shin-Heng Chiou,[3] Alicia N. Schep,[3] Jessika Baral,[3] Cécile Hamard,[7] Martine Antoine,[7] Marie Wislez,[7] Christina S. Kong,[4] Andrew J. Connolly,[4] Kwon-Sik Park,[8] Julien Sage,[2,3,5,6,11] William J. Greenleaf,[1,3,9,11,*] and Monte M. Winslow[2,3,4,5,11,*]

[1]Biophysics Program, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Cancer Biology Program, Stanford University School of Medicine, Stanford, CA 94305, USA
[3]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA
[5]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305, USA
[6]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA
[7]Service de Pneumologie, Hôpital Tenon-APHP, Université Paris 6 Pierre et Marie Curie, 75020 Paris, France
[8]Department of Microbiology, Immunology, and Cancer Biology, University of Virginia School of Medicine, Charlottesville, VA 22908, USA
[9]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA
[10]Co-first author
[11]Co-senior author
*Correspondence: wjg@stanford.edu (W.J.G.), mwinslow@stanford.edu (M.M.W.)
http://dx.doi.org/10.1016/j.cell.2016.05.052

SUMMARY

Metastases are the main cause of cancer deaths, but the mechanisms underlying metastatic progression remain poorly understood. We isolated pure populations of cancer cells from primary tumors and metastases from a genetically engineered mouse model of human small cell lung cancer (SCLC) to investigate the mechanisms that drive the metastatic spread of this lethal cancer. Genome-wide characterization of chromatin accessibility revealed the opening of large numbers of distal regulatory elements across the genome during metastatic progression. These changes correlate with copy number amplification of the *Nfib* locus, and differentially accessible sites were highly enriched for Nfib transcription factor binding sites. Nfib is necessary and sufficient to increase chromatin accessibility at a large subset of the intergenic regions. Nfib promotes pro-metastatic neuronal gene expression programs and drives the metastatic ability of SCLC cells. The identification of widespread chromatin changes during SCLC progression reveals an unexpected global reprogramming during metastatic progression.

INTRODUCTION

Phenotypic changes that occur during development and disease progression are driven by gene expression changes that are themselves governed by regulatory states encoded within the nucleoprotein structure of chromatin (Voss and Hager 2014). During development and differentiation, tens of thousands of regulatory elements change from inactive to active states (or vice versa), eliciting a concerted transformation of gene expression programs that control cell phenotypes (Zhu et al., 2013). Numerous targeted methods of probing this landscape, from chromatin immunoprecipitation approaches to assays measuring DNA methylation, have produced insight into the dimensions of this regulation (Schones and Zhao 2008). Chromatin accessibility, or the genome-wide accounting of loci accessible for transcription factor binding, has been identified as perhaps the single most relevant genomic characteristic correlated with biological activity at a specific locus (Thurman et al., 2012).

Recent work has begun to catalog chromatin state changes between normal and cancer cells and to define the chromatin landscape of several cancer cell lines (Simon et al., 2014; Stergachis et al., 2013). The phenotypic changes associated with metastasis likely require widespread changes in gene expression programs that drive invasion, migration, dissemination, and colonization (Sethi and Kang 2011). However, the specific regulatory changes driving the transition of primary tumors to cells capable of metastatic spread remain largely unexplored.

Small cell lung cancer (SCLC) is a high-grade neuroendocrine carcinoma that accounts for ~15% of all lung cancers and causes over 200,000 deaths worldwide each year (Kalemkerian et al., 2013). The ability of SCLC cells to leave the primary tumor and establish inoperable metastases is a major cause of death and a serious impediment to successful therapy (van Meerbeeck et al., 2011). Molecular analysis of metastatic progression of human cancer is limited by the difficulty in accessing tumor samples at defined stages. This problem is especially true for SCLC, since patients with metastatic disease rarely undergo surgery. Genetically engineered mouse models of human SCLC recapitulate the genetics, histology, therapeutic response, and highly metastatic nature of the human disease (Meuwissen et al., 2003; Schaffer et al., 2010). These models recapitulate cancer progression in a controlled manner and allow isolation of primary tumors and metastases directly from their native microenvironment.

Please cite this article in press as: Denny et al., Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility, Cell (2016), http://dx.doi.org/10.1016/j.cell.2016.05.052

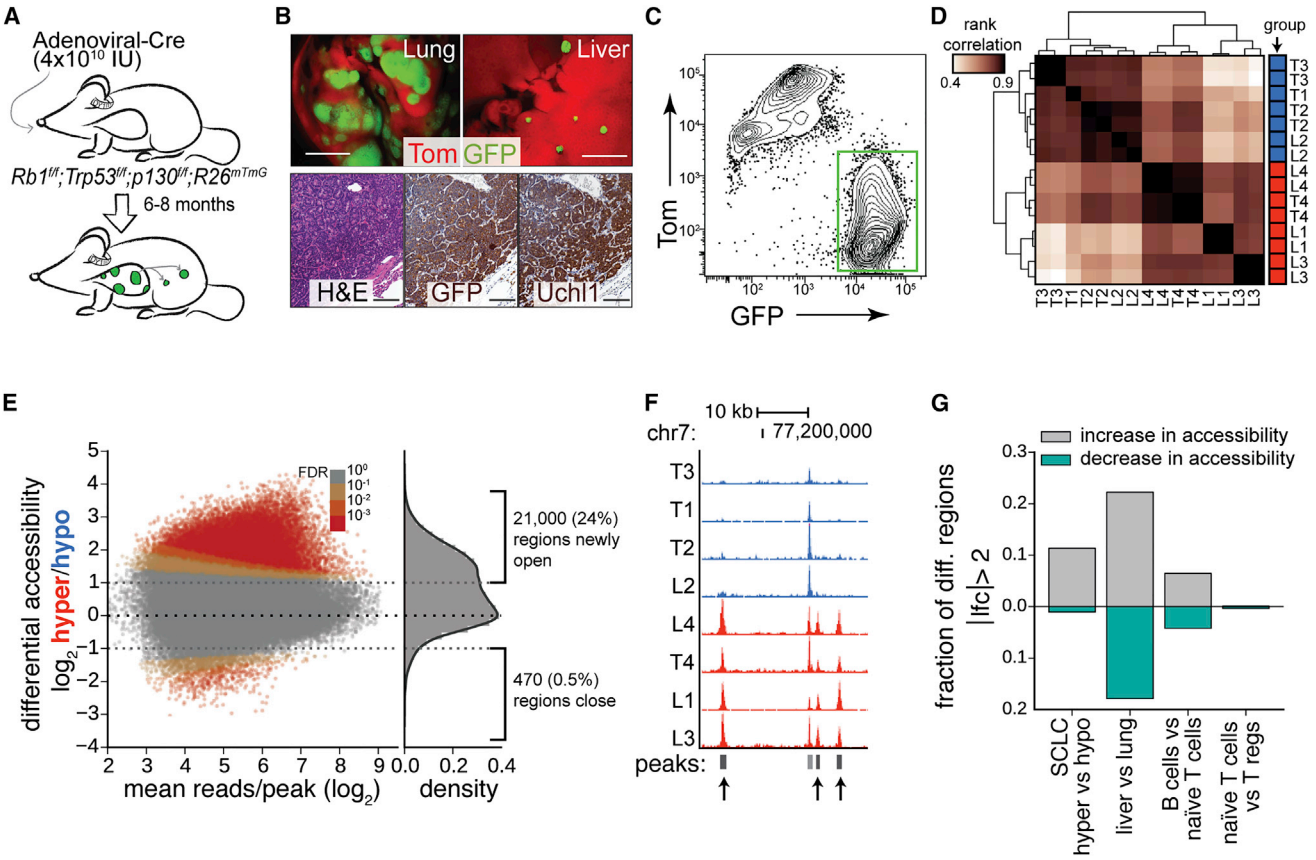


Figure 1. SCLC Exists in Two Distinct Chromatin Accessibility States

(A) Genetically engineered mouse model of SCLC.

(B) GFP^{positive} primary tumors and metastases within Tomato (Tom)^{positive} normal tissues (top). GFP and Uchl1 IHC and H&E are shown. Top scale bars, 5 mm. Bottom scale bars, 100 μm.

(C) Representative FACS plot of a dissociated tumor. FSC/SSC-gated, Lineage^{negative}, and viable (DAPI^{negative}) cells are shown. GFP^{postive}Tom^{negative} SCLC cells are indicated.

(D) Correlation of chromatin accessibility of eight samples by ATAC-seq analysis. T, primary lung tumor; L, liver metastasis.

(E) Differential accessibility (log_2 fold change in reads per accessible region) plotted against the mean reads per region. FDR is the false discovery rate that the absolute value of log_2 fold change is >1.

(F) Insertion tracks of samples at an example locus on chromosome 7. Differentially open regions are marked with arrows.

(G) Fraction of total regions differentially accessible in multiple tissue and cell-type comparisons. lfc, log_2 fold change.

See also Figure S1.

Here, we analyzed SCLC cells from primary tumors and metastases to identify global changes in chromatin accessibility during metastatic progression. We uncovered an unexpectedly dramatic increase in accessibility that occurs during malignant progression. We determined that high expression of a single transcription factor, Nfib, alters chromatin state globally and enacts a program of gene expression that promotes multiple steps of the metastatic cascade.

RESULTS

Identification of Two Distinct Chromatin Accessibility Landscapes within SCLC

To specifically mark cancer cells, we bred a Cre-reporter allele (*R26*^{mTmG}) into the *Trp53*$^{f/f}$;*Rb1*$^{f/f}$;*p130*$^{f/f}$ mouse model of human SCLC (Muzumdar et al., 2007; Schaffer et al., 2010). Adenoviral-Cre inhalation by *Trp53*$^{f/f}$;*Rb1*$^{f/f}$;*p130*$^{f/f}$;*R26*$^{mT/mG}$ (*TKO-mTmG*) mice initiates cancer development. Primary tumors and metastases are Tomato^{negative}, GFP^{positive}, have histologic features of human SCLC, and express markers of SCLC (Figures 1A and 1B) (Schaffer et al., 2010). These mice develop multifocal metastases in the liver, one of the most common sites of SCLC metastasis in humans (Figure 1B) (Nakazawa et al., 2012; Meuwissen et al., 2003). We isolated cells from individual primary tumors and metastases by fluorescence-activated cell sorting (FACS; Figures 1C, S1A, and S1B).

We used the assay of transposase-accessible chromatin using sequencing (ATAC-seq) to determine the genome-wide chromatin accessibility landscape in primary tumors and liver metastases from *TKO-mTmG* mice (Buenrostro et al., 2013).

We isolated cancer cells from one large primary tumor and one liver macro-metastasis from each of four mice. All ATAC-seq samples were enriched for reads at transcription start sites (TSSs) and exhibited the expected periodicity of insert length (Figures S1C–S1E). Hierarchical clustering based on the correlation of accessibility separated the samples into two groups: one containing the majority of primary tumors and the other containing the majority of metastases (Figure 1D). The first principal component of variation also separated the samples into the same two groups and explained 58% of the variance (Figures S1F and S1G).

We assessed differential accessibility between these two groups and discovered that ~24% of all accessible regions were >2-fold more accessible in the predominantly metastatic group (Figures 1E, 1F, S1H, S2A, and S2B). Conversely, only ~0.5% of peaks were >2-fold more accessible in the primary-tumor-enriched group, and thus we defined these chromatin states as ''hyper-accessible'' and ''hypo-accessible.''

To put the magnitude of these chromatin differences in context, we reanalyzed our ATAC-seq data and published DNase-seq data from distinct tissues and cell types (Vierstra et al., 2014). Interestingly, the fraction of peaks that differ between the hypo- and hyper-accessible SCLC cell states is smaller than that detected between lung and liver, but greater than between lymphocyte subsets. The nearly unidirectional increase in accessibility during SCLC progression is also a distinct feature of this cell state change (Figure 1G).

Differentially Accessible Regions Are in Gene Deserts, Evolutionarily Conserved, and Shared by Other Cell Types

The differentially accessible regions were largely gene distal, with relatively few promoter-proximal regions exhibiting differential accessibility (Figures 2A and S2C). Large genomic domains (on the order of 1 to 10 Mb) were more likely than others to have differentially open regions (over 90% of the newly open peaks were in these domains, which contain less than 40% of constitutively open peaks, $p < 1 \times 10^{-12}$ by Fisher's exact test) (Figure 2B). Domains enriched for differentially accessible peaks were characterized by significantly lower gene density and late replication timing (Figures 2C, 2D, and S2D). Gene-poor domains have been classically thought to be less important in regulating gene expression. However, the differentially open intergenic regions were enriched for evolutionarily conserved sequences, suggesting that a subset of these regions may be regulatory elements that promote SCLC progression (Figure 2E).

To determine whether these differentially accessible regions are common to other cell types or represent a pattern of accessibility unique to metastatic SCLC, we compared our chromatin accessibility data to DNase-seq from other cell types and tissues (Vierstra et al., 2014). The constitutively open regions were not dramatically enriched in any particular cell type (Figure S2E). However, the differentially open regions had the greatest overlap (>50%) with open regions in brain tissues (Figure 2F), suggesting that neuroendocrine SCLC tumor cells may co-opt regulatory elements found in neuronal tissue during cancer progression.

Differentially Accessible Regions Are Highly Enriched for NFI Binding Motifs

To identify potential drivers of this dramatic difference in accessibility, we determined the enrichment of transcription factor motifs within the differentially accessible regions, compared to those within constitutively open regions. The most highly enriched motif was the binding site for the NFI family of DNA binding factors (Figure 3A), which can bind either a full site of two inverted 5-bp repeats or a half site with lower affinity (Meisterernst et al., 1988). De novo motif enrichment also assembled the NFI half-site motif, and one or more de novo-assembled half sites were present in over 90% of the differentially open regions (Figures 3B and S3A). These data suggest that NFI transcription factors might play an important role in the large-scale chromatin accessibility changes observed during SCLC progression.

The Hyper-Accessible Samples Are Characterized by Nfib Copy Number Amplification

NFI transcription factors (Nfia, Nfib, Nfic, and Nfix) are widely expressed and the DNA-binding domain is distinct from other DNA-binding factors (Gronostajski 2000). We quantified DNA copy number based on ATAC-seq reads outside of accessible regions and found that the hyper-accessible samples specifically had genomic amplification of the *Nfib* locus (Figure 3C) (Dooley et al., 2011). DNA copy number gains at *Mycl1* or any other loci did not correlate with the hypo- or hyper-accessible states (Table S1).

Nfib Is Highly Expressed in Invasive Primary Tumors and Metastases, and Nfib^{high} Cells Specifically Have the Ability to Disseminate

To investigate whether high Nfib expression is associated with metastatic progression of SCLC, we performed immunohistochemical (IHC) staining for Nfib on tumors from *TKO-mTmG* mice at different stages of progression. Early neuroendocrine hyperplasias and most primary lung tumors expressed low or undetectable levels of Nfib (Figures 3D, 3E, and S3B). Conversely, invasive SCLC within the pulmonary lymphatic vessels, as well as lymph node, liver, and other distant metastases, were >95% Nfib^{high}, and these cells had more advanced cytology (Figures 3D, 3E, and S3B). Rare liver macro-metastases were Nfib^{low}, suggesting that other mechanisms can drive metastatic ability. This likely explains why one metastasis in our initial ATAC-seq dataset was in the hypo-accessible group.

Despite the presence of Nfib^{pos} invasive primary tumors and Nfib^{pos} metastases, it remained unclear whether only Nfib^{pos} cells can initially disseminate or whether both Nfib^{neg} and Nfib^{pos} cells can disseminate with selection for Nfib^{pos} cells occurring at a later step of the metastatic cascade. We examined Nfib expression at the single-cell level by immunofluorescence on FACS-purified SCLC cells. As expected, SCLC cells from individual primary tumors were either homogeneously Nfib^{neg} or a mixture of Nfib^{pos} and Nfib^{neg} cells, while cancer cells from metastases were nearly homogeneously Nfib^{pos} (Figure 3F). Disseminated tumor cells in the pleural fluid were also almost universally Nfib^{pos}, suggesting that the Nfib^{pos} cell state has the unique ability to overcome the initial hurdles that limit systemic spread.

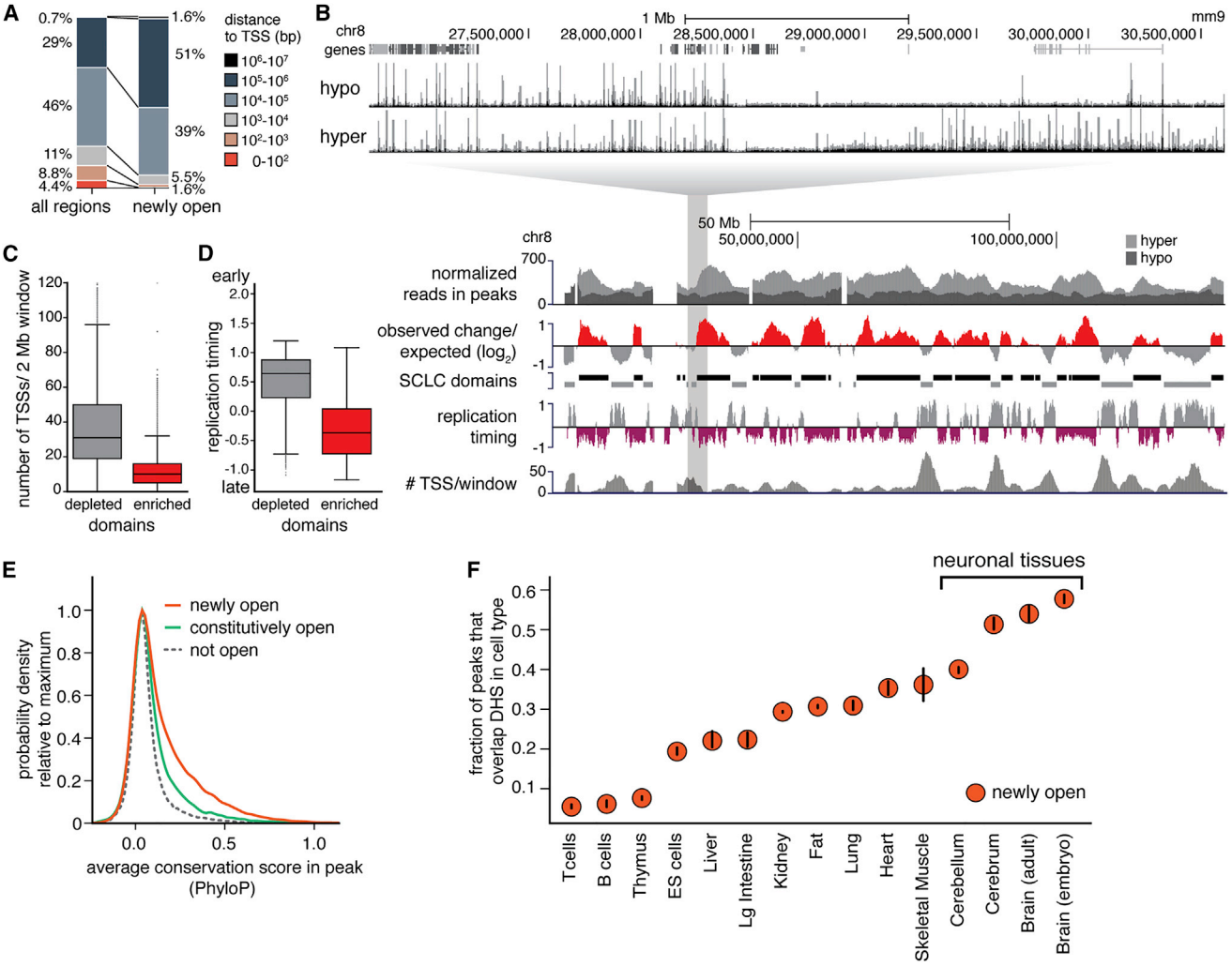


Figure 2. Changes in Chromatin Accessibility Are Gene Distal, Late Replicating, Evolutionarily Conserved, and Enriched for Regulatory Regions in Neuronal Tissue

(A) Distance to closest transcription start sites (TSSs) of all accessible regions and differentially open regions.

(B) Insertion tracks of merged hypo- and hyper-accessible samples. (Bottom) Reads in distal peaks are binned into 2 Mb windows. Examining differential read counts in windows identifies domains enriched for differentially open regions (black bars) and depleted for differentially open regions (gray bars, lower). Replication timing (ES-TT2 cells) and the number of TSSs per window are also shown.

(C) Number of TSSs per 2 Mb window in domains either enriched or depleted for differentially open regions in SCLC. Enriched domains have fewer TSSs ($p < 1 \times 10^{-100}$ by Mann-Whitney U test).

(D) Average replication timing per 2 Mb window (ES-TT2 cells) in domains either enriched or depleted for differentially open regions in SCLC. Enriched domains have significantly later (more negative) replication timing ($p < 1 \times 10^{-100}$ by Mann-Whitney U test).

(E) Average sequence conservation (phyloP) in differentially open intergenic regions, constitutively open intergenic regions, and closed regions (5 kb downstream of any accessible region, same window size). Differentially open, intergenic regions are twice as likely to have higher average sequence conservation (>0.2) than constitutively open, intergenic regions ($p = 1 \times 10^{-114}$ by Fisher's exact test).

(F) Overlap of differentially accessible regions with DNase hypersensitive sites (DHS) from other cell types. Mean overlap with DHS peak calls are shown. Bars represent 95% confidence intervals.

See also Figure S2.

NFIB Is Highly Expressed in a Subset of Human SCLC Primary Tumors and Metastases

NFIB is genomically amplified in 5%–15% of human primary SCLC tumors, consistent with high *NFIB* RNA expression in a subset of primary SCLC samples and cell lines (Figures S3C and S3D) (Dooley et al., 2011; George et al., 2015; Peifer et al., 2012; Rudin et al., 2012; Barretina et al., 2012). By IHC, NFIB was moderately to highly expressed in ~50% of primary human SCLCs, as well as in lymph node and distant metastases, with a trend toward higher expression in distant metastases (Figures S3E and S3F). Patients with stage IV SCLC had tumors with significantly higher NFIB levels than did stage

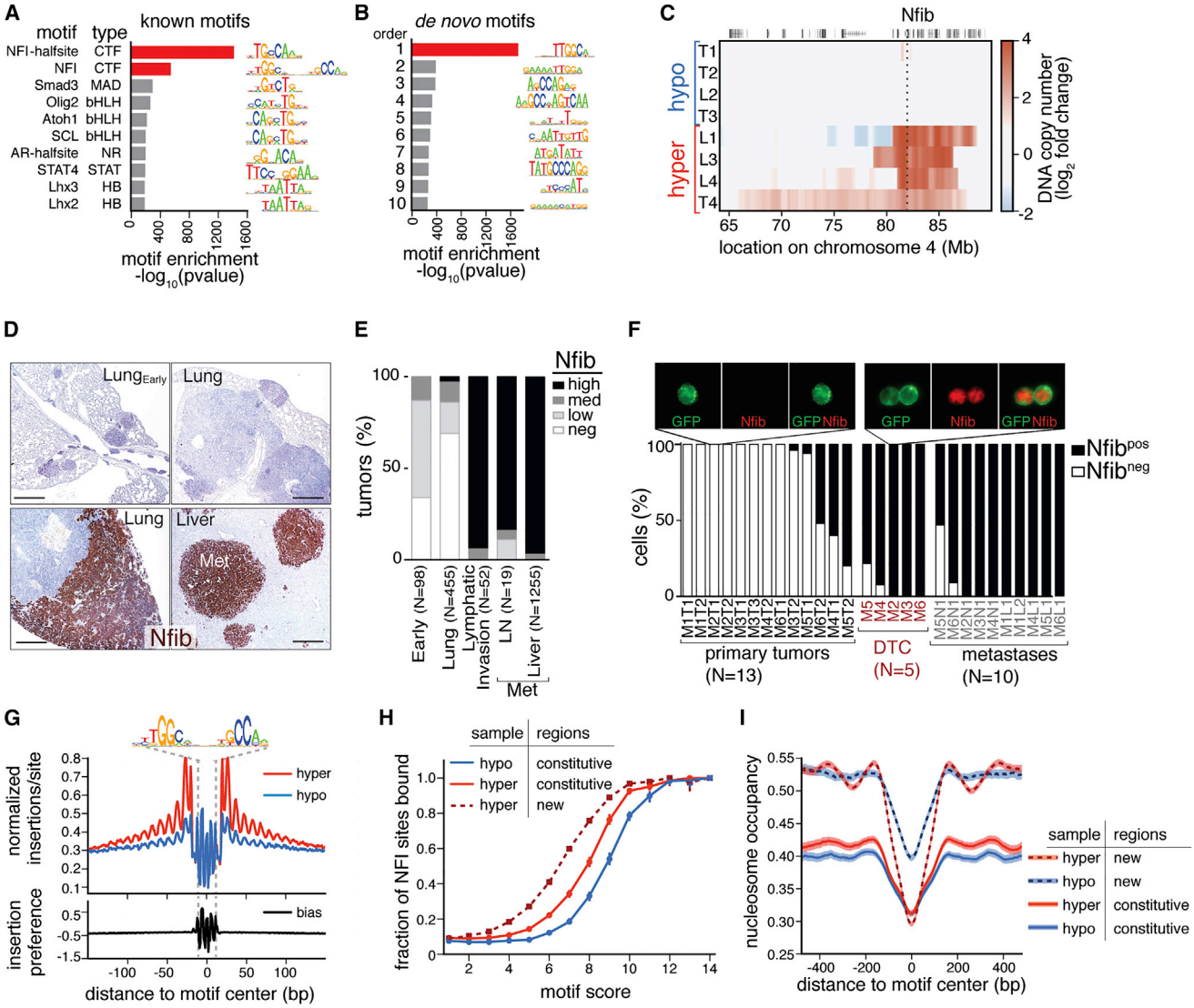


Figure 3. Increased Nfib Expression in Invasive SCLC Leads to Increased Nfib Binding in Differentially Open Regions and Occupancy of Less Canonical Sites

(A and B) Motif enrichment in newly open regions compared to other accessible regions. (A) Top known motif enrichments and (B) de novo motif enrichments.

(C) Copy number amplification of the *Nfib* locus inferred from ATAC-seq.

(D) Representative IHC for Nfib on tumors at different stages of SCLC progression from *TKO-mTmG* mice. Lung^{Early} are neuroendocrine hyperplasias. Scale bars, 500 μm.

(E) Nfib expression at different stages of SCLC progression. The number of tumors in each group is indicated. The percent of Nfib^{high} tumors in early hyperplasias and lung tumors versus lymph node (LN) and liver metastases is significantly different ($p < 1 \times 10^{-16}$ by Fisher's exact test).

(F) Immunofluorescence for Nfib on FACS-isolated GFP^{pos} cancer cells from primary tumors, disseminated tumor cells (DTCs), and metastases. A representative SCLC cell from a primary tumor and DTCs from the pleural cavity are shown. The percent of Nfib^{high} DTCs versus Nfib^{high} lung tumor cells is significantly different ($p < 0.0001$).

(G) (Top) ATAC-seq footprint at the NFI full site. Insertions per site are normalized to have the same average number of insertions 200–500 bp away from motif center. (Bottom) Modeled insertion bias of Tn5 around NFI full sites.

(H) Occupancy of NFI full sites in the merged hyper- and hypo-accessible samples in bins of motif score (log odds similarity to the consensus motif). Error bars represent 95% confidence intervals.

(I) Nucleosome occupancy around NFI full sites in differentially open and constitutively open regions in hypo- and hyper-accessible samples. Shaded areas represent 95% confidence intervals.

See also Figure S3 and Table S1.

I and stage II patients, and patients with higher NFIB had a trend toward shorter overall and progression-free survival (Figures S3G–S3I).

Genomic Regions Containing NFI Motifs Have Increased Occupancy in Hyper-Accessible Samples

To approach the question of how increased expression of Nfib could lead to widespread opening of genomic loci, we employed transcription factor footprinting, which uses ATAC-seq reads to infer binding based on the steric hindrance between transcription factor complexes and the Tn5 transposase used to fragment the genome (Buenrostro et al., 2013). We detected an extremely clear footprint of transcription factor occupancy around the aggregated NFI full sites in accessible regions, suggesting that the NFI-DNA interaction is long-lived (Sung et al., 2014) (Figure 3G). None of the other transcription factors motifs enriched in the hyper-accessible state showed clear differential footprinting in the hypo- versus hyper-accessible chromatin state (data not shown).

We observed a clear pattern of ATAC-seq fragment sizes around NFI motif sites, allowing us to use this information in addition to insertion frequency to infer NFI occupancy at individual sites (Figures S3J and S3K). We divided NFI sites based on their similarity to the consensus NFI position weight matrix (a proxy for biochemical affinity) (Benos et al., 2002) and compared the inferred occupancies of sites within each bin for the hypo- and hyper-accessible chromatin states. The hyper-accessible chromatin state was characterized by increased occupancy of sites, particularly at sites of intermediate affinity, supporting a model in which increased expression of Nfib drives binding to less canonical sites (Figure 3H). NFI sites in differentially accessible regions had even higher occupancy than constitutively open sites, suggesting that they are especially amenable to Nfib binding, beyond what would be expected from their motif quality alone (Figure 3H).

Occupancy by NFI Depletes Nucleosomes and Changes Local Chromatin Architecture

Nucleosomes and transcription factors compete for access to DNA such that increased occupancy of NFI sites may deplete nucleosomes, thereby inducing accessibility changes at NFI sites genome-wide (Bell et al., 2011). By inferring nucleosome occupancy using the insert sizes of the ATAC-seq reads (Schep et al., 2015), we found that differentially open regions were more depleted of nucleosomes around NFI sites in hyper-accessible samples (Figure 3I). The amount of nucleosome depletion was dependent on the occupancy of NFI motifs, such that higher affinity NFI sites were associated with greater depletion of nucleosomes, but lower affinity sites were associated with greatest differential nucleosome depletion between the two chromatin states (Figures S3L and S3M).

These data support a model in which Nfib stably binds DNA and directly competes with nucleosomes to stabilize an accessible chromatin configuration. These results suggest that Nfib upregulation may initiate and maintain chromatin state transitions, ultimately leading to transcriptional changes associated with a gain in metastatic ability.

SCLC Cell Lines Recapitulate the Differential Chromatin States Associated with Nfib Expression

To determine whether SCLC cell lines maintain the chromatin landscapes of the in vivo tumors, we assessed Nfib expression and the chromatin state of six murine SCLC cell lines (Figures 4A–4C). The two cell lines with highest Nfib protein expression (KP1 and 16T) had *Nfib* genomic amplification (Figures 4A and 4B). Hierarchical clustering and principal component analysis of the chromatin accessibility landscape of the cell lines clearly separated the Nfib^{high} from Nfib^{low} lines (Figures 4C and S4A), and the Nfib^{high} cell lines had increased accessibility relative to the Nfib^{low} cell lines (Figures 4D, 4E, and S4B). The changes in accessibility were highly correlated with those in ex vivo samples, and the differentially open regions were similarly gene-distal and enriched for NFI motifs ($r = 0.72$; Figures 4F and S4C–S4E). Thus, cell lines faithfully recapitulate the dramatic chromatin accessibility changes found in vivo and further support the role of Nfib in defining their chromatin states.

Increased Direct Binding of Nfib to Differentially Accessible Sites

We performed chromatin immunoprecipitation (ChIP-seq) for Nfib on three SCLC cell lines, 16T (Nfib^{high}), KP1 (Nfib^{high}), and KP22 (Nfib^{low}) and compared Nfib occupancy between the Nfib^{high} and Nfib^{low} cell lines (Figures 4G and 4H). Peaks that were differentially occupied had, on average, lower motif scores than peaks that were constitutively occupied in the Nfib^{high} and Nfib^{low} cell lines, consistent with increased expression of Nfib enabling binding of lower affinity sites (Figure 4I). Differential accessibility correlated with differential ChIP signal (Figures 4J and S4F), confirming that a high fraction of differential accessibility is attributable to differential occupancy by Nfib. Sites with greater Nfib ChIP signal in the Nfib^{high} lines were specifically enriched for many other transcription factor binding motifs, suggesting that other transcription factors may play a role upstream or downstream of Nfib action (Figures S4G and S4H).

Nfib Maintains the Hyper-Accessible Chromatin State

To assess the requirement for sustained Nfib expression to maintain an open chromatin state and metastatic ability, we stably knocked down *Nfib* in two Nfib^{high} SCLC cell lines (Figures 5A, S5A, and S5B). ATAC-seq on shControl and shNfib cell lines uncovered extensive changes in chromatin accessibility (Figures 5B and S5C). Regions that close upon *Nfib* knockdown were generally anticorrelated with accessibility changes between hypo- and hyper-accessible ex vivo samples ($r = -0.55$). Indeed, the majority of regions (82%) that were differentially open in the hyper accessible class relied on Nfib to maintain accessibility (Figure 5C).

The regions that decrease in accessibility in *Nfib* knockdown cells were generally gene-distal and highly enriched for NFI sites (Figures S5D and S5E). The large majority (75%) of these regions showed differential Nfib occupancy based on Nfib ChIP-seq, and, on average, they had increased nucleosome occupancy at NFI motif sites (Figure 5D). These results demonstrate that Nfib maintains chromatin accessibility and local chromatin architecture.

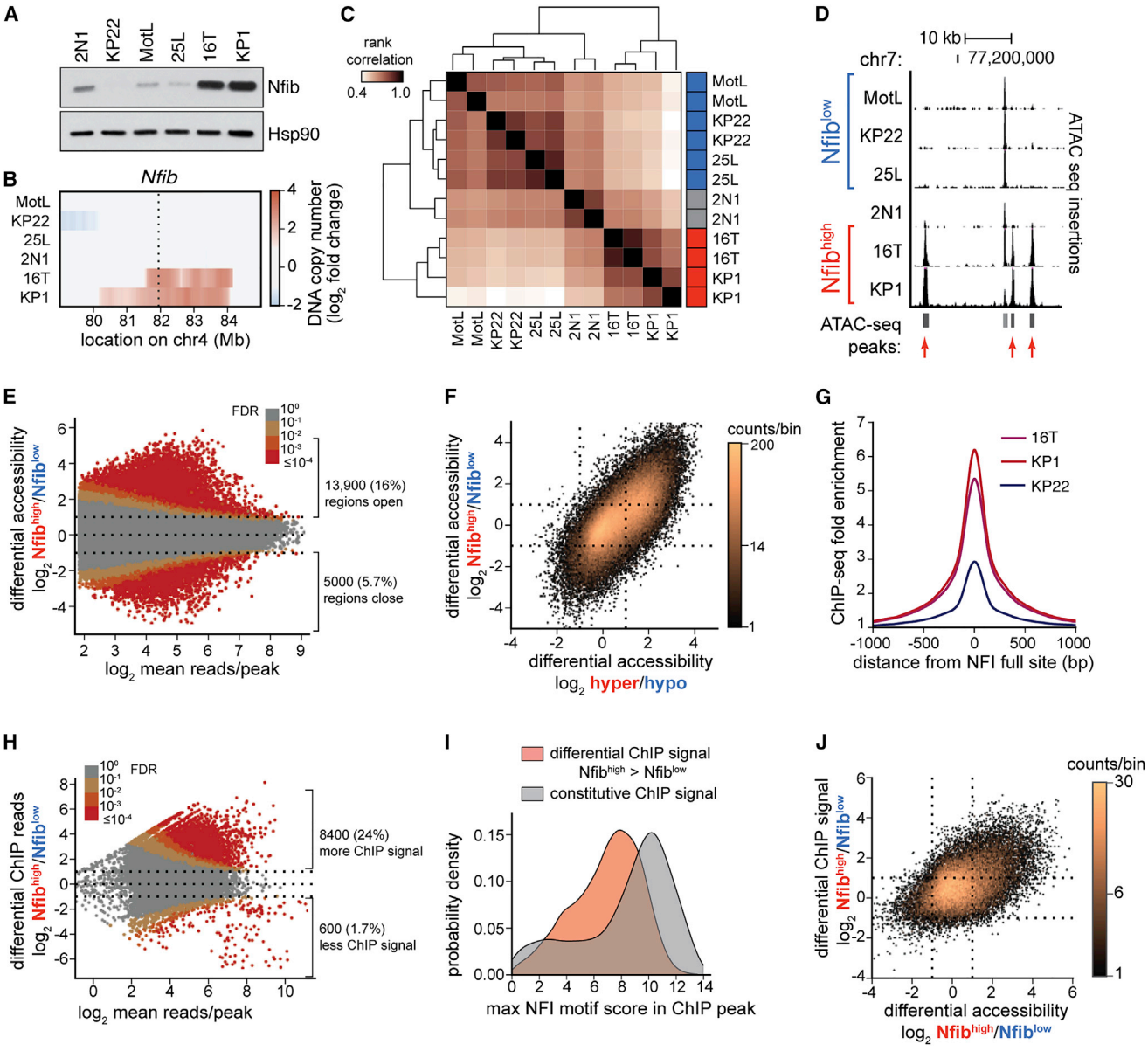


Figure 4. Analysis of SCLC Cell Lines Confirm the Relationship between Nfib Expression, Nfib Binding by ChIP, and Chromatin Accessibility State

(A) Immunoblot analysis of Nfib expression in six murine SCLC cell lines. Hsp90 shows loading.

(B) Copy number amplification of Nfib inferred from ATAC-seq.

(C) Correlation of chromatin accessibility across cell lines and technical replicates. Nfib^{high} and Nfib^{low} cell lines are indicated in red and blue, respectively.

(D) ATAC-seq insertion tracks of cell lines. Differentially accessible regions are indicated with arrows.

(E) Differential accessibility (log_2 fold change in reads per accessible region) between Nfib^{high} and Nfib^{low} cell lines, plotted against the mean reads per region. FDR is the false discovery rate that the absolute value of log_2 fold change is >0.5. Values beside brackets indicate the number of significantly changed peaks.

(F) Correlation of differential accessibility between Nfib^{high} and Nfib^{low} cell lines and hypo- and hyper-accessible ex vivo samples (r = 0.72).

(G) Fold enrichment above input of Nfib ChIP around NFI motif sites for three cell lines.

(H) Log_2 fold change in Nfib ChIP-seq reads per merged ChIP peak versus the mean number of reads per peak. FDR is the false discovery rate that the absolute value of log_2 fold change is >0.5.

(I) Distribution of motif scores of sites within Nfib ChIP peaks, either those that gave more signal in the Nfib^{high} than the Nfib^{low} cell lines or those that were not significantly different (constitutive). The maximum scoring NFI full site within each ChIP peak was used. Constitutive peaks have a higher motif score than differential peaks (Mann-Whitney $p < 1 \times 10^{-100}$).

(J) Correlation of differential ChIP signal in accessible regions (from ATAC-seq) and differential accessibility of Nfib^{high} and Nfib^{low} cell lines (r = 0.35).

See also Figure S4.

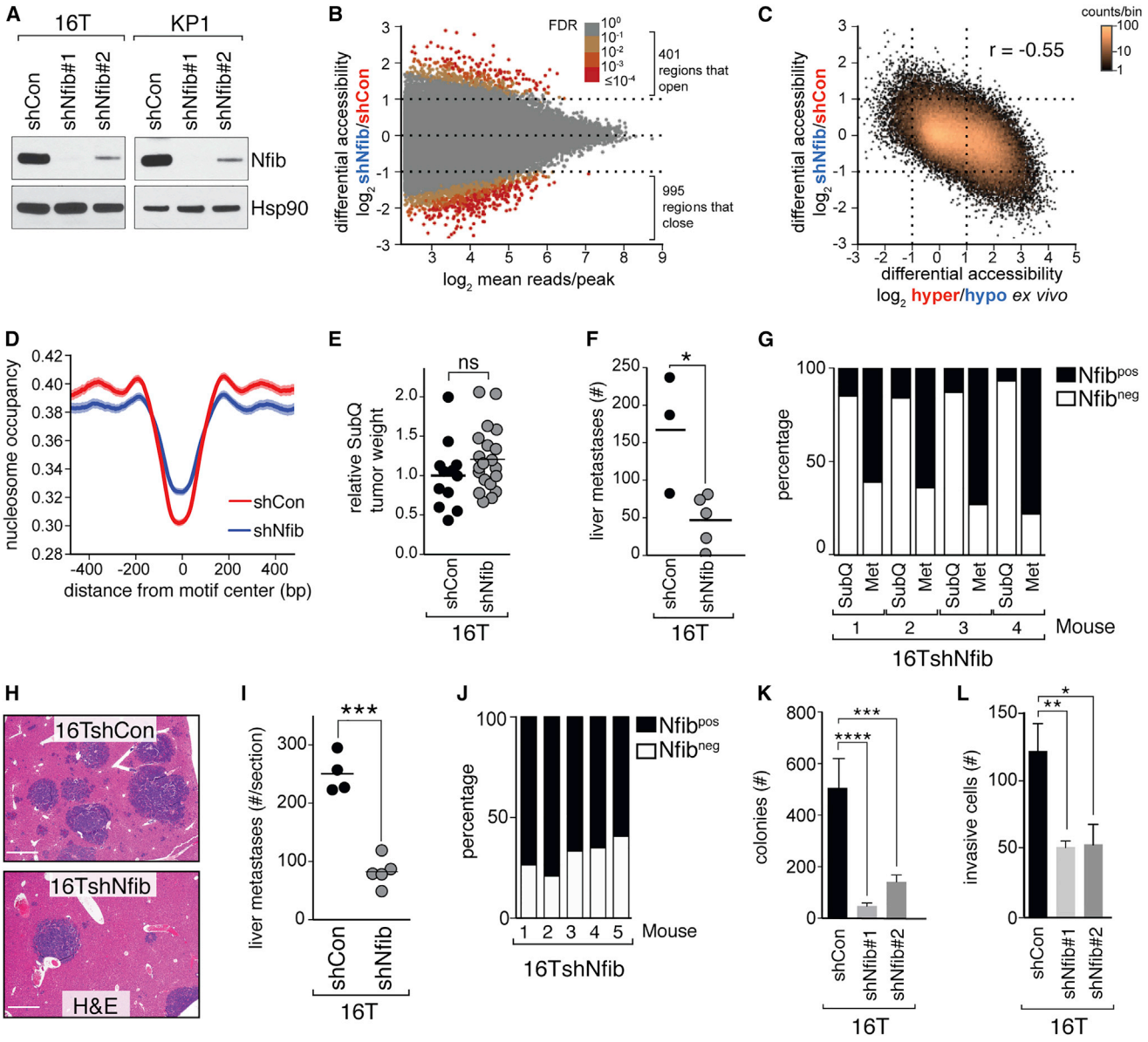


Figure 5. Nfib Maintains Chromatin Accessibility at a Subset of Genomic Regions and Is Required for Metastatic Ability

For a Figure360 author presentation of Figure 5, see http://dx.doi.org/10.1016/j.cell.2016.05.052#mmc4

(A) Immunoblot for Nfib in two Nfib^{high} SCLC cell lines ±*Nfib* knockdown (shNfib). Hsp90 shows loading. shCon, shControl.

(B) Differential accessibility with shNfib#1 and control in combined 16T and KP1 cell lines.

(C) Correlation of differential accessibility with *Nfib* knockdown and differential accessibility of hyper- and hypo-accessible ex vivo samples ($r = -0.55$).

(D) Nucleosome occupancy around NFI sites in shNfib and control cells. Shaded regions indicate 95% confidence intervals.

(E) Weight of the subcutaneous (SubQ) tumors that formed from 16T cells ±*Nfib* knockdown 4 weeks after transplantation. Each dot represents a tumor, and the line indicates the mean. ns, not significant.

(F) Number of liver metastases from SubQ 16T tumors ±*Nfib* knockdown. Scale bars, 500 μm. Each dot represents a mouse, and the line indicates the mean. *$p < 0.025$.

(G) Percent of SubQ tumor area that expresses Nfib (Nfib^{pos}) and percent of liver metastases (Met) that are Nfib^{pos}. Paired analysis of Nfib^{pos} areas in SubQ and Met tumors is significant ($p < 0.0025$).

(H and I) Representative H&E images (I) and quantification (H) of liver metastases 3 weeks after intravenous transplantation. Scale bars, 500 μm. Each dot represents a mouse, and the line is the mean. ***$p < 0.001$.

(J) Percent of Nfib-positive liver metastases (assessed by IHC) after intravenous transplantation of 16TshNfib cells.

(K) Anchorage-independent growth of an Nfib^{high} SCLC cell line ±*Nfib* knockdown. Mean ± SD is shown. ****$p < 0.0001$, ***$p < 0.001$.

(L) Matrigel migration assay of an Nfib^{high} SCLC cell line ±*Nfib* knockdown. Mean ±SD is shown. **$p < 0.01$ *$p < 0.02$.

See also Figure S5.

Nfib Is Required for Metastatic Ability In Vivo

Given the expression of Nfib in invasive SCLC and metastases, as well as the dramatic changes in global open chromatin structure upon *Nfib* knockdown, we tested whether Nfib is required for metastatic ability. Subcutaneous growth of Nfib^{high} SCLC cell lines led to the formation of spontaneous liver metastases. *Nfib* knockdown did not significantly alter subcutaneous tumor growth; however, mice with 16TshNfib tumors had 3-fold fewer liver metastases than mice with 16TshControl tumors (Figures 5E, 5F, and S5F). Because the selective pressure of tumor growth and the metastatic process can select for cancer cells that have escaped knockdown, we assessed Nfib expression in the subcutaneous tumors and metastases. In 16TshNfib subcutaneous tumors, 10%–20% of cells had lost *Nfib* knockdown and were Nfib^{high}, suggesting a small but likely meaningful selective advantage for Nfib-expressing cells (Figure 5G). Conversely, >60% of the metastases in mice with 16TshNfib tumors expressed Nfib (Figure 5G). This dramatic selection for Nfib expression in metastases further underscores the strong metastatic fitness advantage conferred by Nfib.

Subcutaneous growth of a second Nfib^{high} SCLC cell line (KP1) was likewise unaffected by *Nfib* knockdown, and KP1shNfib tumors also seeded ~2.5-fold fewer liver metastases than KP1shControl tumors (Figures S5G and S5H). The Nfib^{pos} cells that escaped *Nfib* knockdown outcompeted Nfib^{neg} cells to an even greater extent within KP1shNfib subcutaneous tumors than within 16TshNfib subcutaneous tumors, and the metastases that formed were almost exclusively Nfib^{pos} (Figure S5I).

We further investigated the role of Nfib in conferring metastatic ability using intravenous transplantation of 16TshControl and 16TshNfib cells. 16TshNfib cells seeded fewer metastases than control cells, and more than half of the metastases that formed from 16TshNfib cells escaped *Nfib* knockdown (Figures 5H–5J), indicating that Nfib is also required for later stages of the metastatic process.

Nfib Is Required for Clonal Growth and the Invasive Ability of SCLC Cells

Nfib knockdown reduced proliferation but had no consistent impact on cell death under standard culture conditions (Figures S5J–S5Q). Because these SCLC cell lines grow in culture as floating spheres, reduced proliferation suggested that Nfib may influence anchorage-independent growth. To directly assess clonal growth in anchorage-independent conditions, we plated shControl and shNfib cells in soft agar. *Nfib* knockdown greatly reduced the ability of SCLC cells to expand into colonies under these conditions (Figures 5K, S5R, and S5S). *Nfib* knockdown also greatly reduced migration (Figures 5L, S5T, and S5U). These results indicate that Nfib controls several cellular phenotypes that likely cooperate to drive metastatic ability.

Nfib Is Sufficient to Open a Subset of Distal Regulatory Regions

To determine whether Nfib is also sufficient to drive these changes in chromatin state and metastatic ability, we generated an Nfib^{low} cell line with doxycycline-inducible *Nfib* expression (KP22-TRE-Nfib; Figure 6A). ATAC-seq on KP22-TRE-Nfib and KP22-TRE-empty cells after doxycycline treatment indicated that Nfib expression was sufficient to rapidly drive chromatin opening of ~1,800 regions (Figures 6B and S6A). These regions were highly enriched for loci that were more open in the hyper-accessible ex vivo samples, as well as for regions that had reduced accessibility in *Nfib*-knockdown cells (Figures 6C and S6A). The regions that opened with Nfib expression were largely gene-distal, were enriched for NFI motifs, and had high overlap with the differential Nfib ChIP signal in the Nfib^{high} versus the Nfib^{low} lines (60% overlap, $p < 1 \times 10^{-300}$ by hypergeometric test; Figures S6B and S6C).

Unlike *Nfib* knockdown, which reduced accessibility at many regions, Nfib overexpression increased accessibility at only a subset of sites that had slightly higher motif scores and slightly higher overlap with putative regulatory elements in other cell types (Figures 6C, S6D, and S6E). The affinity of NFI factors to DNA has been shown to be reduced by up to 300-fold by the presence of nucleosomes (Blomquist et al., 1996), suggesting NFI is not a classically defined "pioneer" transcription factor capable of evicting nucleosomes directly from nucleosome-bound DNA (Zaret and Mango 2016). The sites sensitive to Nfib induction had modest but above-background accessibility in the Nfib^{low} cell line prior to induction of Nfib expression (Figures 6D and 6E). Those sites that were newly open in the hyper-accessible samples, but not sensitive to Nfib overexpression, were not significantly open relative to the background in this cell line (Figure 6E), suggesting that Nfib sensitivity may require a permissive level of accessibility to allow Nfib-mediated changes in the nearby chromatin landscape. Further supporting this hypothesis, the hypo-accessible samples had modest accessibility above background in the sites that were differentially open (Figure S6F).

To extrapolate our findings to human SCLC, we tested whether NFIB is sufficient to increase chromatin accessibility in human SCLC cells. Expression of NFIB in two NFIB^{low} human SCLC cell lines increased chromatin accessibility at a subset of regions in both cell lines (Figures S6L and S6M). Consistent with the effect of Nfib in the mouse SCLC cell lines, NFIB expression predominantly led to increased accessibility at distal regulatory regions that were enriched for NFI motifs (Figure S6M). Human cell lines with NFIB overexpression showed increased occupancy of NFI sites, to an even greater degree than in mouse cell lines (Figure S6N).

Nfib Is Sufficient to Drive Clonal Growth In Vitro and Metastatic Ability In Vivo

To determine whether Nfib expression is also sufficient to increase clonal growth in cell culture, we quantified the ability of KP22 cells with either inducible or constitutive Nfib expression to form colonies in anchorage-independent conditions. In both cases, Nfib expression greatly increased the clonal growth ability of SCLC cells (Figures 6F and S6J). Nfib expression did not consistently alter cell death in vitro, but increased proliferation under standard suspension growth conditions (data not shown). NFIB expression in human SCLC cells also increased cell growth under standard suspension growth conditions, as well as in anchorage-independent conditions (Figures S6O–S6R).

Consistent with our functional data from *Nfib* knockdown cells, KP22-Nfib cells formed ~3-fold more liver metastases after intravenous transplantation relative to controls (Figures 6G–6I). Notably, the few metastases that formed from KP22 control cells

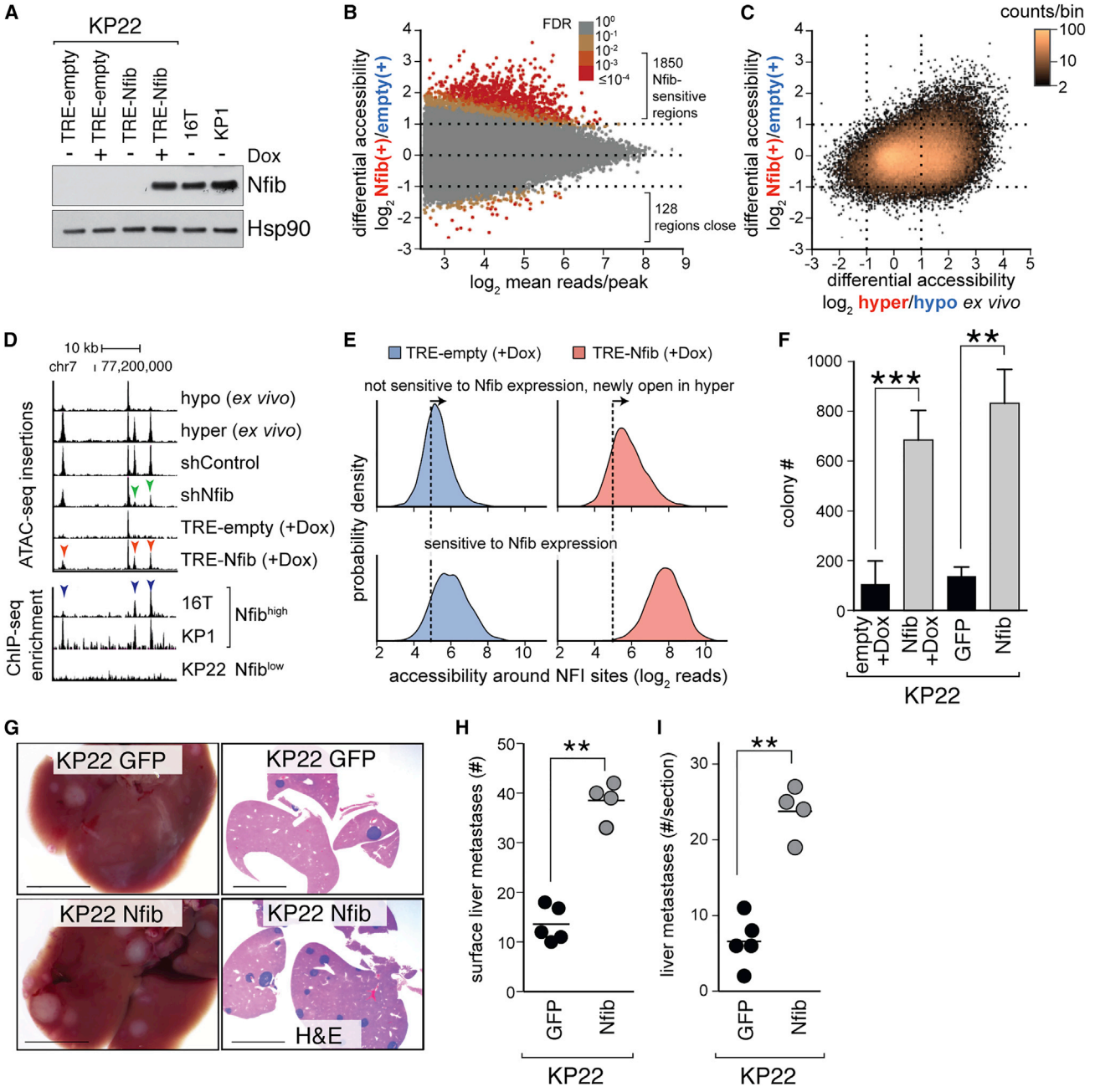


Figure 6. Nfib Is Sufficient to Open a Subset of Sensitive Genomic Regions and to Drive Metastatic Ability

(A) Expression of Nfib by immunoblot in an Nfib^{low} cell line (KP22) with doxycycline (Dox)-inducible expression of Nfib. Two Nfib^{high} cell lines 16T and KP1 are shown. Hsp90 shows loading.

(B) Differential accessibility in Nfib-expressing and control samples.

(C) Correlation of differential accessibility with Nfib expression and differential accessibility of hyper- and hypo-accessible ex vivo samples (r = 0.26).

(D) (Top) ATAC-seq insertions and (bottom) Nfib ChIP-seq enrichment above input at an example locus. Green arrows highlight regions that close with *Nfib* knockdown; red arrows highlight regions that open with Nfib expression; and blue arrows show regions with differential ChIP-seq signal.

(E) Distribution of accessibility (log_2 read count) in 1,000 bp windows around NFI motif sites ±Nfib induction. Panels show motif sites specifically more open in hyper-accessible ex vivo samples or sensitive to Nfib overexpression in KP22 cell line. The dashed line indicates the average number of reads around motif sites that were not accessible.

(F) Colony formation of KP22 cells with induced (Nfib+Dox) and constitutive Nfib expression in anchorage-independent conditions. Mean ± SD is shown. ***p < 0.0005 **p < 0.001.

(legend continued on next page)

had high Nfib expression (data not shown), suggesting a strong selection for Nfib^{high} cells during the post-intravasation steps of the metastatic cascade. Nfib expression did not change subcutaneous tumor growth, confirming that Nfib has a limited effect on SCLC growth in vivo (Figure S6K; data not shown).

Nfib Promotes Neuronal Gene Expression Programs

Our results establish that Nfib modulates accessibility of many distal, putative regulatory regions of the genome. To identify what gene expression programs enable the metastatic phenotype of Nfib^{high} tumors, we examined gene expression changes upon Nfib overexpression and knockdown (Figure S7A). Approximately 500 genes were significantly upregulated with Nfib overexpression and a similar number of genes changed after *Nfib* knockdown (Figures S7B–S7D).

Almost all (92%) of the genes that significantly changed with both overexpression and knockdown changed reciprocally (Figure 7A). The genes that were upregulated with Nfib overexpression and downregulated with *Nfib* knockdown were enriched for gene ontology annotations associated with neuronal function, including axon guidance, synapse organization, and regulation of nervous system development (Figures 7B and S7E).

Gene Expression Changes Are Associated with Distal Regions that Become Accessible upon Nfib Upregulation

Interestingly, we observed Nfib-driven neuronal-associated gene expression programs and a high fraction of Nfib-driven chromatin accessibility changes overlap with putative regulatory elements found in brain tissue (>70%; Figures 7C and S6E). Thus, two distinct datasets (gene expression and chromatin accessibility) link neuronal programs to the metastatic ability of this neuroendocrine cancer.

Nfib appears to drive chromatin accessibility at largely gene-distal regions of the genome. Examining 100 kb windows around genes that were upregulated with Nfib overexpression and downregulated with *Nfib* knockdown uncovered a strong enrichment for gene-distal regions that increased in accessibility with Nfib overexpression (Figure 7D). Having only an Nfib ChIP peak at the promoter (in any cell line) did not have a significant effect on gene expression with Nfib overexpression (Figure 7E), indicating that direct action at the promoters is not the primary mode of gene regulation by Nfib. Collectively, our data support a model in which Nfib stabilizes the open chromatin state at distal regulatory elements in SCLC cells, thereby inducing a gene expression program related to neuronal-associated processes that drive metastatic ability (Figures 7F and 7G).

DISCUSSION

Nfib Drives Multiple Steps of the Metastatic Process

Metastasis is a low-probability, multistep process in which cancer cells from primary tumors must invade the local tissue, disseminate, survive in circulation, extravasate into a secondary site, and expand into a tissue-destructive metastasis. The early spread and high frequency of metastasis in SCLC patients could suggest that SCLC cells inherently possess the capacity to metastasize (Hou et al., 2012). In contrast, our work shows that SCLC tumors can gain metastatic ability through a dramatic remodeling of their chromatin state (Figure 7G). Our results highlight the power of combining genome-wide molecular approaches with genetic model systems to uncover hidden variations in cellular states.

Our data demonstrate that a single transcription factor, Nfib, promotes the ability of SCLC cells to perform several of the requisite steps to metastasize, including invasion, dissemination, and clonal growth. We confirmed moderate to high NFIB expression in human SCLC and showed that NFIB overexpression increases anchorage-independent growth and induces chromatin changes in human SCLC cell lines (Dooley et al., 2011). These data suggest that a subset of human SCLCs likely use NFIB-driven changes in chromatin state as their route for metastasis, although alternative pathways likely exist. A more complete understanding of all pathways by which human SCLC gains metastatic ability, NFIB-driven or otherwise, would allow for better patient stratification and the improved division of SCLC into subtypes based on their pro-metastatic programs.

The Stable and Highly Metastatic SCLC State Drives Rampant Metastatic Spread

Across many cancer types, loss of differentiation correlates with aggressiveness and poor patient outcome; however, in SCLC, the link between dedifferentiation and aggressiveness has remained unclear. Amplification of pluripotency factors drives aggressive SCLC (Rudin et al., 2012; Lin et al., 2012), and inhibition of a histone lysine demethylase, which is known to maintain pluripotency in human embryonic stem cells, delays SCLC growth (Mohammad et al., 2015). Conversely, lineage differentiation pathways driven by Ascl1 and NeuroD1 also promote SCLC tumorigenesis and progression (Augustyn et al., 2014; Osborne et al., 2013). Our data suggest an alternate model, in which partial transdifferentiation driven by Nfib creates a neuronal program of gene expression that drives metastatic ability (Figures 7B, 7G, and S7E).

Unlike the epithelial-mesenchymal transition, which is proposed to create a transient, highly metastatic state (Yang and Weinberg 2008), Nfib amplification, overexpression, and the resulting chromatin changes create a stable and highly metastatic state maintained in metastases. This finding may shed light on why SCLC is so widely metastatic in patients: established Nfib-driven metastases remain in a highly metastatic state and may continuously seed additional secondary metastases (Hou et al., 2012).

Mechanism of Targeting Nfib to Functional Sites

Nfib's inability to bind nucleosome-bound DNA suggests that it does not act as a canonical pioneer factor (Zaret

(G) Light (left) and H&E (right) image of liver metastases of KP22 ± Nfib expression after intravenous transplantation. Scale bars, 5 mm.
(H–I) Number of surface liver metastases (H) and liver metastases quantified by histology (I). Each dot represents a mouse, and the line is the mean. ***$p < 0.001$.
See also Figure S6.

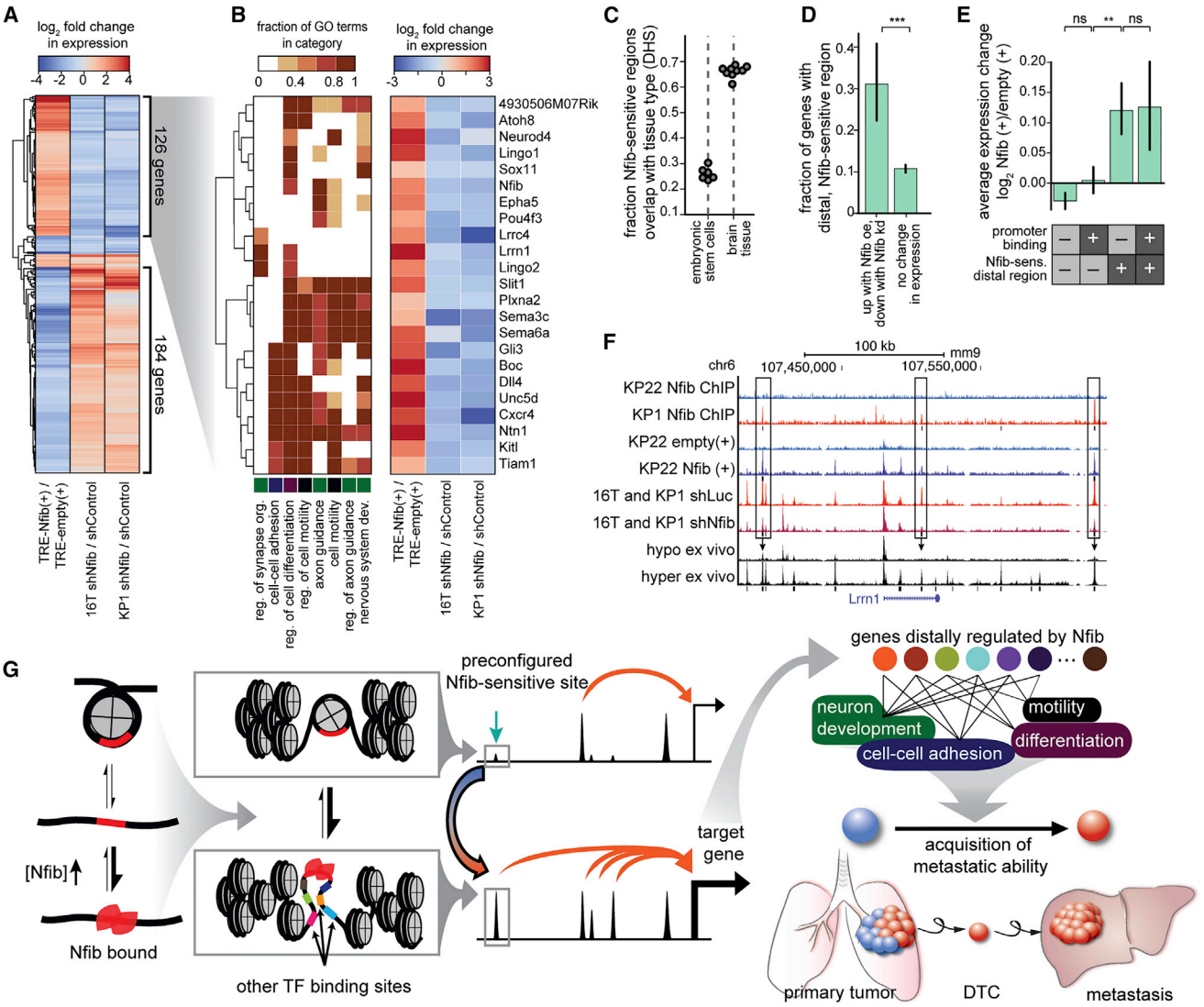


Figure 7. Gene Expression Changes in Response to Nfib Promote Neuronal State

(A) Log_2 fold change of all genes that significantly change (absolute value of log_2 fold change > 0.5 at FDR < 0.1) with both Nfib knockdown and overexpression by RNA-seq.

(B) Gene ontology (GO) terms (merged into categories) and log_2 fold change in expression of a subset of genes.

(C) Overlap of Nfib-sensitive regions with DNase-hypersensitive sites (DHS) in other tissues. Dots represent multiple ES cell lines, plus technical replicates, or adult and embryonic brain, plus technical replicates.

(D) Fraction of genes that have at least one distal, Nfib-sensitive region within 100 kb of transcription start site. oe, overexpression, KD, knockdown. Error bars are 95% confidence intervals. ***$p < 1 \times 10^{-10}$ by Fisher's exact test.

(E) Average change in expression with Nfib overexpression of four gene categories. Promoter binding, genes with promoter-proximal ChIP peaks within 1 kb of TSS. Nfib-sens. distal region, genes with a distal, Nfib-sensitive region within 100 kb of the TSS that opens with Nfib overexpression. Error bars are 95% confidence intervals. ns, not significant; **$p < 1x10^{-7}$ by Mann-Whitney U test.

(F) Example locus showing Nfib ChIP-seq (top two panels) and ATAC-seq data. Black squares below each set of tracks represent significantly changed regions.

(G) Model for Nfib-dependent metastatic progression.

See also Figure S7 and Table S2.

and Mango 2016; Blomquist et al., 1996). Our data suggest that Nfib-sensitive sites exist in a "preset" configuration that permits binding of these motifs when Nfib is upregulated. On average, these preset sites are characterized by modest, but above-background accessibility and an absence of Nfib binding (Figures 6D and 6E). It is notable that such large numbers of preset sites exist in SCLC cells. The ease with which Nfib can reconfigure the chromatin state of SCLC cells indicates that the initial cancerous state may be prone to conversion toward a metastatic state.

Several mechanisms could establish this preset architecture. Foxa transcription factors are pioneer factors that interact with

NFI in other cell types, and Foxa motifs are enriched around the NFI sites more bound in the Nfib^{high} cell lines (Figure S4J) (Grabowska et al., 2014). A fraction of sites, therefore, may rely on Foxa factors to maintain a permissive chromatin state around NFI sites, marking them for opening on Nfib upregulation. Histone turnover in active regulatory regions may also enable sensitivity to Nfib expression, possibly through the action of chromatin remodelers (Hebbar and Archer 2003; Deal et al., 2010).

Molecular Mechanism of Chromatin Opening and Gene Activation

How does Nfib establish an open chromatin architecture after binding? Nfib gives an especially strong "footprinting" signal in our ATAC-seq data, a feature associated with a slow DNA binding off-rate (Sung et al., 2014). Furthermore, the high affinity of NFI to its consensus full site (Meisterernst et al., 1988) implies a long half-life similar in magnitude to the longest-lived transcription factors (Sung et al., 2016). This distinguishes Nfib from other transcription factors that are thought to have short residence time on DNA and are insufficient to compete with nucleosomes or other transcription factors (Sung et al., 2016; Voss and Hager 2014). The long residence time of Nfib on DNA likely allows the adaptation of nearby chromatin and enforces a prototypical chromatin architecture with positioned nucleosomes. This architecture is supported by the depleted nucleosome occupancy and positioned proximal nucleosomes around NFI sites in the hyper-accessible state (Figure 3I). Many additional motifs are enriched around newly open sites, and binding at these motifs may help stabilize an open structure (Figures 3A, S4J, and S6B). Our data do not rule out Nfib-induced destabilization of higher-order nucleosome interactions (Alevizopoulos et al., 1995) or recruitment of chromatin remodelers (Hebbar and Archer 2003), but rather suggests that these mechanisms could complement the direct nucleosome depletion and positioning effects of Nfib.

Our data indicate that Nfib establishes a pro-metastatic gene expression program through stabilizing chromatin accessibility at distal regulatory elements. Increased Nfib occupancy in the absence of increased accessibility is only weakly associated with increased gene expression, while Nfib-driven increases in distal accessibility are associated with upregulation of nearby genes (Figures 7D, S7E, and S7F). Nfib has diverse functions and has been shown to both promote cell differentiation during development and maintain populations of stem cells in adult tissues (Harris et al., 2015; Gronostajski 2000; Chang et al., 2013). An implication of our study is that maintenance of open chromatin architecture by Nfib may promote varied functional outcomes through combinatorial binding of nearby transcription factors.

Nfib Regulates Neuronal Guidance and Migration Pathways

Our data suggest that incorporating features of a neuronal-like program into the SCLC state promotes metastatic ability. Nfib is important for brain development and neuronal migration, and several of the neuronal-associated Nfib-regulated genes and pathways (including Cxcr4, the Eph receptors, Semaphorins, Netrin-1, and Slit/Robo families) have been implicated in invasion and metastasis of other cancer types (Steele-Perkins et al., 2005; Betancourt et al., 2014; Burger et al., 2003; Vanharanta et al., 2013; Chen 2012; Foley et al., 2015; Ko et al., 2014; Zhou et al., 2011). Our results suggest that overlapping components of a neuronal program with the transformed SCLC state can drive the insidious metastatic ability of this cancer type.

Epigenetic machinery has recently emerged as a promising therapeutic target, and preclinical results from modulating components of these pathways have been encouraging (Mohammad et al., 2015). Progression to pro-metastatic epigenetic states may create collateral therapeutic vulnerabilities or provide direct targets to limit the development of initial metastases or halt the propagation of secondary metastases. Given the dramatic remodeling of the gene regulatory state driven by a single factor in SCLC, we speculate that other major cancer types may also gain metastatic proclivity through large-scale remodeling of their chromatin state.

EXPERIMENTAL PROCEDURES

Mouse Model

All experiments were performed in accordance with Stanford University's Animal Care and Use Committee guidelines. *Trp53^{flox}*, *Rb1^{flox}*, *p130^{flox}*, and *R26^{mTmG}* mice have been described previously (Schaffer et al., 2010; Muzumdar et al., 2007). Tumors were initiated by inhalation of Adeno-CMV-Cre (University of Iowa Vector Core).

ATAC-Seq Library Preparation and Differential Accessibility

5×10^4 cells were from ex vivo tumors or from cell culture directly. ATAC-seq libraries were generated as described previously (Buenrostro et al., 2013). Peaks were called on the merged set of all ex vivo ATAC-seq reads using MACS2 (Zhang et al., 2008) and filtered to remove putative copy number altered regions. The number of reads/peak was determined for each sample using bedtools multicov, and the relative sequencing depth was estimated using a set of "housekeeping" peaks at transcription start sites of genes that were uniformly expressed across *TKO-mTmG* tumors. Differential accessibility was assessed using DESeq2 (Love et al., 2014). Unless otherwise stated, regions were called differentially accessible if the absolute value of the log2 fold change was 0.5 at an FDR <0.1. Visualizations of insertion tracks were smoothed by 150-bp sliding windows (20-bp step size).

Transcription Factor Occupancy

For each motif site, the distribution of fragments as a function of fragment size and midpoint position relative to the motif center was fit to a mixture model that finds the optimal mixture parameter between two generating distributions: (1) the distribution from an ideal "bound" model modulated by local sequence bias and (2) the distribution from an "unbound" model in which fragments are distributed based only on the local sequence bias. Local sequence bias was modeled (Schep et al., 2015), and the ideal "bound" model was found by aggregating reads around motif sites with high motif scores (>7.6) and divided by the expectation given sequence bias at those sites.

Nfib Knockdown and Overexpression

Stable Nfib knockdown cell lines were generated using lentiviral pLKO/PuroR vectors. Knockdown was confirmed by qRT-PCR and western blotting. Inducible and constitutive Nfib-expressing human and mouse cells were generated using lentiviral vectors.

Transplantation and Cell Culture Assays

For intravenous transplantation, 2×10^4 cells were injected into the lateral tail vein of NOD/SCID/γc (NSG) mice. For subcutaneous injection, 5×10^4 cells were resuspended in 100 μl PBS and mixed with 100 μl Matrigel (Corning, 356231) with four injection sites per mouse. Cell culture assays were performed using standard protocols.

RNA-Sequencing Expression Analysis

RNA-sequencing (RNA-seq) libraries were prepared using the Illumina TruSeq Kit (v.2), according to the manufacturer's instructions. RNA-seq reads were separately aligned to the mouse genome (mm10) using TopHat. Read counts within merged exons (RefSeq) were found and compared using DESeq2. Size factors were estimated from cqnnorm using the exon length and average GC content as covariates in the normalization.

ACCESSION NUMBERS

The accession number for the ATAC-seq, ChIP-seq, and RNA-seq sequencing data reported in this paper is GEO: GSE81258.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.05.052.

AUTHOR CONTRIBUTIONS

Conceptualization, S.K.D., D.Y., J.S., M.M.W., and W.J.G.; Formal Analysis, S.K.D., D.Y., M.W., C.S.K., A.J.C., J.S., M.M.W., and W.J.G.; Investigation, S.K.D., D.Y., C.-H.C., J.J.B., J.S.L., B.M.G., S.-H.C., A.N.S., J.B., C.H., M.A., C.S.K., and A.J.C.; Resources, M.W., C.S.K., K.-S.P., and W.J.G.; Writing – Original Draft: S.K.D., D.Y., J.S., M.M.W., and W.J.G.; Writing – Reviewing and Editing, all authors; Supervision, M.W., J.S., M.M.W., and W.J.G.; Project Administration: J.S., M.M.W., and W.J.G.; Funding Acquisition: J.S., M.M.W., and W.J.G.

ACKNOWLEDGMENTS

We thank Pauline Chu for technical assistance, Sean Dolan and Alexandra Orantes for administrative support, David Feldser, Nadine Jahchan, Youcef Ouadah, Mark Krasnow, Gerald Crabtree, Namita Bisaria, and members of the M.M.W., W.J.G., and J.S. laboratories for helpful comments. We thank the Stanford Shared FACS Facility. This work was supported by a Stanford Cancer Institute Cancer Biology Seed Grant (to M.M.W., W.J.G., and J.S.) and, in part, by a Stanford Cancer Institute support grant (NIH P30-CA124435), NIH grants (P50HG007735 and R01CA206540), a Baxter Foundation Faculty Scholar Grant, the Rita Allen Foundation, and the Human Frontier Science Program. S.K.D. was supported by a Stanford Biophysics training grant (T32 GM008294) and by the NSF GRFP. D.Y. was supported by a Stanford Graduate Fellowship and by a TRDRP Dissertation Award (24DT-0001). C.-H.C. was funded by an American Lung Association Fellowship. J.J.B. was supported by an NIH grant (F32-CA189659). B.M.G. was supported by a fellowship from the Hope Funds for Cancer Research (HFCR-15-06-07). J.S. is the Harriet and Mary Zelencik Scientist in Children's Cancer and Blood Diseases. W.J.G. is the scientific co-founder of Epinomics.

Received: March 28, 2016
Revised: May 9, 2016
Accepted: May 16, 2016
Published: June 30, 2016

REFERENCES

Alevizopoulos, A., Dusserre, Y., Tsai-Pflugfelder, M., von der Weid, T., Wahli, W., and Mermod, N. (1995). A proline-rich TGF-beta-responsive transcriptional activator interacts with histone H3. Genes Dev. *9*, 3051–3066.

Augustyn, A., Borromeo, M., Wang, T., Fujimoto, J., Shao, C., Dospoy, P.D., Lee, V., Tan, C., Sullivan, J.P., Larsen, J.E., et al. (2014). ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers. Proc. Natl. Acad. Sci. USA *111*, 14788–14793.

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

Bell, O., Tiwari, V.K., Thomä, N.H., and Schübeler, D. (2011). Determinants and dynamics of genome accessibility. Nat. Rev. Genet. *12*, 554–564.

Benos, P.V., Bulyk, M.L., and Stormo, G.D. (2002). Additivity in protein-DNA interactions: how good an approximation is it? Nucleic Acids Res. *30*, 4442–4451.

Betancourt, J., Katzman, S., and Chen, B. (2014). Nuclear factor one B regulates neural stem cell differentiation and axonal projection of corticofugal neurons. J. Comp. Neurol. *522*, 6–35.

Blomquist, P., Li, Q., and Wrange, O. (1996). The affinity of nuclear factor 1 for its DNA site is drastically reduced by nucleosome organization irrespective of its rotational or translational position. J. Biol. Chem. *271*, 153–159.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Burger, M., Glodek, A., Hartmann, T., Schmitt-Gräff, A., Silberstein, L.E., Fujii, N., Kipps, T.J., and Burger, J.A. (2003). Functional expression of CXCR4 (CD184) on small-cell lung cancer cells mediates migration, integrin activation, and adhesion to stromal cells. Oncogene *22*, 8093–8101.

Chang, C.-Y., Pasolli, H.A., Giannopoulou, E.G., Guasch, G., Gronostajski, R.M., Elemento, O., and Fuchs, E. (2013). NFIB is a governor of epithelial-melanocyte stem cell behaviour in a shared niche. Nature *495*, 98–102.

Chen, J. (2012). Regulation of tumor initiation and metastatic progression by Eph receptor tyrosine kinases. Adv. Cancer Res. *114*, 1–20.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science *328*, 1161–1164.

Dooley, A.L., Winslow, M.M., Chiang, D.Y., Banerji, S., Stransky, N., Dayton, T.L., Snyder, E.L., Senna, S., Whittaker, C.A., Bronson, R.T., et al. (2011). Nuclear factor I/B is an oncogene in small cell lung cancer. Genes Dev. *25*, 1470–1475.

Foley, K., Rucki, A.A., Xiao, Q., Zhou, D., Leubner, A., Mo, G., Kleponis, J., Wu, A.A., Sharma, R., Jiang, Q., et al. (2015). Semaphorin 3D autocrine signaling mediates the metastatic role of annexin A2 in pancreatic cancer. Sci. Signal. *8*, ra77.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

Grabowska, M.M., Elliott, A.D., DeGraff, D.J., Anderson, P.D., Anumanthan, G., Yamashita, H., Sun, Q., Friedman, D.B., Hachey, D.L., Yu, X., et al. (2014). NFI transcription factors interact with FOXA1 to regulate prostate-specific gene expression. Mol. Endocrinol. *28*, 949–964.

Gronostajski, R.M. (2000). Roles of the NFI/CTF gene family in transcription and development. Gene *249*, 31–45.

Harris, L., Genovesi, L.A., Gronostajski, R.M., Wainwright, B.J., and Piper, M. (2015). Nuclear factor one transcription factors: Divergent functions in developmental versus adult stem cell populations. Dev. Dyn. *244*, 227–238.

Hebbar, P.B., and Archer, T.K. (2003). Nuclear factor 1 is required for both hormone-dependent chromatin remodeling and transcriptional activation of the mouse mammary tumor virus promoter. Mol. Cell. Biol. *23*, 887–898.

Hou, J.M., Krebs, M.G., Lancashire, L., Sloane, R., Backen, A., Swain, R.K., Priest, L.J., Greystoke, A., Zhou, C., Morris, K., et al. (2012). Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. J. Clin. Oncol. *30*, 525–532.

Kalemkerian, G.P., Akerley, W., Bogner, P., Borghaei, H., Chow, L.Q., Downey, R.J., Gandhi, L., Ganti, A.K., Govindan, R., Grecula, J.C., et al.; National Comprehensive Cancer Network (2013). Small cell lung cancer. J. Natl. Compr. Canc. Netw. *11*, 78–98.

Ko, S.Y., Blatch, G.L., and Dass, C.R. (2014). Netrin-1 as a potential target for metastatic cancer: focus on colorectal cancer. Cancer Metastasis Rev. *33*, 101–113.

Lin, C.Y., Loven, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. Cell *151*, 56–67.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Meisterernst, M., Gander, I., Rogge, L., and Winnacker, E.L. (1988). A quantitative analysis of nuclear factor I/DNA interactions. Nucleic Acids Res. *16*, 4419–4435.

Meuwissen, R., Linn, S.C., Linnoila, R.I., Zevenhoven, J., Mooi, W.J., and Berns, A. (2003). Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. Cancer Cell *4*, 181–189.

Mohammad, H.P., Smitheman, K.N., Kamat, C.D., Soong, D., Federowicz, K.E., Van Aller, G.S., Schneck, J.L., Carson, J.D., Liu, Y., Butticello, M., et al. (2015). A DNA hypomethylation signature predicts antitumor activity of LSD1 inhibitors in SCLC. Cancer Cell *28*, 57–69.

Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. Genesis *45*, 593–605.

Nakazawa, K., Kurishima, K., Tamura, T., Kagohashi, K., Ishikawa, H., Satoh, H., and Hizawa, N. (2012). Specific organ metastases and survival in small cell lung cancer. Oncol. Lett. *4*, 617–620.

Osborne, J.K., Larsen, J.E., Shields, M.D., Gonzales, J.X., Shames, D.S., Sato, M., Kulkarni, A., Wistuba, I.I., Girard, L., Minna, J.D., and Cobb, M.H. (2013). NeuroD1 regulates survival and migration of neuroendocrine lung carcinomas via signaling molecules TrkB and NCAM. Proc. Natl. Acad. Sci. USA *110*, 6524–6529.

Peifer, M., Fernández-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T., et al. (2012). Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat. Genet. *44*, 1104–1110.

Rudin, C.M., Durinck, S., Stawiski, E.W., Poirier, J.T., Modrusan, Z., Shames, D.S., Bergbower, E.A., Guan, Y., Shin, J., Guillory, J., et al. (2012). Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat. Genet. *44*, 1111–1116.

Schaffer, B.E., Park, K.S., Yiu, G., Conklin, J.F., Lin, C., Burkhart, D.L., Karnezis, A.N., Sweet-Cordero, E.A., and Sage, J. (2010). Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. Cancer Res. *70*, 3877–3883.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. *25*, 1757–1770.

Schones, D.E., and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. Nat. Rev. Genet. *9*, 179–191.

Sethi, N., and Kang, Y. (2011). Unravelling the complexity of metastasis - molecular understanding and targeted therapies. Nat. Rev. Cancer *11*, 735–748.

Simon, J.M., Hacker, K.E., Singh, D., Brannon, A.R., Parker, J.S., Weiser, M., Ho, T.H., Kuan, P.F., Jonasch, E., Furey, T.S., et al. (2014). Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res. *24*, 241–250.

Steele-Perkins, G., Plachez, C., Butz, K.G., Yang, G., Bachurski, C.J., Kinsman, S.L., Litwack, E.D., Richards, L.J., and Gronostajski, R.M. (2005). The transcription factor gene Nfib is essential for both lung maturation and brain development. Mol. Cell. Biol. *25*, 685–698.

Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell *154*, 888–903.

Sung, M.-H., Guertin, M.J., Baek, S., and Hager, G.L. (2014). DNase footprint signatures are dictated by factor dynamics and DNA sequence. Mol. Cell *56*, 275–285.

Sung, M.-H., Baek, S., and Hager, G.L. (2016). Genome-wide footprinting: ready for prime time? Nat. Methods *13*, 222–228.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

van Meerbeeck, J.P., Fennell, D.A., and De Ruysscher, D.K.M. (2011). Small-cell lung cancer. Lancet *378*, 1741–1755.

Vanharanta, S., Shu, W., Brenet, F., Hakimi, A.A., Heguy, A., Viale, A., Reuter, V.E., Hsieh, J.J., Scandura, J.M., and Massagué, J. (2013). Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. Nat. Med. *19*, 50–56.

Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R.S., Stehling-Sun, S., Sabo, P.J., Byron, R., Humbert, R., et al. (2014). Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science *346*, 1007–1012.

Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat. Rev. Genet. *15*, 69–81.

Yang, J., and Weinberg, R.A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Dev. Cell *14*, 818–829.

Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr. Opin. Genet. Dev. *37*, 76–81.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

Zhou, W.-J., Geng, Z.H., Chi, S., Zhang, W., Niu, X.F., Lan, S.J., Ma, L., Yang, X., Wang, L.J., Ding, Y.Q., and Geng, J.G. (2011). Slit-Robo signaling induces malignant transformation through Hakai-mediated E-cadherin degradation during colorectal epithelial cell carcinogenesis. Cell Res. *21*, 609–626.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642–654.

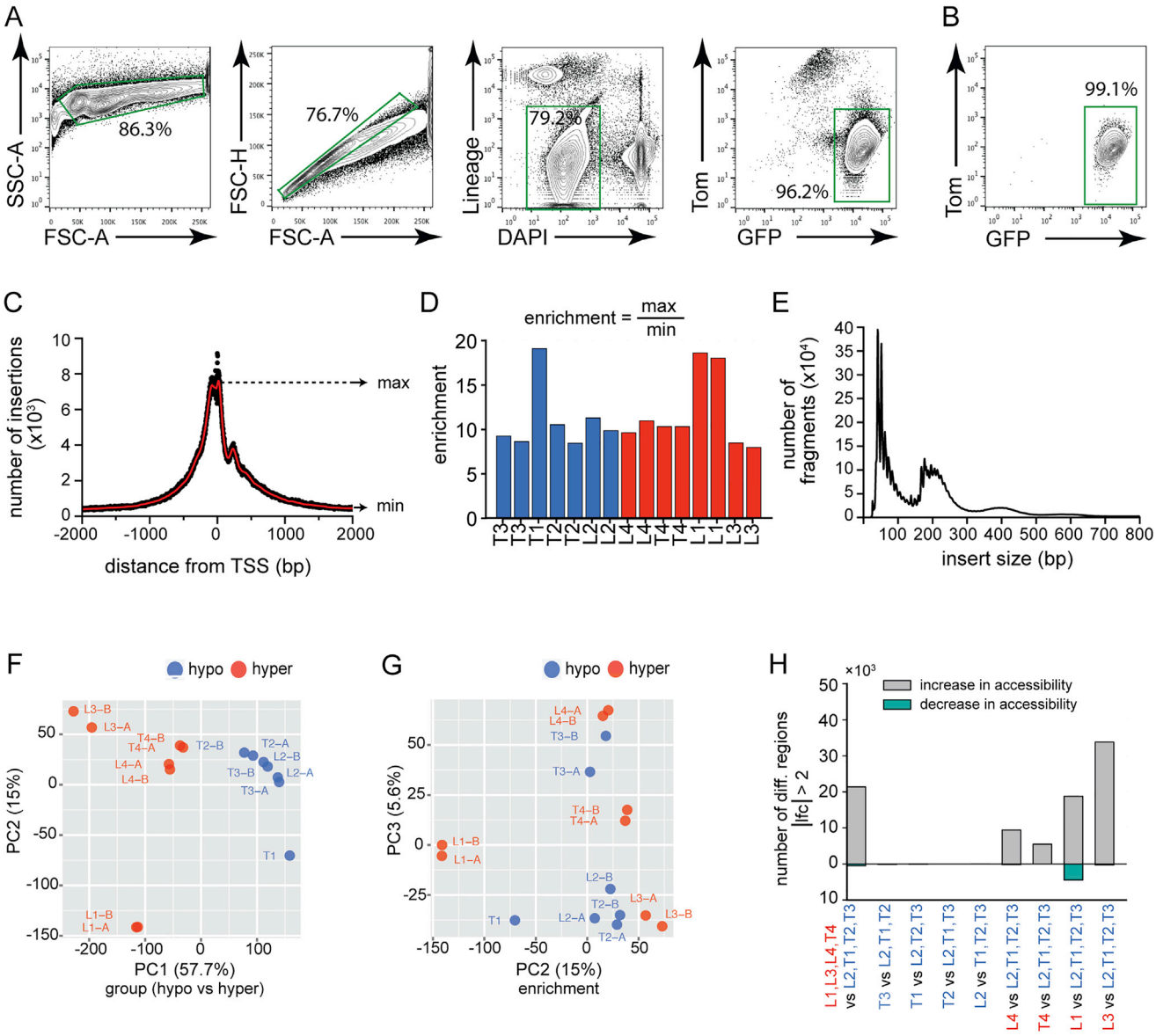


Figure S1. Analysis of Chromatin Accessibility on Purified SCLC Cells Highlights Two Distinct Chromatin States, Related to Figure 1

(A) FACS gating strategies for isolation of SCLC cells at high purity. FSC-A/FSC-gate enriches for single cells. DAPI stains dead cells. Lineage marker includes mouse CD45, CD31, Ter119 and F4/80. $GFP^{postive}Tom^{negative}$ SCLC cells are sorted.

(B) Post-sort flow cytometry confirms high purity of isolated cells. On average, > 97% purity of cancer cells was confirmed by genomic PCR for unrecombined alleles (data not shown).

(C) Representative plot of aggregate signal around transcription start sites (TSS) shows high enrichment for reads located around TSS. The small bump at around 250 bp is evidence of the +1 nucleosome, and confirms that our ATAC-seq libraries are sensitive to nucleosome-level chromatin structure. Red line shows smoothed profile using a running average in 50 bp windows.

(D) Enrichment scores for 15 SCLC ATAC-seq libraries. Enrichment score is defined as the fold enrichment at TSS compared to background, where background is defined as the minimum number of smoothed, aggregate signal in 4000 bp window centered at the TSS.

(E) Representative insert size distribution. Clear modulation of signal is evident for mono- and di- nucleosomes.

(F and G) Principal component analysis of regularized reads per peak of samples and technical replicates. The first principal component (F) separates the hyper- and hypo-accessible groups, the second principal component (F) and (G) is associated with samples L1 and T1, both of which have much higher enrichment than the rest of the samples. Principal component 3 does not correlate with any technical difference and only explains < 6% of the total variation.

(H) Number of differentially (diff.) accessible regions when comparing any of the hyper-accessible samples to the hypo-accessible samples. Many regions open, and very few close. Comparing any of the hypo-accessible samples to the other hypo-accessible samples uncovers essentially no regions that significantly change. lfc = $\log_2$ fold change.

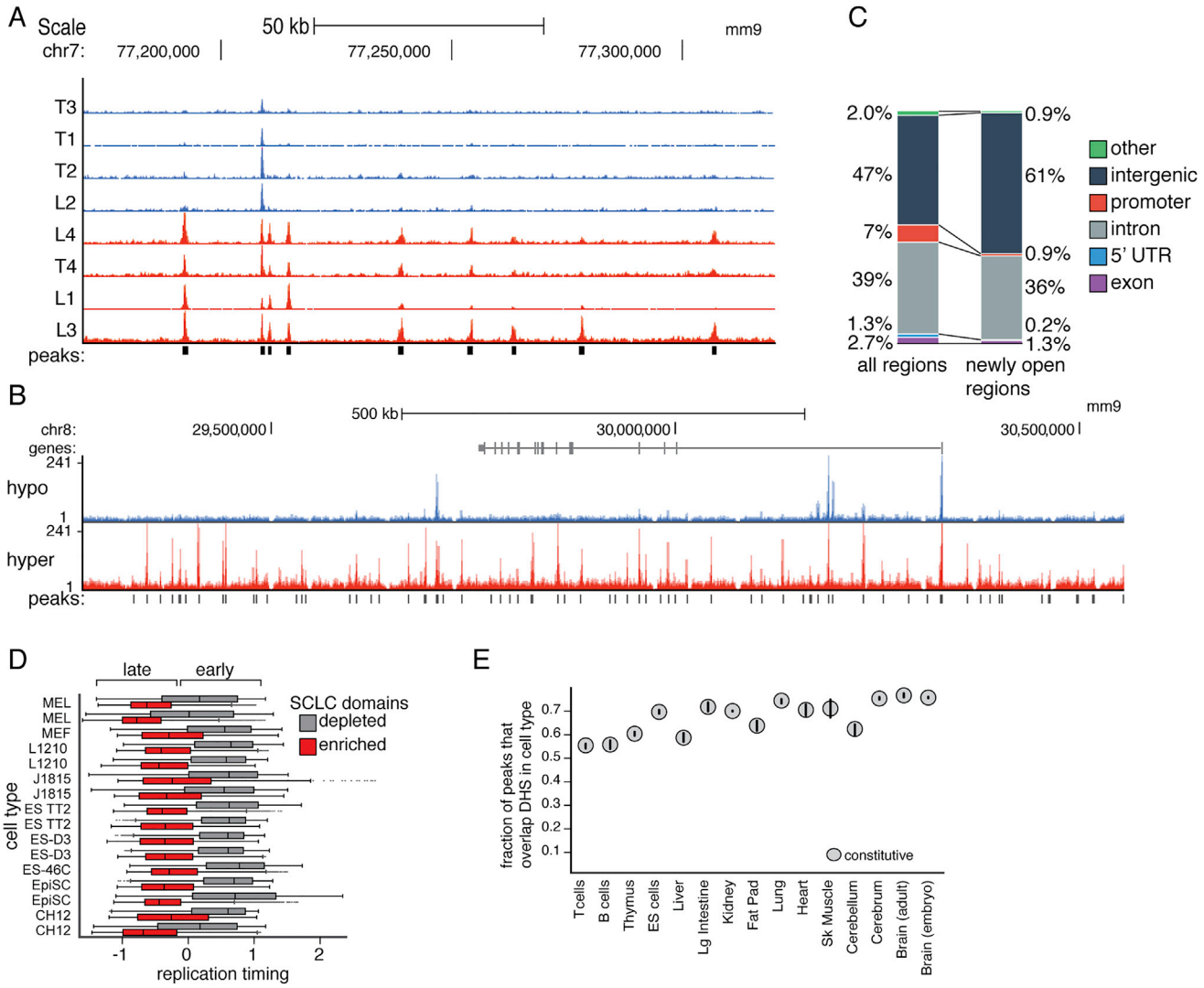


Figure S2. Changes in Chromatin Accessibility Are at Intergenic Late-Replicating Regions, Related to Figure 2

(A) Smoothed insertions of all 8 samples at an example locus (150 bp windows, 20 bp step size), showing that hyper-accessible samples have many regions of increased accessibility relative to hypo-accessible samples.

(B) Smoothed insertions of merged hypo- and hyper-accessible samples at a locus highly enriched for number of newly open peaks (150 bp windows, 20 bp step size).

(C) Annotations of all accessible regions compared to regions that are newly open in hyper-accessible samples. Promoter annotations are depleted in the newly open regions.

(D) Average replication timing of different cell types by repliChip (Stamatoyannopoulos et al., 2012) in 2 Mb windows in regions enriched or depleted for newly open peaks in SCLC. For all cell types, replication timing is significantly later in regions enriched for newly open peaks ($p < 1x10^{-12}$ by Mann Whitney U test).

(E) Overlap of constitutively accessible regions with DNase hypersensitive sites (DHS) from other cell types. Shown is the mean overlap with DHS peak calls. Black bars represent 95% confidence intervals on the mean overlap given technical replicates of DHS. Constitutively open peaks tend to be shared with many tissue types (50%–70% for all cell types).

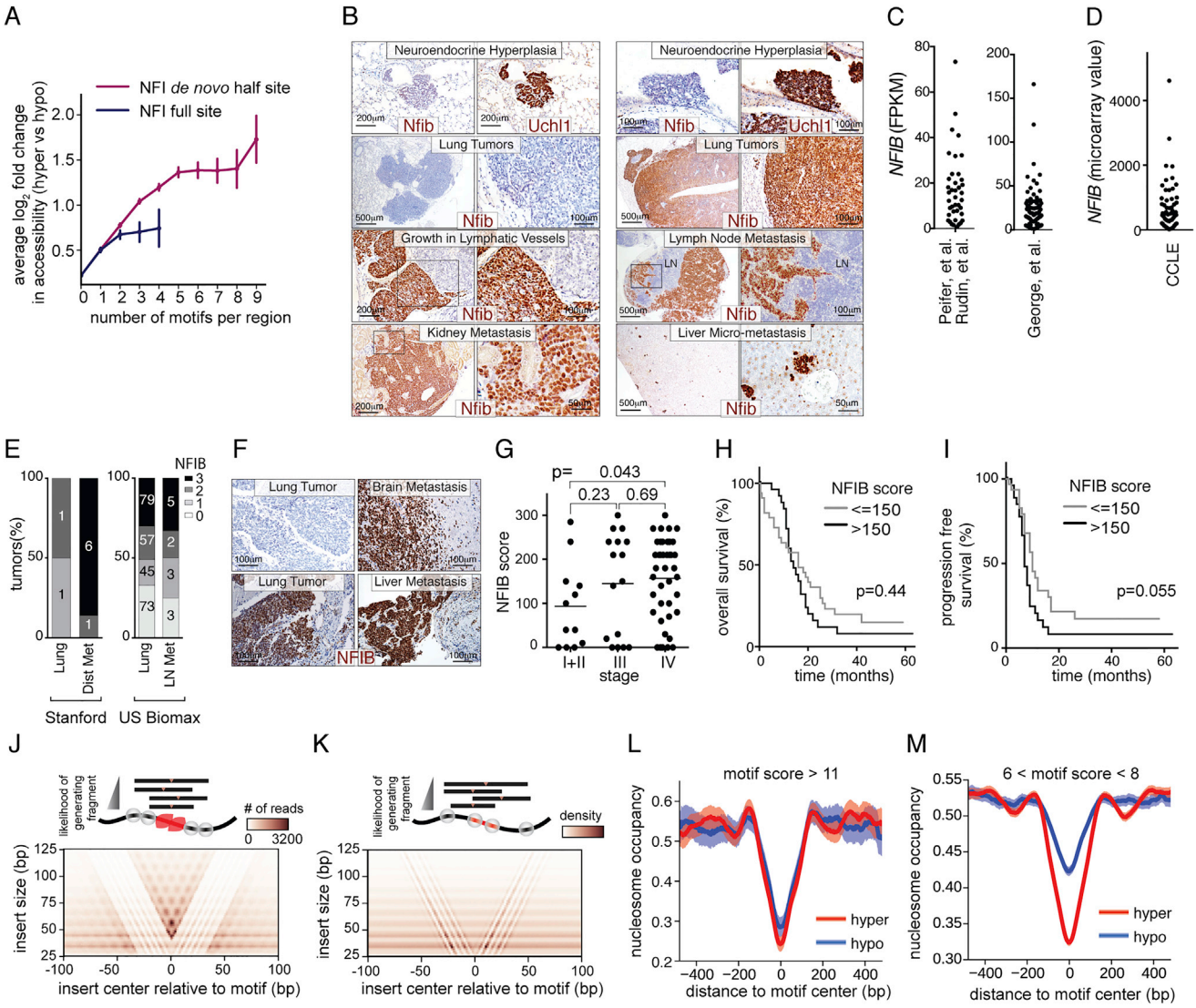


Figure S3. Nfib Is Highly Expressed in Advanced SCLC in a Mouse Model and Has Diverse Expression in Human SCLC, Related to Figure 3

(A) Average $\log_2$ fold change in accessibility of regions is modulated by the number of NFI half sites or full sites. (Purple) Number of half sites versus accessibility change of regions with no full sites ($r = 0.33$, $p < 1\times10^{-12}$). (Blue) Number of full sites versus accessibility change of regions with no half sites ($r = 0.18$, $p = 5\times10^{-135}$). Error bars are 95% confidence intervals on the mean.

(B) Representative IHC for Nfib and the SCLC marker Uchl1 on neuroendocrine hyperplasia (2 months after tumor initiation) as well as lung tumors, lymphatic invasion in the lung, and metastases from $p53^{f/f};Rb1^{f/f};p130^{f/f};R26^{mTmG}$ mice (6-8 months after tumor initiation). At the histological level, Nfib-negative tumor areas have a mixed histology of frequent well-polarized rosettes and solid sheets of cells, and uniform oval nuclei, while Nfib-positive tumor areas are sheets of cells with no rosettes or polarized cells, frequent necrosis, cells with scant cytoplasm, mildly increased nuclear size, high nuclear/cytoplasmic ratio, nuclear molding, higher mitotic index compared to the Nfib-negative area, and numerous discohesive cells along the borders.

(C and D) NFIB mRNA expression (in FPKM) is diverse in human SCLC and very high in a subset of tumors. Each dot represents a sample. (C) RNA-seq data from Peifer et al. (2012) and Rudin et al. (2012) (13+28 samples) and George et al. (2015) (81 samples). (D) Microarray data from the cancer cell line encyclopedia (CCLE; Barretina et al., 2012) on 52 human SCLC cell lines.

(E) Expression of NFIB protein detected by IHC on human SCLC primary tumors and metastases. The number of human SCLC samples in each group is indicated. Tumor sections from Stanford include primary tumors from lung and distant metastases from multiple organs (Dist. Met), and tissue microarrays from USBiomax include primary tumors and lymph node metastases (LN).

(F) Representative IHC for NFIB on human SCLC primary lung tumors (one negative and one positive example are shown) and brain and liver metastases. Scale bars = 100 μm.

(G) Z-score of NFIB expression in staged SCLC patients. Stage IV SCLC patients have a trend toward having higher NFIB expression than Stage I-II patients, $p < 0.05$.

(H and I) Patients with SCLC that express higher NFIB have a trend toward (H) shorter overall survival (time since diagnosis), and (I) progression-free survival (time since treatment start). p by log-rank (Mantel-Cox) test.

(legend continued on next page)

(J and K) Occupancy at individual NFI sites was inferred from ATAC-seq Vplots that show insert sizes and positions relative to NFI full sites. At every individual site, ATAC-seq inserts were compared to the (J) aggregate Vplot of all inserts around NFI sites (full sites, score > 7.6) and the (K) theoretical density of inserts given a model for Tn5 insertion bias. Cartoon shows Tn5 (gray circles) and NFI dimer (red), black bars represent ATAC-seq inserts.
(L and M) Nucleosome depletion around NFI full sites at different motif score thresholds in the newly open peaks. Higher motif scores are more likely to be bound by Nfib. At high affinity NFI sites (L), nucleosome depletion is approximately the same in the hypo- and hyper-accessible samples, whereas at intermediate affinity sites (M), nucleosomes are more depleted in hyper-accessible than hypo-accessible samples. Shaded area represents 95% confidence interval on the mean occupancy.

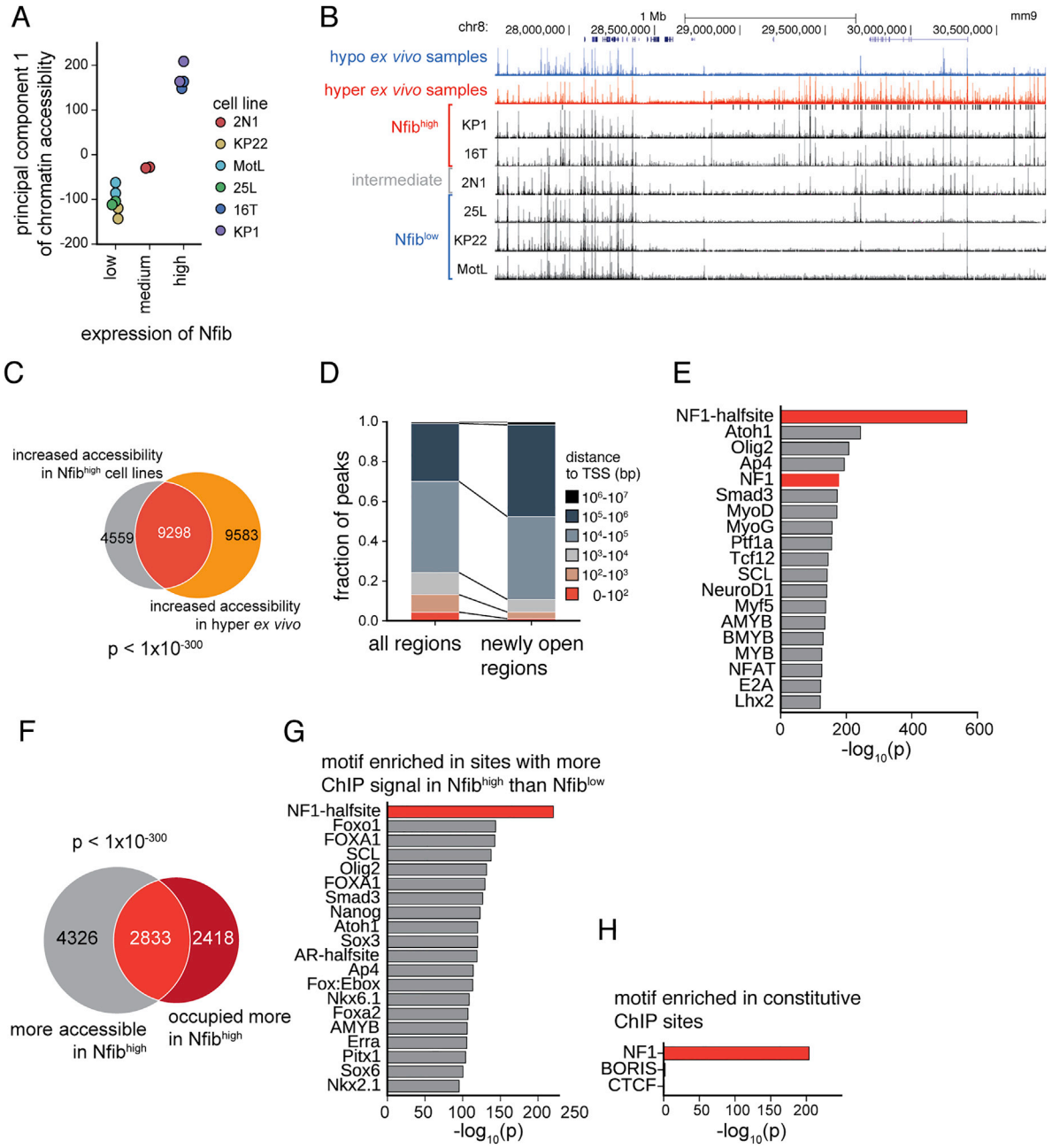


Figure S4. Differential Open Regions in Nfib^{high} Cell Lines Have Similar Features to Newly Open Regions in Hyper-Accessible Ex Vivo Samples, Related to Figure 4

(A) First principal component of variation of chromatin accessibility, plotted against expression level of Nfib based on western blot (Figure 4A). Amount of Nfib expression is highly correlated with chromatin accessibility state. First principal component explains 40% of the variance in the chromatin accessibility of the cell lines.

(B) Smoothed insertion tracks of cell lines shows cell lines have similar accessibility profiles to ex vivo samples. Peaks are scaled such that the 95th percentile of reads per bp in peaks is proportional to max, and smoothed by 150 bp windows, 20 bp step size. Black bars under ex vivo samples show the newly open peaks in the hyper-accessible samples.

(C) High overlap of differentially accessible peaks in cell lines and in ex vivo samples. p by hypergeometric test.

(D) Peaks that are newly accessible in cell lines are largely gene distal relative to all peaks.

(E) Motif enrichment in newly open peaks compared to unchanged peaks. Known motif enrichments indicate a strong enrichment for NFI full sites and half sites.

(F) Overlap of peaks that are more accessible in Nfib^{high} cell lines and have higher ChIP signal in Nfib^{high} cell lines. p by hypergeometric test.

(G) Motif enrichment in ChIP peaks with differential ChIP signal compared to those with the constitutive ChIP signal.

(H) Motif enrichment in ChIP peaks with constitutive ChIP signal compared to those with the differential ChIP signal.

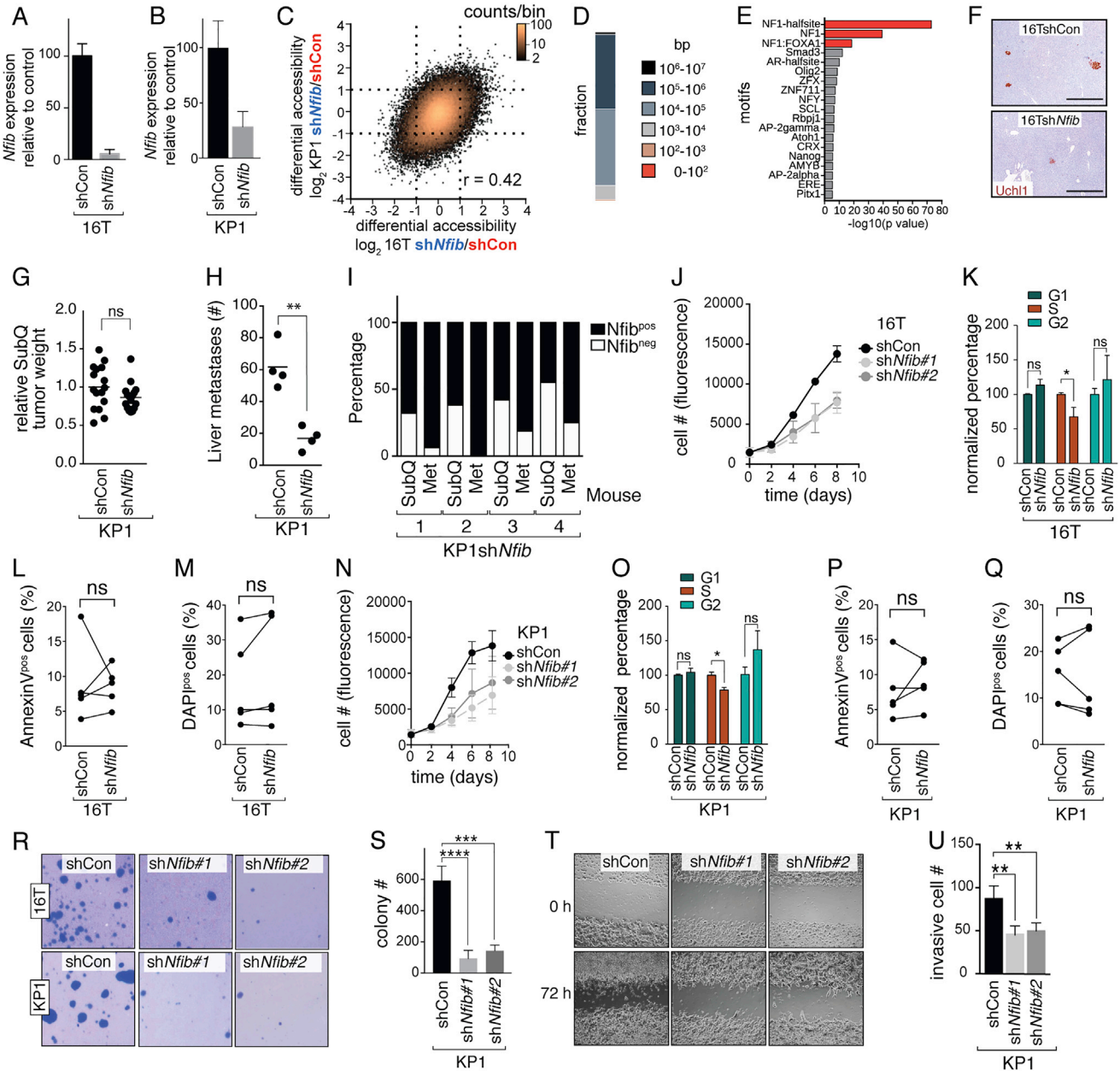


Figure S5. Nfib Is Required to Sustain Chromatin Accessibility, Metastatic Ability, Clonal Growth Potential, and Migratory Ability, Related to Figure 5

(A and B) Efficient *Nfib* knockdown in 16T (A) and KP1 (B) cells, quantified by qPCR, using *Gapdh* as an endogenous control for normalization. Mean ± SD of triplicate is shown. shCon = shControl.

(C) Regions that change in accessibility with *Nfib* knockdown in 16T and KP1 cells generally correlate (r = 0.42).

(D) Distance to nearest transcription start site (TSS) of regions that decrease in accessibility in *Nfib* knockdown cells.

(E) Motif enrichment in regions that close with *Nfib* knockdown compared to regions that do not change.

(F) Immunohistochemistry for the SCLC marker Uchl1 highlights metastases seeded from 16TshNfib or 16TshControl subcutaneous tumors. Scale bars = 500mm. Mice with shNfib subcutaneous tumors have fewer liver metastases (quantification is in Figure 5F).

(G) *Nfib* knockdown in KP1 cells does not significantly affect subcutaneous (SubQ) tumor growth. Each dot represents a subcutaneous tumor and the line indicates the mean. ns = not significant.

(H) Mice with KP1shNfib subcutaneous tumors have fewer liver metastases. Each dot represents a mouse and the line is the mean. **p < 0.004 by Student's test.

(I) Quantification of the percent of KP1shNfib subcutaneous tumor area (SubQ) that expresses Nfib ($Nfib^{pos}$) and the percent of liver metastases (Met) that are Nfib-positive. Data from the 4 mice with shNfib tumors that developed liver metastases indicate most metastases have escaped *Nfib* knockdown. The percent of subcutaneous tumors that are $Nfib^{pos}$ is less than the percent of metastases that are $Nfib^{pos}$ (p < 0.008 by Student's test).

(legend continued on next page)

(J–M) *Nfib* knockdown in 16T cells reduces proliferation but does not consistently change cell death under standard floating culture conditions. (J) AlamarBlue growth assay. Mean ± SEM of 3 replicate experiments is shown. (K) BrdU&7-AAD staining for cell cyle, (L) AnnexinV staining for apoptotic and dead cells, and (M) DAPI staining for dead cells are shown. Mean ± SD of 3-6 replicate experiments is shown. In (L) and (M), pairwise shCon and shNfib samples analyzed on the same day are connected by a line. ns = not significant, *$p < 0.05$ by Student's t test.

(N–Q) Nfib knockdown in KP1 cells reduces proliferation but does not consistently change cell death under standard culture conditions. (N) AlamarBlue growth assay. Mean ± SEM of 3 replicate experiments is shown. (O) BrdU&7-AAD staining for cell cyle, (P) AnnexinV staining for apoptotic and dead cells, and (Q) DAPI staining for dead cells are shown. In (M) and (N), error bars indicate standard deviation. In (O) and (P), pairwise shCon and shNfib samples analyzed on the same day are connected by a line. Mean ± SD is shown. ns = not significant, *$p < 0.05$ by Student's t test.

(R) Representative images of 16T (top) and KP1 (bottom) soft agar colonies that form from shCon and shNfib cells.

(S) Quantification of the number of colonies from KP1 shCon and shNfib is shown as mean ± SD. ****$p < 0.0001$, ***$p < 0.0005$.

(T) Representative images of the migration of 16T shCon and sh*Nfib*. The migration of cells into the cell-free gap after 72 hr is greatly reduced by knockdown of *Nfib*. Quantification is shown in Figure 5L.

(U) Quantification of cancer cell migration of KP1 shCon and sh*Nfib*. *Nfib* knockdown reduces SCLC migration. Mean ± SD is shown. **$p < 0.005$.

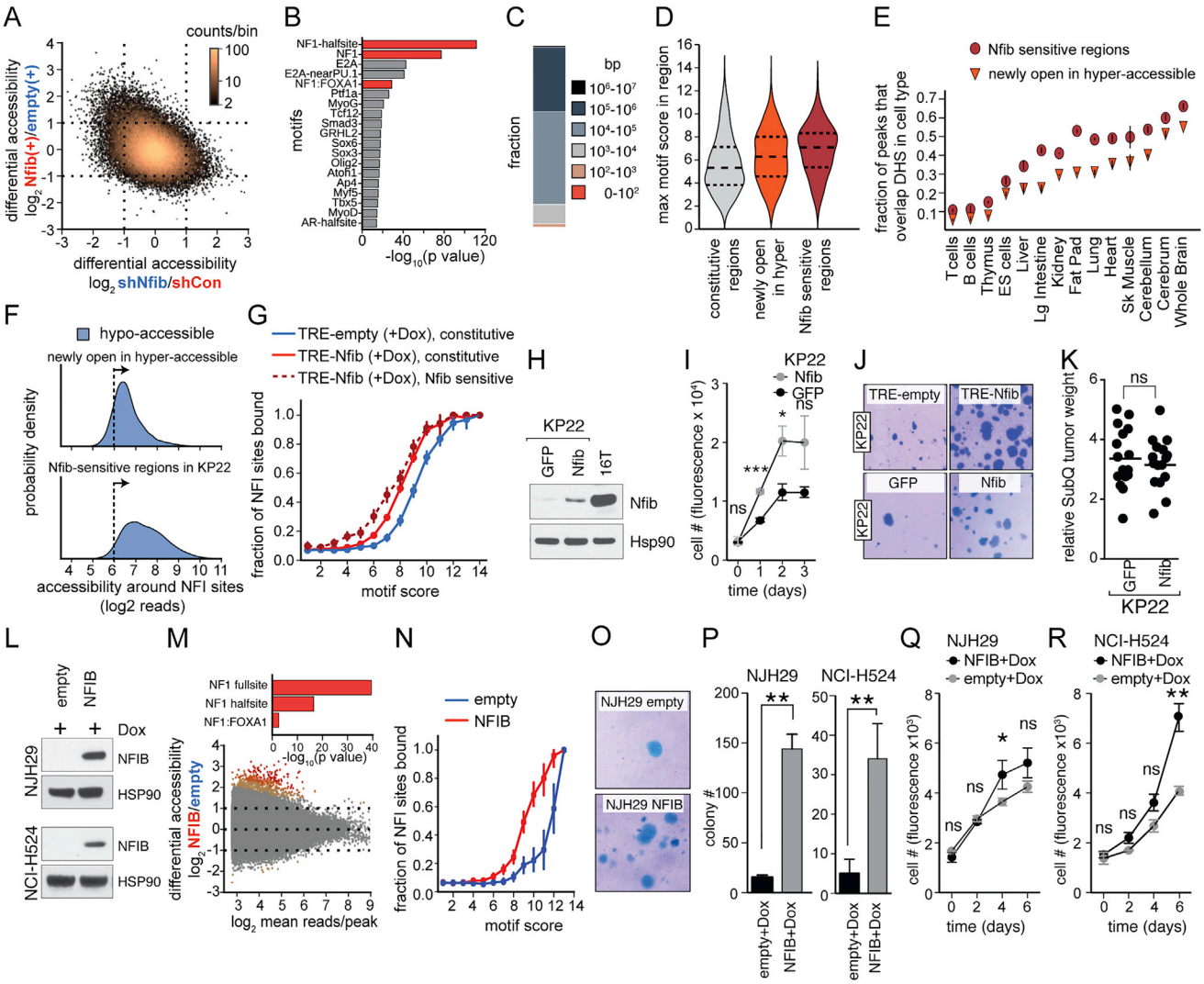


Figure S6. Nfib Expression Increases Clonal Growth Ability in Mouse and Human Cell Lines, Related to Figure 6

(A) Highly significant overlap between regions that close with *Nfib* knockdown and those that open with Nfib expression (r = - 0.37; p < 1e-300 by hypergeometric test).

(B) Motif enrichment in regions that open in response to Nfib expression (showing top 20).

(C) Distance to nearest transcription start site (TSS) of regions that are sensitive to Nfib overexpression. Very few promoter peaks are sensitive.

(D) Motif score in regions that do not change in accessibility, regions that are specifically more open in hyper-accessible ex vivo samples, and regions that are sensitive to Nfib overexpression in KP22 cell line. Sites that are directly sensitive to Nfib expession have higher motif scores on average than those that are newly open in hyperaccessible samples.

(E) Overlap with DNase hypersensitive sites (DHS) found in other cell types of the regions that are specifically more open in hyper-accessible ex vivo samples, and regions that are sensitive to Nfib expression in KP22 cell line.

(F) Accessibility (log2 read count) in 1000 bp windows around NFI motif sites in hypo-accessible ex vivo samples. Dotted line indicates the average "background" number of reads around sites that were not within accessible regions found in SCLC. Sites that are newly open in hyper-accessible samples were also slightly open in the hypo-accessible samples.

(G) Occupancy of NFI full sites in KP22 with and without induction, in bins of motif score. The TRE-Nfib with doxycycline induction is divided into sites within constitutively open peaks and sites within Nfib-sensitive peaks.

(H) Expression of Nfib from a constitutive expression vector. The Nfib^{high} cell line 16T is shown for comparison and Hsp90 serves as a loading control.

(I) Nfib expression increases proliferation of KP22 cells under standard suspension growth condition. Mean ± SEM of triplicate experiments is shown. ns = not significant, ***p < 0.001, *p < 0.05 by Student's t test.

(J) Representative images of soft agar colonies from KP22 cell with inducible (top) and constitutive Nfib expression (bottom).

(K) Nfib expression in KP22 cells does not significantly affect subcutaneous (subQ) tumor growth. Each dot represents a subcutaneous tumor and the bar indicates the mean. ns = not significant by Student's t test.

(L) Inducible expression of NFIB in two NFIB low human SCLC cell lines (NJH29 and NCI-H524) with doxycycline (Dox). HSP90 serves as a loading control.

(legend continued on next page)

(M) Differential accessibility driven by NFIB expression in NFIB^{low} human SCLC cell lines. Merged data from NJH29 and NCI-H524 is shown. (Inset) The three significant motif enrichments in regions that open in human SCLC cell lines.
(N) Occupancy of NFI full sites in human cell lines expressing NFIB and control. Sites are more occupied with NFIB expression, though the overall occupancy is low.
(O and P) NFIB expression increases soft agar colony formation in human cell lines. (O) Representative images of soft agar colonies from human NJH29 cells with NFIB expression. (P) Quantification of the colony number is the mean ± SD of triplicate experiments. **$p < 0.01$ by Student's t test.
(Q and R) NFIB expression increased proliferation of human cell lines under standard suspension growth conditions. Mean ± SEM of triplicate experiments is shown. (Q) NJH29 cells. (R) NCI-H524 cells. ns = not significant, *$p < 0.05$, **$p < 0.01$ by Student's t test.

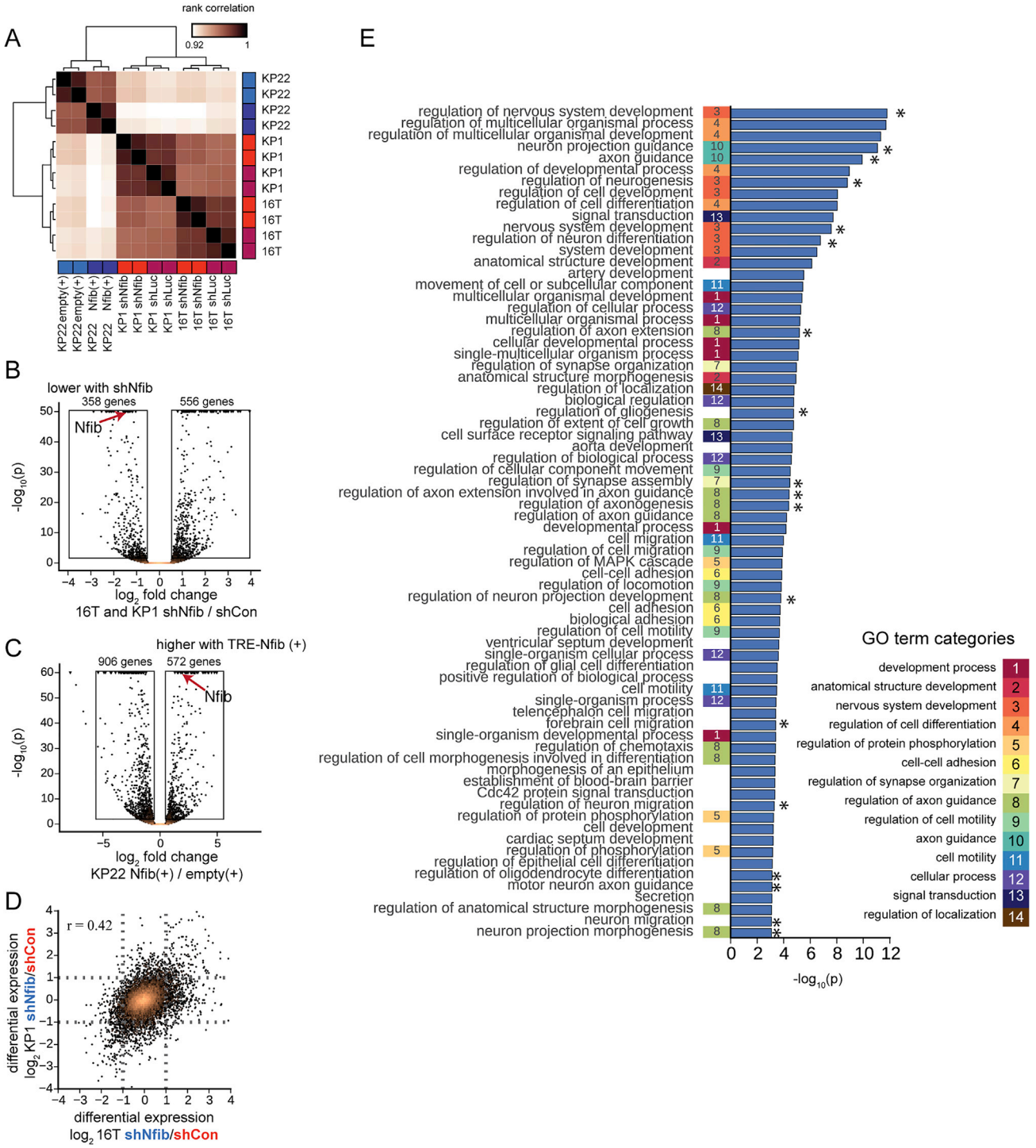


Figure S7. Gene Expression Changes in Response to Nfib Implicate Promotion of Neuronal State, Related to Figure 7

(A) Correlation of regularized counts within exons of RefSeq genes in 3 cell lines, Nfib is either upregulated or downregulated, and two technical replicates are shown for each. The two Nfib^{high} cell lines (16T, KP1) have a more similar program of gene expression than the one Nfib^{low} cell line (KP22)

(B and C) Significance of change versus the $\log_2$ fold change in expression upon knockdown (B) and overexpression (C) of Nfib. Nfib is indicated with red arrow.

(D) Differential expression of genes of two Nfib^{high} cell lines upon knockdown of Nfib are correlated (r = 0.42).

(E) GO term enrichments for genes that were upregulated with Nfib overexpression and downregulated with *Nfib* knockdown, versus genes that did not change in expression. Marked are 14 categories based on the correlation of genes giving rise to each GO term (Supplemental Experimental Procedures). Black asterisks show GO terms associated with neural phenotypes.

Cell, Volume *166*

Supplemental Information

Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility

Sarah K. Denny, Dian Yang, Chen-Hua Chuang, Jennifer J. Brady, Jing Shan Lim, Barbara M. Grüner, Shin-Heng Chiou, Alicia N. Schep, Jessika Baral, Cécile Hamard, Martine Antoine, Marie Wislez, Christina S. Kong, Andrew J. Connolly, Kwon-Sik Park, Julien Sage, William J. Greenleaf, and Monte M. Winslow

SUPPLEMENTAL METHODS

Mouse model, tumor initiation, and cancer cell isolation

Trp53^{flox}, *Rb1*^{flox}, *p130*^{flox} mice have been described (Schaffer et al. 2010; Sage et al. 2003; Jonkers et al. 2001). We further incorporated a *R26*^{mTmG} allele to enable high-purity isolation of cancer cells. Adeno-CMV-Cre (Ad-Cre) was from the University of Iowa Vector Core. Mice were infected intratracheally with 4 x10^8 Ad-Cre to initiate tumors (Schaffer, 2010; DuPage, 2009). To increase the number of neuroendocrine hyperplasia at early time points (2-3 months after tumor initiation), mice were infected mice with 2 x10^{10} Ad-Cre. Primary tumors and metastases were dissected and dissociated using collagenase IV, dispase, and trypsin at 37^{o}C for 30 minutes. After dissociation the samples are continually on ice, in contact with ice-cold solutions, and in the presence of 2mM EDTA and 1U/ml DNase to prevent aggregation. Cells were stained with antibodies to CD45 (30-F11), CD31 (390), F4/80 (BM8), and Ter119 (all from BioLegend) to exclude hematopoietic and endothelial cells. DAPI was used to exclude dead cells. Cell sorting was performed on FACSAria^{TM} sorters (BD Biosciences).

Histology and immunohistochemistry

Lung, liver and other tumor samples were fixed in 4% formalin and paraffin embedded. Hematoxylin and Eosin staining was performed using standard methods. Percent tumor area was calculated using ImageJ. For IHC, we used antibodies to Nfib (1:1000, abcam ab186738); Uchl1 (1:500, sigma HPA005993), and GFP(1:500, abcam ab6673).

Human SCLC Immunuohistochemistry

Human SCLC samples were obtained from the Stanford Pathology files. TMAs used for analysis were LC814a, LC817, LC818, LC802a, LC10010b, LC245 from US Biomax. Immunohistochemical stain for NFIB (abcam ab186738, 1:1000, citrate) was performed on 4 μM-thick sections cut from the formalin-fixed paraffin-embedded tissue blocks. NFIB expression was scored by a board-certified pathologist on a three point scale as follows: 0 = negative or weak staining of less than 10% cells, 1 = weak staining of more than 10% cells, 2 = moderate intensity staining, 3 = strong intensity staining.

French Group survival method

We selected patients who had diagnosis of small cell lung cancer at Hôpital TENON, Assistance Publique-Hôpitaux de Paris, France, between January 2010 and January 2013, from the database of Pathology Service (code "small cell lung cancer"). 73 consecutive patients were identified, whose tissue was available for immunohistochemistry. For each patient, medical file and clinical characteristics were available.

Clinical and biological characteristics of 73 consecutive patients with diagnosed small-cell lung cancer at Hôpital Tenon between 2010 and 2013 are shown:

Patient characteristics (n=73)	Variables : % (n) median [1st quartile-3rd quartile]
Sex Male Female	 61.6% (45) 38.4% (28)
Median age at diagnosis (years)	63.5 [58.2 –73.1]

Tobacco Non-smoker or former smoker Active smoker	 2.7% (2) 97.3% (71)
Histology « pure » SCLC « mixed » SCLC	 83.6% (61) 16.4% (12)
TNM Staging I II III IV **Extension** Localised Extended	 12.3% (9) 5.5% (4) 21.9% (16) 60.3% (44) 34.2% (25) 65.8% (48)
First treatment Surgery Radio-chemotherapy Chemotherapy No treatment	 9.6% (7) 8.2% (6) 76.7% (56) 5.5% (4)
Survival (months) Progression-free survival Overall survival	 7 [3 – 11] 13 [5 –22]
Immunohistochemistry Expression of NFIB Z score : % x (intensity) Expression of HES-1 (%)	 150 [40 - 240] 0 [0 – 5]

Statistics

All data are in (%) or median [first quartile – third quartile] unless otherwise specified. HES1 and NFIB were semi-quantitative variables. We used Mann-Whitney test to compare two quantitative variables, and Kruskall-Wallis test to compare three or more quantitative variables. We used Kaplan-Meier with log-rank for univariate survival analyses and Cox model with log-rank test for multivariate survival analyses. For all the rest of the in vitro and in vivo experiments, we used Student's t test unless otherwise noted. Analyses were performed using Prism 5.0 (Graphpad Software Inc., San Diego, CA) and SPSS22.0 (SPSS Inc., Chicago, III).

Immunofluorescence

FACS-sorted cancer cells from primary tumors, DTCs, and liver metastases were cytospun to glass slides at 500rpm for 5 minutes. Cells were fixed with 4% PFA for 15min, and stained with *Nfib* antibody (Abcam, ab186738) 1:1000 and goat anti rabbit secondary antibody (Invitrogen). For imaging, membrane GFP staining was confirmed to indicate a DTC and Nfib expression was checked through the far-red channel using fluorescence scope (Leica). Nfib staining was quantified by counting directly under the scope. On average, we quantified 30-100 cells per sample based on how many cells we harvested.

ATAC-seq library preparation and sequencing

ATAC-seq libraries were generated as described in (Buenrostro et al. 2013). In short, 50K cells were isolated from sorted populations of *ex vivo* samples or directly from cell culture. These were washed in PBS once and then resuspended in lysis buffer (10mM Tris-HCl, pH 7.5, 10mM NaCl, 3mM MgCl2, an 0.1% IGEPAL CA-630) and centrifuged for 10 minutes at 500x g. Transposition was initiated by adding 1x TD buffer (Illumina) with 2.5ul of Tn5 transposase

(Illumina) in 50ul total. Transposition was allowed to proceed for 30 minutes at 37 degrees in a thermomixer shaking at 300 rpm. Libraries from human cell lines often had high percentages of mitochondrial reads; we reduced this by adding Tween-20 to the lysis and transposition buffers at 0.1%.

Transposition reactions were cleaned up with MinElute columns. Libraries were amplified with Illumina Nextera sequencing primers in the same manner as in (Buenrostro et al. 2013). Libraries were quantified by qPCR against PhiX standard (Illumina).

Libraries were sequenced on an Illumina Hiseq (Elim Biosciences). Runs consisted of three cycles of 50 x 8 x 50 bp. When samples were in multiple Hiseq lanes, samples within any comparison group (*ex vivo* samples, cell lines, or within modified cell lines) were always present at approximately equal fractions in each lane.

Primer sequences are given below:

Ad1_noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT

Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

ATAC-seq library read counts

Ex vivo samples

sample	total read count	aligning to any chr (%)	aligning to chrM (%)	within peaks (%)
L1-A	7.43E+07	97.52%	28.21%	19.59%
L1-B	7.63E+07	96.35%	26.85%	19.51%
L2-A	4.39E+07	95.75%	49.60%	5.08%
L2-B	3.70E+07	97.60%	46.10%	5.60%
L3-A	3.34E+07	97.38%	20.05%	10.54%
L3-B	4.03E+07	97.18%	18.14%	10.43%
L4-A	3.43E+07	97.88%	19.78%	10.71%
L4-B	5.77E+07	97.57%	23.29%	10.74%
T1	6.79E+07	97.28%	32.66%	15.56%
T2-A	4.79E+07	96.68%	37.97%	6.84%
T2-B	2.79E+07	97.14%	33.66%	7.30%
T3-A	3.14E+07	97.17%	24.73%	10.91%
T3-B	4.48E+07	97.37%	15.74%	10.64%
T4-A	4.30E+07	95.97%	26.86%	8.17%
T4-B	4.56E+07	97.58%	28.85%	7.51%

Cell lines

sample	total read count	aligning to any chr (%)	aligning to chrM (%)	within peaks (%)
16T-A	1.97E+07	96.42%	37.12%	16.42%
16T-B	1.60E+07	97.39%	22.45%	21.35%
25L-A	3.02E+07	95.86%	31.84%	12.80%
25L-B	2.29E+07	97.49%	50.45%	8.87%
2N1-A	2.27E+07	95.13%	47.80%	8.51%
2N1-B	2.22E+07	96.58%	53.50%	5.47%
KP1-A	1.53E+07	90.67%	39.55%	7.97%
KP1-B	1.95E+07	89.63%	29.70%	13.51%
KP22-A	1.36E+07	87.39%	24.10%	13.41%
KP22-B	1.74E+07	84.32%	23.91%	12.51%
MotL-A	2.13E+07	75.13%	34.36%	4.71%

MotL-B	2.23E+07	76.39%	15.04%	8.85%

Modified cell lines

sample	total read count	aligning to any chr (%)	aligning to chrM (%)	within peaks (%)
16T-sh2-A	2.93E+07	98.50%	16.65%	10.20%
16T-sh2-B	2.80E+07	98.29%	17.85%	9.04%
16T-shLuc-A	1.49E+07	98.32%	13.35%	16.78%
16T-shLuc-B	4.71E+07	98.22%	15.48%	12.84%
KP1-sh2-A	1.28E+07	86.45%	29.84%	9.61%
KP1-sh2-B	1.52E+07	86.10%	35.70%	8.83%
KP1-shLuc-A	5.60E+06	88.71%	41.44%	6.80%
KP1-shLuc-B	1.21E+07	91.65%	38.64%	8.03%
KP22-empty-1d-A	2.09E+07	98.74%	44.43%	4.86%
KP22-empty-1d-B	1.90E+07	98.73%	37.20%	5.78%
KP22-empty-2d-A	1.72E+07	98.39%	29.26%	6.46%
KP22-empty-2d-B	2.21E+07	98.24%	22.72%	7.02%
KP22-empty-1w-A	1.82E+07	98.84%	36.65%	6.95%
KP22-empty-1w-B	1.64E+07	98.76%	36.09%	6.18%
KP22-empty-2w-A	2.20E+07	98.86%	46.03%	5.85%
KP22-empty-2w-B	1.90E+07	98.89%	45.49%	6.97%
KP22-nfib-1d-A	1.93E+07	98.69%	35.14%	11.05%
KP22-nfib-1d-B	1.88E+07	98.68%	34.57%	8.98%
KP22-nfib-2d-A	4.69E+07	98.57%	35.59%	8.13%
KP22-nfib-2d-B	3.36E+07	98.46%	34.97%	8.62%
KP22-nfib-1w-A	1.88E+07	98.81%	50.83%	6.19%
KP22-nfib-1w-B	1.86E+07	98.84%	53.03%	5.74%
KP22-nfib-2w-A	1.57E+07	98.75%	44.11%	6.69%
KP22-nfib-2w-B	1.66E+07	98.94%	50.88%	6.91%

Human overexpression

	total read count	aligning to any chr (%)	aligning to chrM (%)	within peaks (%)
NJH29_empty_A	2.18E+07	96.65%	34.01%	13.67%
NJH29_nfib_A	2.30E+07	97.23%	42.47%	11.46%
NCI-H524_empty_A	1.30E+07	97.91%	9.26%	19.94%
NCI-H524_empty_B	1.10E+07	97.98%	13.11%	21.28%
NCI-H524_empty_C	2.35E+07	96.94%	8.74%	22.20%
NCI-H524_nfib_A	1.21E+07	98.22%	11.73%	20.10%
NCI-H524_nfib_B	9.38E+06	98.09%	11.70%	19.51%
NCI-H524_nfib_C	2.24E+07	97.14%	43.71%	11.40%

Processing reads

Fastq files were aligned to the mouse genome (mm9) with bowtie2 (version 2.0.0-beta7 for *ex vivo* and unmodified cell lines and version 2.1.0 for cell line knockdown/overexpression and controls) with all default settings except a maximum fragment length of 2000 bp enforced (command line parameter –X 2000) (Langmead & Salzberg 2012). Bam files were processed by removing reads that: 1) did not align to the nuclear genome, 2) aligned to mitochondrial genome, or 3) were PCR duplicates. The last step was done using picard tools MarkDuplicates program (version picard-tools-1.77 for for *ex vivo* and unmodified cell lines, and version picard-tools-1.117 for cell line knockdown/overexpression and controls) (https://broadinstitute.github.io/picard/).

Peak calling on mouse tissue/cell lines

MACS2 (version 2.0.10.20131216 beta) was used to call peaks on merged bam file from all *ex vivo* samples (L1, L2, L3, L4, T1, T2, T3, T4) to get broad peak calls (Zhang et al. 2008). No model was applied for the offset of 5' and 3' strands; and the default lambda was applied for local background subtraction. Duplicates were not removed in this step, having been previously removed by picard (see above). The command used was:

macs2 callpeak --nomodel --broad --keep-dup all

This initial set of 261,515 peaks was subsequently filtered for low false discovery rate (qvalue < 1e-5), resulting in 99,819 peaks total. Blacklisted regions were removed (https://sites.google.com/site/anshulkundaje/projects/blacklists) using bedtools (Quinlan & Hall 2010).

Despite having called peaks using only reads with high mapping quality (>q30), there were a few likely artifactual peaks, which we surmised to have resulted from the alignment of reads with high similarity to mitochondrial sequences that were aligned to the nuclear genome. We identified such artifacts based on their having a large fraction of poor mapping quality reads; therefore, any peaks that contained more than 70% poorly mapping reads (<q30) were also removed (~700 peaks).

Finally, we observed that there were regions that had copy number amplifications (CNA) evident even in the ATAC-seq data. These were called from the ATAC-seq background signal (see next section), and peaks overlapping these regions were removed (~8000 peaks). The final set was composed of 91,401 peaks, 88,900 of which were on the autosomes.

Peak calling on human cell lines

For human cell lines, peaks were called on merged bam files using the same parameters as for mouse. Peaks intersecting blacklisted regions were similarly removed, resulting in 192,254 peaks. When assessing differential accessibility, only the top 100K peaks with the highest number of normalized read counts in the human cell lines were used.

Identification of copy number alterations from ATAC-seq data

From examining the read distributions at chromosome-wide scale, it was clear that certain large regions (>10Mb) had large fold amplifications. We hypothesized that these regions had copy number amplifications, supported by the fact that these regions corresponded to well documented CNAs found in mouse SCLC models (Dooley et al. 2011; McFadden et al. 2014; Calbo et al. 2011).

These regions were identified by assessing the number of 'background' reads--i.e. reads that were outside of the called peaks before any filtering (the original set of ~200,000 peaks)—within sliding windows of 1Mb (step size 500kb). For each sample, at each step, the fold-change above the median of all windows was evaluated. If fold-change was greater than three for any of the tissue or cell line samples, those windows were termed 'amplified'. Any peaks that fell within that window were discarded (<10% of total peaks).

Normalizing read counts and evaluating significantly altered peaks

The number of reads that fell into each peak for each sample was evaluating using the multicov module of bedtools (Quinlan & Hall 2010). Only reads with high mapping score (>q30) were included in this analysis. Samples could have different read counts per peak for multiple reasons, including differential efficiency of the ATAC-seq protocol in assaying regions of open chromatin, differential quantification/clustering efficiency on the sequencer, or differential chromatin accessibility across samples. Only differential chromatin accessibility represents a biological feature of interest.

Assessing differential read counts requires an underlying model for how the read count corresponds to the feature of interest, i.e. accessibility. Variability in quantification and clustering efficiency on the sequencer generates different numbers of reads per sample that do not reflect underlying accessibility. To account for these technical differences, per-sample scale factors are introduced to normalize for sequencing depth.

Many methods of finding the appropriate scale factors are suggested in the literature. We used a trimmed-mean approach, whereby a set of low-variance peaks ("null" peaks) was identified, and the mean read counts in this set were used to determine the per-sample scale factors. The peaks forming this set were selected by annotating peaks with expression data from tumors and metastases from same mouse model (28 samples total from 9 mice). Peaks containing transcription start sites (from RefSeq) for genes that were both expressed (median FPKM > 5) and had low variance (variance in FPKM < 0.1) were assumed to accurately reflect the change in read count due to technical differences in sample preparation. These criteria produced a set of 4,001 null peaks. The sum of reads within these peaks, normalized by the geometric means, was used as scale factors for each sample.

Human cell lines lacking expression data were normalized by the total reads in all TSS peaks.

To evaluate whether the mean per peak was significantly different in any sample set, DEseq2 was used (Love et al. 2014).

For the ex vivo samples, peaks that were greater than 2 fold changed at FDR < 0.1 were considered newly open peaks for the tissue samples, and peaks that did not have log2 fold change significantly different from zero were termed constitutive peaks.

Because of lower overall sequencing depth of cell lines, these and subsequent comparisons were defined as changed if the absolute value of the log2 fold change was greater than 0.5 at FDR < 0.1.

For modified cell lines, the DESeq2 model was built using both the knockdown and overexpression libraries. Multiple batches of the knockdown cell lines were taken into account by adding this batch variable to the model. Though multiple time points were taken for the overexpression dataset, we found no significant accessibility differences when explicitly adding time as independent variable to our linear model in DESeq2. To identify accessibility differences due to Nfib overexpression, all samples expressing cDNA for Nfib were compared to all samples with induction of the empty vector.

For human overexpression cell lines, the model was slightly different, with batch, cell line, and Nfib expression induction all given as predictors in DESeq2's model. The reported values were the log2 fold-changes estimated to be due to overexpression.

Principal components analysis and clustering

A regularized-log transformation was applied to the scaled reads/peak with the package DEseq2 (Love et al. 2014). This is intended to flatten the relationship between average openness and variance in openness, in order to weight both highly open and lowly open peaks similarly in the principal components analysis. Principal component analysis was then done in R with the command *prcomp* and default settings. The *ex vivo* samples were clustered using the correlation of regularized-log-transformed counts.

Standardized pipeline for comparing ATAC-seq to Dnase accessibility calls

Peaks calls from DNase-seq on diverse mouse tissues were downloaded from Mouse

ENCODE (Vierstra et al. 2014). Peak calls from all tissues were merged with our SCLC ATAC peak calls, and overlapping regions were merged. The number of reads within peaks was found for all tissues. For each tissue comparison (lung versus liver, T cell versus B cell, etc.), only the 50,000 peaks with the highest total read count were considered. The sum of reads in transcription start site peaks, normalized by geometric means, was used as scale factors in DEseq2, and log2 fold changes were found for each comparison separately. The reported fraction of differentially open peaks was determined using threshold that peaks changed in openness at least 2 fold.

TSS annotations

Transcription start site annotations were made using RefSeq (mm9) downloaded from UCSC table browser (Karolchik et al. 2004).

Finding large windows enriched or depleted for change

Windows were defined across the genome by tiling a 2 Mb with 100 kb step size. The count per peak for gene distal peaks (>5 kb from nearest TSS) of the merged hypo- and hyper-accessible samples were summed within each window, and the fold change between these two samples within each window was calculated. To determine the amount of change expected, the count per peak was randomly shuffled, and the fold change in peak count was recalculated. This was repeated 10 times to get an average expected change per window given randomly distributed fold change in peak counts. From the observed versus expected fold change, we calculated the log2 observed/expected change. When this was value was greater than 0.1, this window center was determined to be ‘enriched for change’, when it was less than -0.1, the window was determined to be ‘depleted for change’. Window centers were merged if within 1 Mb, and merged windows were kept if they were at least 5 Mb.

Comparison to replication timing

RepliChip data was downloaded from Mouse Encode published datasets (Stamatoyannopoulos et al. 2012). DCC accession numbers are given below:

CH12	wgEncodeEM002962
ES-46C	wgEncodeEM002963
ES-46C	wgEncodeEM002964
ES-D3	wgEncodeEM002965
ES-D3	wgEncodeEM002966
ES-D3	wgEncodeEM002967
ES-D3	wgEncodeEM002976
ES-D3	wgEncodeEM002968
ES-EM5Sox17huCD25	wgEncodeEM002978
ES-EM5Sox17huCD25	wgEncodeEM002977
ES-TT2	wgEncodeEM002969
ES-TT2	wgEncodeEM002970
EpiSC-5	wgEncodeEM002971
EpiSC-7	wgEncodeEM002972
J185a	wgEncodeEM002979
L1210	wgEncodeEM002973
MEF	wgEncodeEM002974
MEL	wgEncodeEM002975

Merging hypo- and hyper-accessible samples

Because sample L1 had many more reads per peak than the other *ex vivo* samples, reads from this sample were subsampled to only take 25% of total reads. The bam files of the subsampled L1 was merged with L3, L4, and T4 to get the hyper-accessible set of reads, and the bam files of L2, T1, T2, and T3 were merged to get the hypo-accessible set of reads, using samtools merge (Li et al. 2009).

Finding overrepresented motifs within changing peaks

HOMER was used to find motif representation changes in the set of newly open peaks using a background set of peaks that did not significantly change, using the parameter "-size given" (Heinz et al. 2010).

Finding NFI motif sites

HOMER annotatePeaks function was used to find NFI full sites within peaks. Unless otherwise stated, the set of motifs used for footprinting, nucleosome occupancy, or accessibility consisted of all motifs with score at least 6.

Finding transcription factor footprints and occupancy

For each motif site, the distribution of fragments as a function of fragment size and midpoint position relative to the motif center (for possible fragments of size 25 to 125 bp centered within 100 bp of the motif center) was fit to a mixture model that finds the optimal relative weights between two generating distributions: 1) the distribution from an ideal "bound" model modulated by local sequence bias, and 2) the distribution from an "unbound" model in which fragments are distributed based only on the local sequence bias and the fragment size distribution observed within peaks. The "bound" model was defined as the aggregate fragment size versus midpoint distribution for NFI motif sites with high motif scores (>7.598) (Fig S3I), divided by the expectation based on sequence bias at those sites, then normalized to sum to 1 (Fig S3J). The maximum likelihood estimate for the proportion of the fragments around a motif site generated from the "bound" model is considered to be the occupancy of that motif site. Classification of sites as bound was based on an occupancy threshold of 0.5, with log-likelihood

ratio greater than 2. Only sites with at least 20 reads with read centers within 100 bp of the motif site were included in the downstream analysis.

We can estimate true and false positive rates given the calls on the merged Nfib^{high} and Nfib^{low} ATAC-seq libraries by comparing to the intersection of the Nfib^{high} cell line ChIP-seq peaks or the KP22 cell line ChIP-seq peaks. We started with the set of NFI fullsite motifs with motif score>0 within accessible regions. We used a set of non-overlapping motifs such that each ChIP-seq peak had at most one site: if multiple sites were present within 500bp, only the site with the highest motif score was kept. Using the cutoffs described above: (only sites with at least 20 reads, occupancy at least 0.5, and log-likelihood ratio greater than 2), we found a true positive rate of 77% of Nfib^{high} cell lines and 84% for Nfib^{low} cell lines. We found a false positive rate of 39% and 21% for Nfib^{high} and Nfib^{low} libraries respectively. The high false positive rate signified that the ATAC-seq data suggested binding events that were not supported by ChIP-seq. However, looking at very low-affinity motifs, where we do not expect binding, we are calling at most 7% of sites as bound (Figure 3H), suggesting that the false positive rate may be much lower, and that our high estimates could be due to suboptimal ChIP-seq sensitivity. Furthermore, the sites that we called as bound but lacked ChIP-seq signal still exhibited strong transcription factor binding footprints (data not shown). We thus hypothesize that these regions were not detected in ChIP-seq due to the specificity of the antibody, which might not capture the binding events of other isoforms of Nfib or of other members of the NFI family.

Nucleosome occupancy determination

The program NucleoATAC was used to determine nucleosome occupancy (Schep et al. 2015). In short, this program implements a mixture model based on the overall fragment length distribution of a given sample. The occupancy measures how likely a given fragment is to come

from nucleosome-free DNA or nucleosomal DNA. This metric is normalized between 0 and 1. Occupancy was called within accessible regions in mouse SCLC. Signal was aggregated in windows 1kb surrounding NFI full sites of motif score greater than 6. If occupancy was not called, i.e. 500bp away from NFI motif site was not within accessible region, it was not included in the analysis.

Nfib knock-down and cDNA expression

For knock down, shRNA were purchased from TRC PLKO library. For shControl, shGFP:5'GCAAGCTGACCCTGAAGTTCAT3'shLuciferase: pLKO (SHC007). For mouse Nfib, shNfib#1: TRCN0000012089, shNfib#2: TRCN0000012092. For cDNA expression, mouse *Nfib* isoform 3(420aa) was cloned for overexpression. *Nfib* isoform 3 is chosen becuase: 1. Majorly expressed isform determined by Western Blot 2. Major expressed isoform predicted by exon reads from mouse RNAseq data 3. The most abundant and the major isoform amplified well). human NFIB cDNA was from hORFome library ID 2723.

Mouse *Nfib* and human *NFIB* were amplied by PrimeSTAR (TaKaRa). Mouse *Nfib* and human *NFIB* cDNA were cloned into Dox inducible vector *pCW22.TRE. cDNA; Ubc-rtTAiresBlastR* vector using Cold Fusion kit (System Biosciences). Mouse *Nfib* was also cloned into *pCDH-CMV-MCS-EF1-puro* (System Biosciences) via BamHI and NotI site.

Primer Sequence (5'-3'):

m*Nfib* Fwd (For *pCW22.TRE. cDNA; Ubc-rtTAiresBlastR*):

GATCGCCTGGAGGTTAACCACCATGATGTATTCTCCCATCTGTC

m*Nfib* Rev (For *pCW22.TRE. cDNA; Ubc-rtTAiresBlastR*):

GATCGCCTGGAGGTTAACCACCATGATGTATTCTCCCATCTGTC

h*NFIB* Fwd (For *pCW22.TRE. cDNA; Ubc-rtTAiresBlastR*):

GAACCGTCAGATCGCCTGGAGGTTAACCACCATGATGTATTCTCCCATCTGTC

h*NFIB* Rev (For *pCW22.TRE. cDNA; Ubc-rtTAiresBlastR*):

GCCAGATCTTGGGTGGGTTAATTAACTAGCCCAGGTACCAGGACTGGCT

m*Nfib* Fwd (For *pCDH-CMV-MCS-EF1-puro*):

CAGTGGATCCCCACCATGATGTATTCTCCCATCTGTCTC

m*Nfib* Rev (For *pCDH-CMV-MCS-EF1-puro*):

ACTGGCGGCCGCCTAGCCCAGGTACCAGGACTGGC

Virus production, infection and selection

Lentivirus were generated using Delta8.2 and VSV-G packaging plamids. MSCV retroviral vectors were generated using pCL-Eco and pseudotyped with VSV-G. For virus production, 5×10^6 HEK293T cells were seeded into 10 cm dishes and transfected with the vector of interest using PEI (Polysciences 23966-2) and appropriate packaging plasmids. Medium was changed 24 h later. Supernatants were collected at 36 and 48 h, passed through a 40μm filter and applied at full concentration to 50% confluent 10cm dishes of target cells. Two days after infection cells were selected with Puromycin (2μg/mL) or Blastcidine (8μg/mL) for at least1 week.

qRT-PCR and Western blotting

RNA extraction was done by Qiagen RNeasy mini kit. cDNA was generated by Reverse transcription kit (Invitrogen, 4368814) from 2μg of RNA input. qPCR was done using SYBR green (Sigma, S9194). Mouse *Nfib* and human *Nfib* gene expression levels were assessed using standard SYBR green qPCR protocols and normalized to mouse *Gapdh* or human *GAPD*. Primer sequences are as follows, from 5' to 3':

m*Nfib* Fwd: GTGTTCAGCCACACCACATC

m*Nfib* Rev: GAGGATTCTTGGCAGGATCA

h*NFIB* Fwd: TTTGTGTCCAGCCACATCAT

h*NFIB* Rev: GTGGCTTGGACTTCCTGATT

h*GAPD* Fwd: GAAGGTGAAGGTCGGAGTCA

h*GAPD* Rev: TTGAGGTCAATGAAGGGGTC

For western blotting, denatured protein samples were run on a 4-12% Bis-Tris gel (NuPage) and transferred onto PVDF membrane. Membranes were immunoblotted using primary antibodies against Hsp90 (1:1000, BD Transduction Laboratories, 610419) and Nfib (1:2000, Abcam, ab186738). Primary antibody incubations were followed by secondary HRP-conjugated anti-mouse (Santa Cruz Biotechnology, sc-2005) and anti-rabbit (Santa Cruz Biotechnology, sc-2004) antibodies and membranes were developed with ECL 2 Western Blotting Substrate (P180196, ThermoScientific Pierce).

Dox induction in vitro

Doxycycline was added to media at a final concentration of 1ug/ml. For NCI-H524 cell line, we used 2ug/ml.

Cell culture assays

All murine and human SCLC cell lines used in this study grow as floating spheres and were cultured in DMEM or RPMI with 10% FBS (Pen/Strep/Glutamine). Human cell lines were originally purchased from ATCC and cell identities were validated by Genetica DNA Laboratories using STR analysis. NJH29 was derived from SCLC patient derived xenograft, which has been described before (Jahchan et al. 2013). All cell lines were confirmed to be mycoplasma-negative (MycoAlert Detection Kit, Lonza). To assess proliferation, 1.5 x 10^6 cells were plated in triplicate wells of a 6-well plate two days before experiments. Cells were labeled with 10μM BrdU for 3 hours followed by anti-BrdU staining using the BD APC flow kit following the manufacturer's instructions.

In vitro cell growth was also assessed using AlamarBlue Cell Viability Reagent (Invitrogen), with 10,000 cells seeded per well of a 96 well plate. Cell growth was measured every two days after seeding using fluorescence plate reader with 530nm excitation and 590nm emission.

To assess cell death, 1.5 x10^6 cells were seeded into 6-well plates in triplicate wells two days before experiments. Cells together with supernatant were collected, spinned done, trypsinized and filtered by 40um filter to get single cell suspension. Cells were stained with AnnexinV (Biolend, 5ul per test) and DAPI (Life Technology D1306, 1:1000) and analyzed by FACS.

For anchorage-independent growth assays, cells from single cell suspension were plated in triplicate wells of a 6-well tissue culture plate in 0.4% agar in culture media on top of a layer of 0.8% agar with culture media. 50,000 cells for murine cell lines or 20,000 cells for human cell lines were plated per well. Cells were allowed to grow at 37 °C for 2–3 weeks. Colonies were

stained with 0.2% crystal violet at room temperature for 30 min and subsequently de-stained with water for several days. Once the colonies were visible by eye, plates were counted using a microscope.

For Matrigel invasion assays: ibidi inserts (80209) were attached to the plate that was pre-coated with poly-lysine (Sigma P6407). 80 $x10^4$ cells in 100µl were seeded to each chamber of the ibidi insert. After 6 hours, ibidi insert is removed and 1ml of 1:1 matrigel (Corning356231)-media mix is added to coat each well. To quantify the level of migration, the number of cells that has migrated into the gap were counted in each field at 10x under the scope.

Transplantation assays

Cells were transplanted into age-matched gender-matched NOD/SCID/γc (NSG) mice (The Jackson Laboratories, Stock number 005557). For 16T cells, $2x10^4$ cells (in 200µl PBS) were transplanted into the lateral tail vein of recipient mice. KP22-TRE-empty and KP22-TRE-Nfib cells were treated with doxycycline for 2 weeks in culture prior to transplantation. $2x10^4$ cells (in 200µl PBS) were transplanted into the lateral tail vein of recipient mice. All mice were analyzed 3 to 4 weeks after transplantation. For subcutaneous injection, cells were resuspended in 100µl PBS and mixed with 100µl growth factor reduced matrigel (Corning, 356231). 4 injection sites per mouse. All experiments were performed in accordance with Stanford University Animal Care and Use Committee guidelines.

RNA-seq library preparation and analysis

RNA-sequencing libraries were prepared using the Illumina TruSeq v2 kit, according to manufacturer's instructions. Total RNA was isolated from each cell line (KP1-shLuc, KP1-

shNfib, 16T-shLuc, 16T-shNfib, KP22-TRE-Empty and KP22-TRE-Nfib) in duplicate using the Qiagen RNAeasy mini kit, resulting in 12 samples. For each sample, 0.5µg of total RNA was used for library construction. The integrity and quality of RNA was assessed prior to library construction using an Agilent Bioanalyzer 2100. High-throughput sequencing was performed on a single lane of an Illumina HiSeq 2000 for 100bp paired end reads.

For analysis of the RNA-seq reads, we performed differential gene and transcript expression analysis of RNA-seq experiments using TopHat and DESeq2 (Trapnell et al., 2012). RNA-seq reads were separately aligned to the mouse genome (mm10) using TopHat. Exons were found from RefSeq genes, and exons with the same start and end sites were merged. The number of reads overlapping each merged exon was found using bedtools multicov (-q30). Finally, this count was summed for exons all mapping to a particular gene. In this way we obtained a count per gene per sample. These counts were normalized for length and GC bias using cqnnorm (Hansen et al. 2012). The sum of all reads within exons was used for the size factors. The length of all exons per gene was summed to get the length covariate, and the GC covariate was the average GC content within exons annotated per gene. The size factors were used to calculate DESeq2 normfactors as recommended in the DESeq2 documentation:

cqnNormFactors <- exp(mycqn$glm.offset)

normFactors <- cqnNormFactors / exp(rowMeans(log(cqnNormFactors)))

where “mycqn” is the cqn results object.

Using these normalization factors and counts, DEseq2 was used to assess differential expression (Love et al. 2014). The knockdown and overexpression samples were analyzed together such that dispersion estimates used all of these libraries. Genes were considered

differentially expressed if the absolute value of log2 fold change was greater than 0.5 between modification (i.e. overexpression or knock-down) and control at FDR < 0.1.

GO terms were found by comparing gene sets to a set of background genes that were tested in DEseq2 but did not change significantly from zero using GORILLA web application using two unranked list of genes (Eden et al. 2009). To avoid redundancy, any "directional" GO term, i.e. that contained "positive" or "negative" regulation was not included in the analysis, which reduced the number of GO terms from 96 to 72.

GO term categories were consolidated by finding the genes that were consistently giving the enriched GO terms. GO terms were clustered by the jaccard index of the vector of genes that gave rise to it, and 16 flat clusters were called from this hierarchical clustering. 14 categories were included in the analysis; the other two were composed of mixed groups that had few genes (less than four) and were thematically unrelated. The name of the GO category was annotated by manual examination of the GO terms within each cluster.

Comparison of genes to ATAC-seq peaks

To make the comparison of number of distal peaks around genes, windows were generated encompassing 100kb upstream and downstream of RefSeq TSS locations for a total window size of 200kb. These windows were intersected with peaks such that every peak's overlap with the TSS windows was annotated. Any peak that was annotated as a promoter by HOMER (-1000bp, +100bp for TSS) was discarded from the analysis. This mapping was then used to find the number of changed peaks for every gene, and the average expression changes for genes that had peaks nearby or not.

Chromatin immunoprecipitation and analysis

Chromatin immunoprecipitation (ChIP) for Nfib was performed as described by the Farnham laboratory (http://farnham.genomecenter.ucdavis.edu/pdf/FarnhamLabChIP%20Protocol.pdf) with several modifications. Cell pellets were weighed after fixation and 100 μl of swelling solution and nuclei lysis buffer was used per 10 mg of pellet weight. Chromatin was fragmented to an average size of 150-300 bp using a Bioruptor 300 (Diagenode) for 170 cycles of 30 seconds at high power setting. For each ChIP, 300 µl of fragmented and pre-cleared chromatin was diluted and incubated overnight at 4°C with 4 µg of antibody. The nucleoprotein complexes were bound by Pansorbin cells (Calbiochem, #507862) and the eluted DNA was digested with RNaseA and Proteinase K before being purified using the QIAquick PCR purification kit (Qiagen). Antibodies used for ChIP: Nfib antibody (Sigma) and normal rabbit IgG (Cell Signaling, #2729) as a control. ChIP product was quantified by qubit. Library was generated following Clontech low input with 1ng of ChIP product. 12 samples were pooled for Illumina Hiseq.

ChIP reads were aligned to mm9 using bowtie2. Bam files were processed the same as ATAC-seq libraries (only reads aligning to autosomes, duplicate reads removed). ChIP peaks were found using the pulldown replicate bam files using MACS2, and using the input bam files as background, i.e.: "callpeak -t 16T_PD1.bam 16T_PD2.bam -c 16T_input1.bam 16T_input2.bam -B --SPMR -g mm"

The peak calls for each of the three cell lines were merged, and counts for the cell lines pulldown and technical replicates were found using bedtools multicov (-q 30) and compared in DEseq2 to find sites with differential ChIP signal (absolute value of log2 fold change > 0.5 at FDR < 0.1). The number of ChIP reads was also found in the accessible regions defined by the

ATAC-seq libraries; this was used to compare differential ChIP signal to differential accessibility.

Calculation of residence time of NFI family members

Reported measurements of in vitro affinity of NFI to the consensus full site range from $5x10^{-11}$ to $5x10^{-12}$ (Meisterernst et al. 1988). Assuming a diffusion-limited on rate of 10^8 $M^{-1}s^{-1}$ (Halford & Marko 2004), and $t_{1/2} = \frac{\log(2)}{K_d k_{on}}$, we calculate a lower bound on the half life of binding to DNA between 2 and 20 minutes.

References

Buenrostro, J.D. et al., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, 10(12), pp.1213–8.

Calbo, J. et al., 2011. A Functional Role for Tumor Cell Heterogeneity in a Mouse Model of Small Cell Lung Cancer. *Cancer Cell*, 19(2), pp.244–256.

Dooley, A.L. et al., 2011. Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes & development*, 25(14), pp.1470–5.

Eden, E. et al., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics*, 10(1), p.48.

Halford, S.E. & Marko, J.F., 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic acids research*, 32(10), pp.3040–52.

Hansen, K.D., Irizarry, R.A. & WU, Z., 2012. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*, 13(2), pp.204–216.

Heinz, S. et al., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4), pp.576–89.

Jahchan, N.S. et al., 2013. A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer discovery*, 3(12), pp.1364–77.

Jonkers, J. et al., 2001. Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nature genetics*, 29(4), pp.418–25.

Karolchik, D. et al., 2004. The UCSC Table Browser data retrieval tool. *Nucleic acids research*, 32(Database issue), pp.D493–6.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), pp.357–9.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), pp.2078–9.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), p.550.

McFadden, D.G. et al., 2014. Genetic and Clonal Dissection of Murine Small Cell Lung Carcinoma Progression by Genome Sequencing. *Cell*, 156(6), pp.1298–1311.

Meisterernst, M. et al., 1988. A quantitative analysis of nuclear factor I/DNA interactions. *Nucleic acids research*, 16(10), pp.4419–35.

Quinlan, A.R. & Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), pp.841–2.

Sage, J. et al., 2003. Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. *Nature*, 424(6945), pp.223–8.

Schaffer, B.E. et al., 2010. Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. *Cancer research*, 70(10), pp.3877–83.

Schep, A.N. et al., 2015. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome research*, 25(11), pp.1757–70.

Stamatoyannopoulos, J.A. et al., 2012. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome biology*, 13(8), p.418.

Vierstra, J. et al., 2014. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science*, 346(6212), pp.1007–12.

Zhang, Y. et al., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), p.R137.