

# SCIENTIFIC REPORTS



OPEN

## p52 expression enhances lung cancer progression

Jamie A. Saxon<sup>1</sup>, Hui Yu<sup>2,3</sup>, Vasily V. Polosukhin<sup>4</sup>, Georgios T. Stathopoulos<sup>5,6</sup>, Linda A. Gleaves<sup>4</sup>, Allyson G. McLoed<sup>1</sup>, Pierre P. Massion<sup>1,4,7</sup>, Fiona E. Yull<sup>1</sup>, Zhongming Zhao<sup>1,2,3</sup> & Timothy S. Blackwell<sup>1,4,7,8</sup>

Received: 17 February 2017

Accepted: 28 March 2018

Published online: 17 April 2018

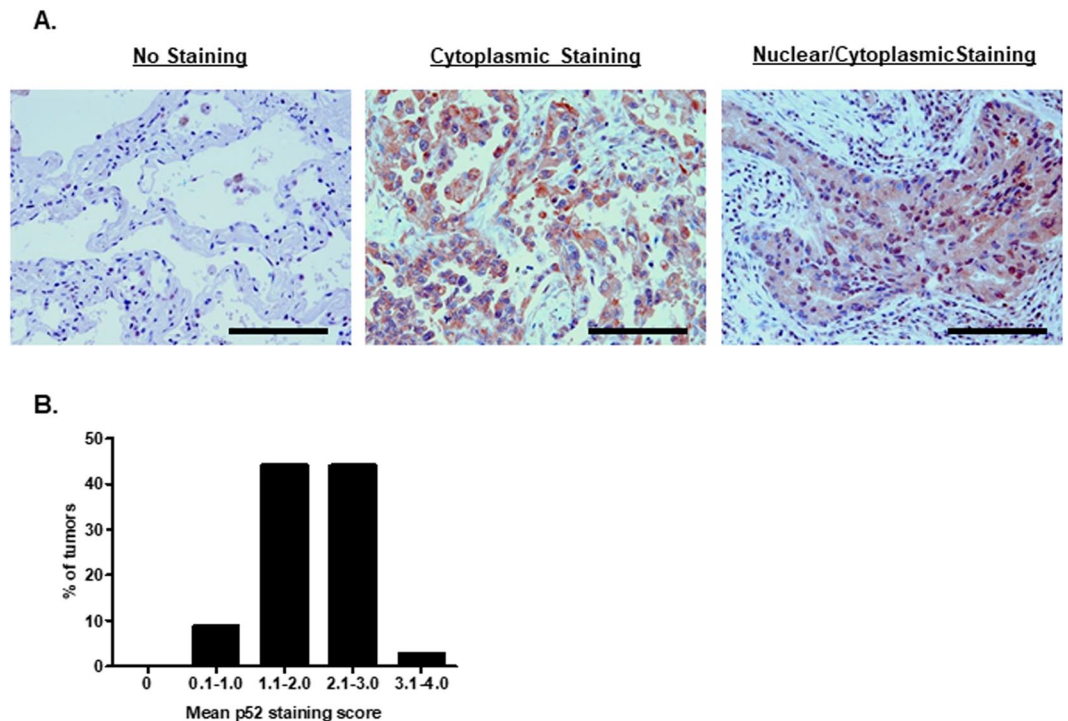
While many studies have demonstrated that canonical NF- $\kappa$ B signaling is a central pathway in lung tumorigenesis, the role of non-canonical NF- $\kappa$ B signaling in lung cancer remains undefined. We observed frequent nuclear accumulation of the non-canonical NF- $\kappa$ B component p100/p52 in human lung adenocarcinoma. To investigate the impact of non-canonical NF- $\kappa$ B signaling on lung carcinogenesis, we employed transgenic mice with doxycycline-inducible expression of p52 in airway epithelial cells. p52 over-expression led to increased tumor number and progression after injection of the carcinogen urethane. Gene expression analysis of lungs from transgenic mice combined with *in vitro* studies suggested that p52 promotes proliferation of lung epithelial cells through regulation of cell cycle-associated genes. Using gene expression and patient information from The Cancer Genome Atlas (TCGA) database, we found that expression of p52-associated genes was increased in lung adenocarcinomas and correlated with reduced survival, even in early stage disease. Analysis of p52-associated gene expression in additional human lung adenocarcinoma datasets corroborated these findings. Together, these studies implicate the non-canonical NF- $\kappa$ B component p52 in lung carcinogenesis and suggest modulation of p52 activity and/or downstream mediators as new therapeutic targets.

Lung cancer is the leading cause of cancer-related death in the U.S, with an estimated 224,390 new lung cancer diagnoses in 2016 and a 5-year survival rate of less than 20 percent<sup>1</sup>. In addition to emphasizing the need for better understanding of tumor biology and development of new therapeutic approaches, available epidemiological data suggest that identification of biomarkers for improved outcome prediction beyond the current staging system would be valuable for guiding appropriate use of available treatment strategies.

Numerous studies have demonstrated the critical role of epithelial NF- $\kappa$ B signaling in lung cancer<sup>2–4</sup>. The NF- $\kappa$ B transcription factor family contains 5 members (p65/RelA, p52, p50, RelB, and c-Rel) and can be activated through either canonical or non-canonical signaling pathways. Canonical pathway activation occurs when inhibitory I $\kappa$ Bs are phosphorylated, releasing p65/p50 heterodimers to translocate into the nucleus. Non-canonical NF- $\kappa$ B signaling is defined by nuclear accumulation of p52, which hinges on proteolytic processing of p100 removing an inhibitory C-terminal domain and resulting in p52 activation. p52 is then able to enter the nucleus, typically as a heterodimer bound to RelB. Although components of both the canonical and non-canonical NF- $\kappa$ B signaling pathways are activated in a number of malignancies, including lung cancer<sup>5–7</sup>, studies of NF- $\kappa$ B in cancer have mainly focused on canonical pathway signaling, leaving the non-canonical pathway largely uninvestigated.

The majority of knowledge regarding *in vivo* functions of p100/p52 relate to its role in lymphoid development and hematopoietic disease. Mice with global knockout of the *Nfkb2* gene, which encodes p100, demonstrate defective dendritic cell function, B-cell maturation, T-cell responses, and secondary lymphoid organ

<sup>1</sup>Department of Cancer Biology, Vanderbilt University, Nashville, TN, 37232, USA. <sup>2</sup>Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, 37203, USA. <sup>3</sup>Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN, 37232, USA. <sup>4</sup>Department of Medicine, Division of Allergy, Pulmonary and Critical Care Medicine, Vanderbilt University, Nashville, TN, 37232, USA. <sup>5</sup>Comprehensive Pneumology Center (CPC) and Institute for Lung Biology and Disease (iLBD), University Hospital, Ludwig-Maximilian University (LMU) and Helmholtz Center Munich, Member of the German Center for Lung Research (DZL), Max-Lebsche-Platz 31, 81377, Munich, Bavaria, Germany. <sup>6</sup>Laboratory for Molecular Respiratory Carcinogenesis, Department of Physiology, Faculty of Medicine, University of Patras, 1 Asklepiou Str., 26504, Rio, Achaia, Greece. <sup>7</sup>Department of Veterans Affairs Medical Center, Nashville, TN, 37232, USA. <sup>8</sup>Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN, 37232, USA. Correspondence and requests for materials should be addressed to T.S.B. (email: [timothy.blackwell@vanderbilt.edu](mailto:timothy.blackwell@vanderbilt.edu))



**Figure 1.** p100/p52 expression is present in human lung adenocarcinomas. (A) Representative images of p100/p52 immunostaining of lung adenocarcinoma and normal lung tissue (40x magnification; scale bar = 100  $\mu$ m). (B) Distribution of p100/p52 immunostaining on a tumor tissue microarray (TMA) from 106 human lung adenocarcinoma samples obtained from 34 patients. Tumor spots were scored as follows: 0 = no staining; 1 = diffuse cytoplasmic staining, no nuclear staining; 2 = dark granular cytoplasmic staining, no nuclear; 3 = nuclear staining in <25% of tumor cells; 4 = nuclear staining in >25% of tumor cells. Each tumor had at least two spots on the TMA, and the mean score was calculated for each tumor based on the individual score for each spot.

development<sup>8,9</sup>, presumably due to the role of p52 in regulating expression of chemokine genes important for normal lymphoid organogenesis<sup>10</sup>. In several hematopoietic malignancies, including multiple myeloma, chronic lymphocytic leukemia, and B- and T-cell cutaneous lymphomas, chromosomal translocations in the *Nfkb2* gene have been identified that truncate the C-terminus, thus removing the inhibitory domain<sup>11-13</sup>. These truncated proteins localize in the nucleus, leading to constitutively active p52<sup>11</sup>. Knock-in mice that express truncated *Nfkb2* develop enlarged lymph nodes and hyperactive T cell responses, providing some insight into the function of activated p52 in hematopoietic diseases<sup>14</sup>. However, these mice die prematurely due to significant gastric hyperplasia, hinting that dysregulated p52 activation could also play a role in epithelial-based disorders.

Increased activation of p52 has been observed in lung<sup>6</sup>, breast<sup>15</sup>, prostate<sup>16,17</sup>, and pancreatic cancers<sup>18</sup>; however, studies on the effects of p52 activation in cancers of epithelial origin have been limited due to lack of appropriate *in vivo* models. Here, we use a transgenic mouse model with inducible expression of p52 in the airway epithelium (designated CCSP-p52) to investigate a potential role for p52 in lung tumorigenesis. Together, these studies define a role for p52 in lung tumorigenesis and identify a p52-associated gene network that is highly correlated with patient prognosis.

## Results

**p100/p52 expression is common in human lung adenocarcinoma.** To examine the distribution of p100/p52 expression in human lung adenocarcinoma, we performed p100/p52 immunostaining on a lung adenocarcinoma tissue microarray (TMA) (Fig. 1A, Table 1). We observed p100/p52 expression in all lung adenocarcinomas tested (34/34) (Fig. 1B). Multiple samples from each tumor were present on the array, and each sample was scored on a 0 to 4 point scale based on the intensity and the localization of p100/p52 staining. We calculated an average p100/p52 expression score for each tumor and found that 47% (16/34) of tumors had mean scores > 2, indicating substantial nuclear staining for p52. In contrast, minimal p100/p52 expression was observed in normal lung parenchyma (Fig. 1A). Together, these observations suggest that p100/p52 expression and p52 activation are common in lung adenocarcinomas.

**CCSP-p52 mice develop an increased tumor burden and more advanced lesions.** In order to investigate the effect of p52 expression during lung tumorigenesis, we employed CCSP-p52 mice, which inducibly express FLAG-tagged murine p52 in Clara cell secretory protein (CCSP)-positive epithelial cells<sup>19</sup>. These mice exhibit normal lung histology after p52 over-expression<sup>19</sup>. To induce tumors in CCSP-p52 mice, we used the lung carcinogen urethane (ethyl carbamate), which causes tumors primarily through *Kras* mutations<sup>20</sup> and, in FVB

Factor	n	Mean $\pm$ SE
Age (years)		
$\leq 60$	12	54.41 $\pm$ 1.31
$> 60$	21	67.95 $\pm$ 0.99
Unknown	1	
Ever Smoker		
Yes	32	
No	1	
Unknown	1	
Pack Years		
$\leq 50$	16	34.99 $\pm$ 2.33
$> 50$	15	109.95 $\pm$ 11.90
Unknown	1	
Stage at Diagnosis		
Stage I (A and B)	22	
Stage II (A and B)	5	
Stage III (A and B)	5	
Unknown	2	

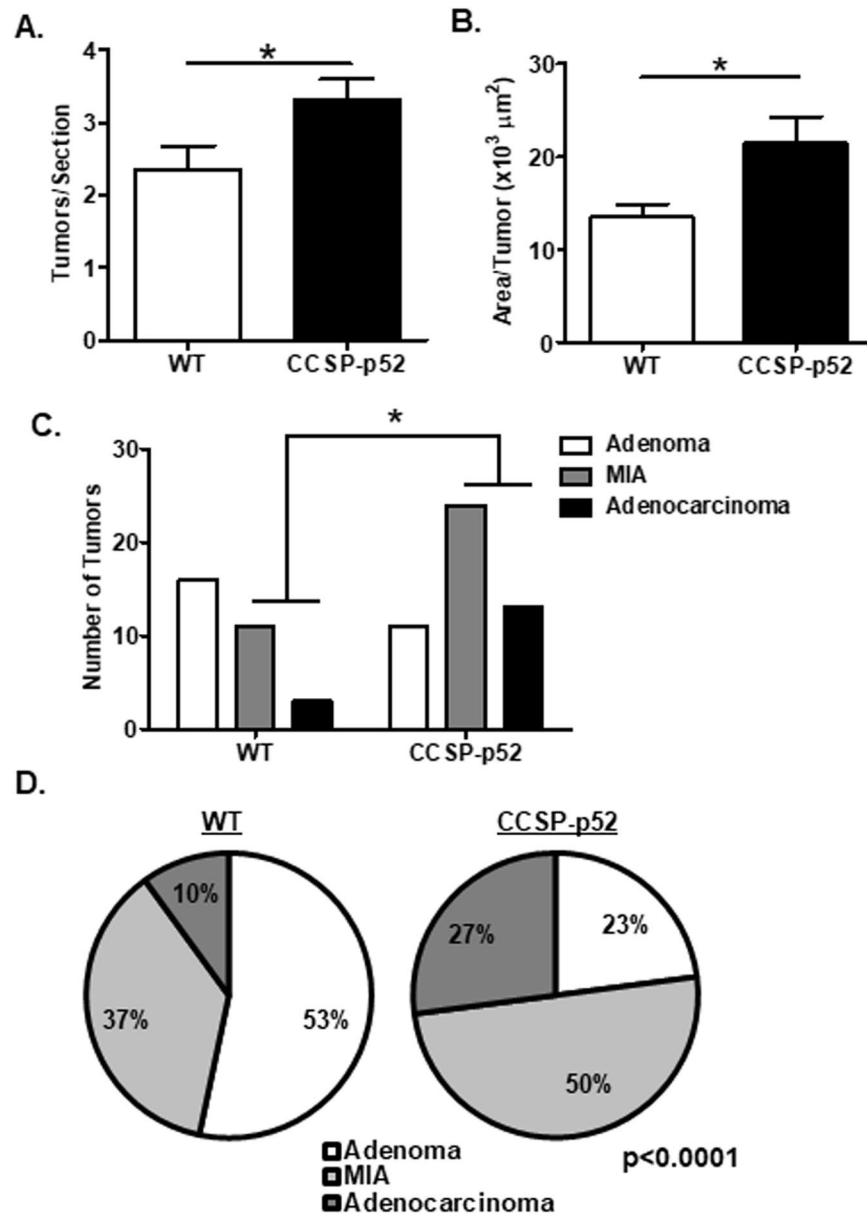
**Table 1.** Clinicopathological characteristics of patient samples.

background mice, induces a mild inflammatory response<sup>2</sup>. We first examined the effects of p52 over-expression on urethane-induced inflammation and atypical adenomatous hyperplasia (AAH) formation. For these experiments, CCSP-p52 and wild-type (WT) mice were treated with dox for one week followed by a single intraperitoneal (IP) injection of urethane (1 g/kg) and remained on dox until euthanized. Total inflammatory cells in bronchoalveolar lavage (BAL) were counted at 10 days, 21 days (3 weeks), and 42 days (6 weeks) after urethane injection. No differences were observed between CCSP-p52 and WT mice at any time point (Supplementary Fig. S1A). In addition, no differences in AAH lesion numbers were detected at 6 weeks post-urethane (Supplementary Fig. S1B).

To investigate later stages of tumor formation, CCSP-p52 and WT mice were treated with dox for one week followed by 4 weekly IP urethane injections. Mice remained on dox treatment until sacrifice 6 months after the first urethane injection. Lung tumor dimensions were measured, and lung tumors enumerated on lung sections, revealing a significant increase in tumor number and tumor size in CCSP-p52 lungs compared to WT (Fig. 2A and B). We further analyzed these tumors by histology, classifying tumors as adenomas, minimally invasive adenocarcinomas (MIAs), or adenocarcinomas. Compared to WT mice, CCSP-p52 mice had an increased proportion of advanced lesions (MIAs and adenocarcinomas,  $X^2 p < 0.0001$ ) (Fig. 2C and D). In comparison, long-term p52 expression alone resulted in sustained p52 expression but no evidence of inflammation or tumor formation (Supplementary Fig. S2). Taken together, these studies demonstrate that p52 expression augments lung tumor formation and contributes to enhanced tumor progression.

**p52 expression enhances proliferation through regulation of cell cycle genes.** To understand the mechanism by which p52 expression augments tumorigenesis, we performed microarray gene expression analysis using mRNA isolated from lungs of CCSP-p52 and WT mice on dox for 1 week. Genes were ranked based on the log fold change (logFC) of expression between WT and CCSP-p52 mice using a cut-off of 0.41 (approximately 30% increase in gene expression in CCSP-p52 mouse lungs) (Fig. 3A). To evaluate the potential relevance of p52-regulated genes in human lung tumors, we focused only on genes that had known human homologs and were present in The Cancer Genome Atlas (TCGA) expression data from lung adenocarcinoma patients. Using these criteria, a total of 71 genes associated with increased p52 expression were identified (Fig. 3A and B; Supplementary Table S1). Quantitative PCR was performed to validate increased expression of selected genes from the 71 gene list in CCSP-p52 lungs (Fig. 3C–E). To elucidate gene functions, we identified biological processes that were enriched in the p52-associated gene set using a Gene Ontology (GO) analysis. Applying a false discovery rate (FDR) of 0.001, we found that p52-associated genes were over-represented in processes related to cell cycle progression (Table 2), suggesting that p52 regulates expression of genes involved in proliferation. To validate this finding, we used a stable p52-expressing rat lung epithelial cell line (RLE-p52) generated previously<sup>19</sup>. We found that BrdU incorporation was enhanced in RLE-p52 cells compared to empty vector (RLE-EV) (Fig. 3F), indicating that p52 stimulates proliferation of lung epithelial cells.

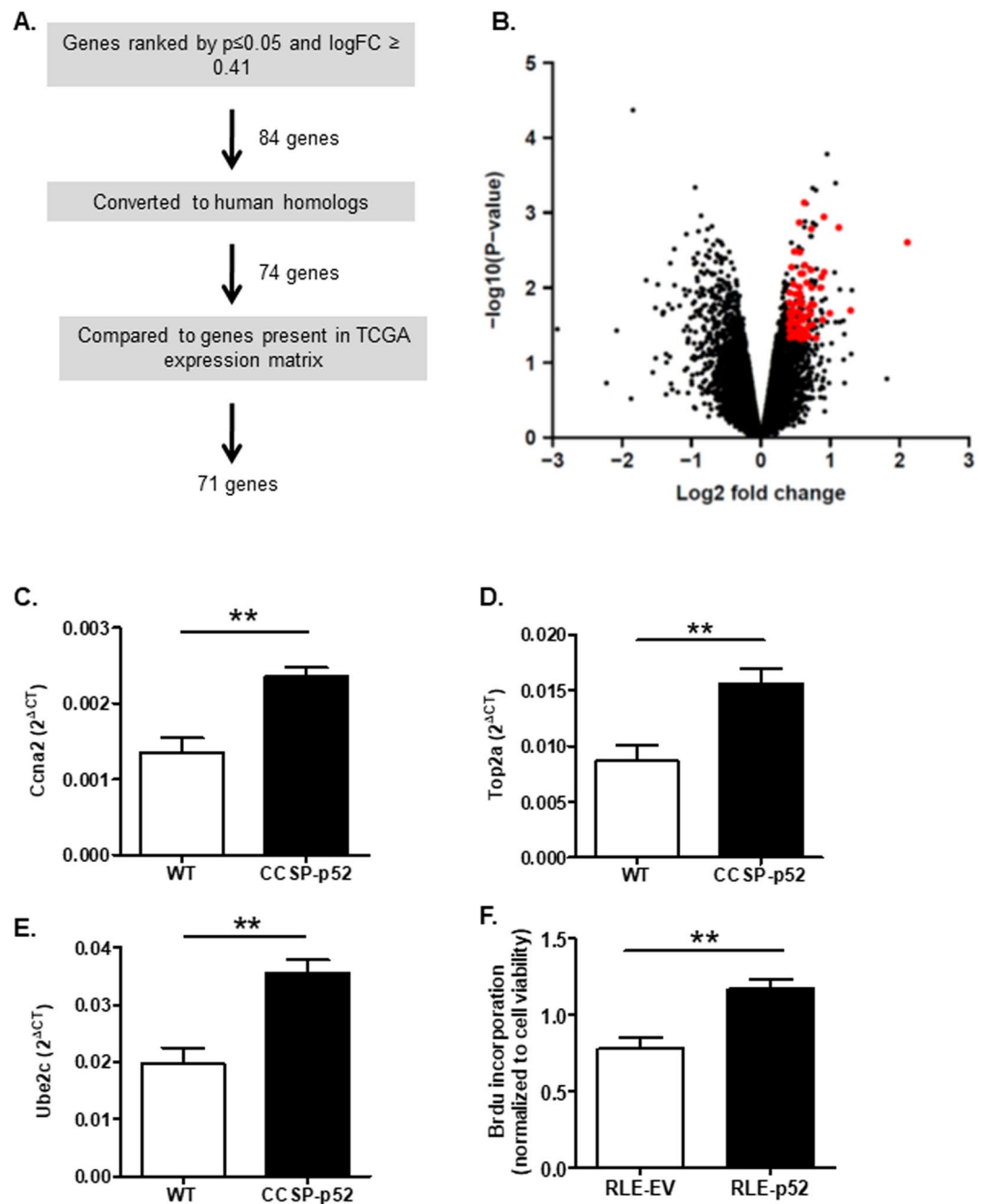
**Expression of p52-associated genes predicts prognosis of lung cancer patients.** To investigate the relevance of p52 expression in human lung cancer, we examined expression of the 71 identified p52-associated genes in matched lung cancer and normal samples. A differential expression test using the paired t-test was performed for each gene in paired tumor and normal samples from 54 lung adenocarcinoma patients available in the TCGA dataset. Up-regulated genes were defined as having both a logFC  $> 1$  and false discovery rate (FDR)  $< 0.01$ . From the set of 20,501 genes, 1,274 up-regulated genes (6.2%) were identified in tumors compared to matched normal samples (Fig. 4A). Of the 71 p52-associated genes, 35 were up-regulated in tumors (49.3%), indicating a significant over-representation of p52-associated genes in tumor samples compared to normal (hypergeometric



**Figure 2.** p52 over-expression results in increased tumor number, size, and progression. (A) Average tumor number per lung section and (B) average tumor area from WT and CCSP-p52 mice 6 months after urethane injection (3 sections/mouse; n = 13–14 mice/group; \*p < 0.05 compared to WT mice.). (C) Number and (D) proportion of lung tumors classified as adenomas, minimally invasive adenocarcinomas (MIA), or adenocarcinomas (n = 13–14 mice/group; \*X<sup>2</sup> p < 0.0001 for comparing total number of advanced lesions [MIAs and adenocarcinomas] between CCSP-p52 and WT mice).

test  $p = 6.0 \times 10^{-25}$ ) (Fig. 4A,B and Supplementary Table S1) and suggesting that p52 could be important in human lung tumors.

Since we observed more advanced tumors in urethane-treated CCSP-p52 mice, we wondered whether increased p52 expression would correlate with a worse prognosis in lung cancer patients. Therefore, we investigated the relationship between increased expression of p52-associated genes and patient outcomes using tumor expression data and clinical information from the TCGA dataset. 437 primary lung adenocarcinoma samples were divided into two equal groups based on the collective expression level of the 71 p52-associated genes by summing expression from all 71 genes (Supplementary Table S2). Kaplan-Meier survival analysis revealed that patients with high expression of p52-associated genes had a significantly worse prognosis (log-rank test  $p = 0.004$ ) (Fig. 4C). To validate this analysis, we divided patients by a second method into two groups based on summed “votes” from the 71 genes. In this analysis, one gene counted as a vote for one sample if its expression in the sample was equal to or higher than the median expression level for this gene across samples, thus preventing a tumor sample with only a few highly-expressed genes from being placed in the “high expression group”. Using this method, Kaplan-Meier survival analysis again revealed that patients with high expression of p52-associated



**Figure 3.** Identification of p52-regulated genes by microarray analysis. (A) Schematic of process used to identify p52-regulated genes from microarray data generated from pooled mRNA samples from lungs of WT and CCSP-p52 mice on dox for 1 week ( $n = 4/\text{group}$ ). (B) Volcano plot of log fold changes and differential expression p-values comparing CCSP-p52 mice to WT. Black dots represent all genes in the expression data matrix. Red dots represent 71 genes identified for further analysis. (C–E) Expression microarray data validation by quantitative PCR of identified p52-regulated genes *Ccna2* (C), *Top2a* (D), and *Ube2c* (E) measured in whole lung RNA from WT and CCSP-p52 mice on dox for 1 week ( $n = 6$  mice/group; \* $p < 0.05$  compared to WT). (F) Quantification of BrdU incorporation of RLE-p52 and RLE-WT cells, normalized to total viable cells (\*\* $p < 0.01$  compared to RLE-EV).

genes had significantly worse outcomes (log-rank test  $p = 0.021$ ) (Supplementary Fig. S3A). Furthermore, survival analysis of only Stage I patients revealed that patients with high expression of p52-associated genes still had a significantly worse prognosis (log-rank test  $p = 0.036$ ) (Fig. 4D). Together, these data suggest that expression of p52-associated genes in tumors of lung cancer patients leads to shorter survival and can be used to predict patient outcome in early stage lung cancer patients.

Since the GO analysis suggested that p52 regulates a number of genes that have highly related functions, we extracted a protein subnetwork connecting the p52-associated genes from a human protein-protein interaction



Gene Ontology term (Biological Process)	Reference set size	Expected Count	Observed count	Adjusted p-value <sup>1</sup>
Cell cycle checkpoint	230	0.98	9	2.59E-05
Mitotic cell cycle phase transition	443	1.88	11	8.83E-05
DNA unwinding involved in DNA replication	8	0.03	3	1.39E-04
Regulation of cell division	221	0.94	8	1.39E-04
Regulation of attachment of spindle microtubules to kinetochore	9	0.04	3	1.81E-04
DNA replication initiation	29	0.12	4	1.88E-04
Regulation of cell cycle process	435	1.85	10	3.68E-04
Chromosome organization	765	3.25	13	4.04E-04
Spindle assembly involved in mitosis	15	0.06	3	7.44E-04

**Table 2.** Gene ontology (GO) analysis of p52 regulated genes (FDR = 0.001). <sup>1</sup>p-values have been adjusted using the Benjamini-Hochberg method.

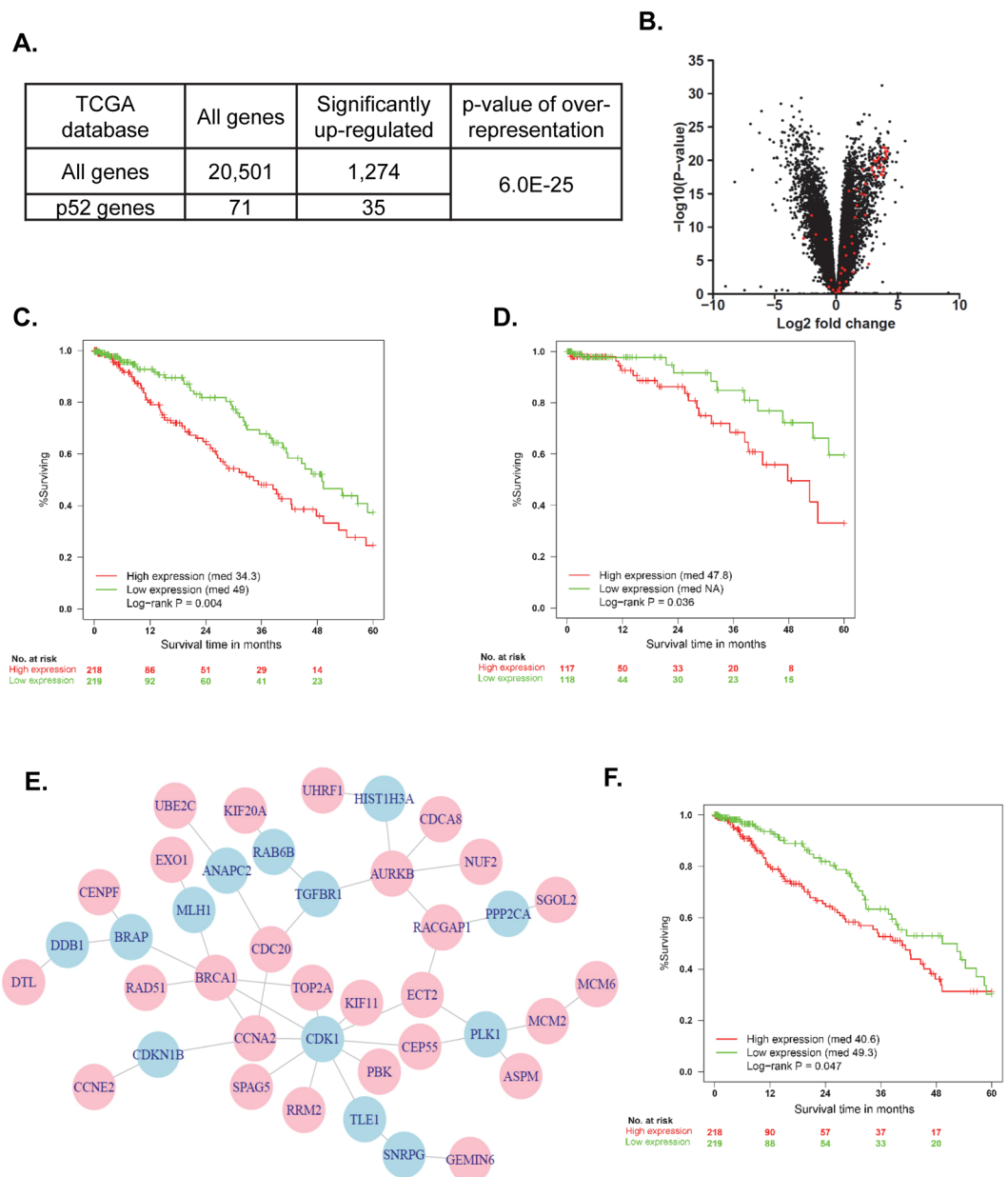
network. The 35 p52-associated genes identified in tumor versus normal samples were used to generate a network consisting of genes from the p52-associated gene list as well as 13 imputed mediator genes (Fig. 4E). To test the strength of this protein interaction network, a survival analysis was performed based only on expression of the 13 imputed mediator genes. High expression of the imputed genes also correlated with shorter survival (log-rank test  $p = 0.047$  by expression summation,  $p = 0.097$  by votes) (Fig. 4F and Supplementary Fig. S3B), supporting the strength of this p52-derived interaction network in predicting patient prognosis.

We further validated our original 71 p52-associated gene signature in additional human lung adenocarcinoma datasets. Using normal and lung adenocarcinoma expression data from the BATTLE study<sup>21,22</sup>, we performed unsupervised hierarchical clustering and found that this p52-associated gene signature segregated normal and lung adenocarcinoma samples (Fisher's exact test  $p = 0.000003$ ) (Fig. 5A). We then performed a differential expression test for each gene in normal and lung adenocarcinoma samples, defining up-regulated genes as having both a  $\log_{2}FC > 1.3$  and  $FDR < 0.01$ . From the entire set of 22,216 genes, 9,832 (44.3%) were significantly up-regulated in tumors compared to matched normal samples whereas 54 (76.1%) of the 71 p52-associated genes were up-regulated in tumors, demonstrating a significant over-representation of p52-associated genes in tumor samples compared to normal in a second independent dataset (Fisher's exact test  $p = 0.0034$ ) (Fig. 5B). When evaluated as a group, mRNA levels of the 71 p52-associated genes were significantly increased in lung adenocarcinoma samples compared to normal lung tissue ( $p < 0.0001$ ) (Fig. 5C). In addition, we assessed the relationship between p52 expression and survival of lung adenocarcinoma patients in a composite dataset that includes expression data from several lung cancer cohorts<sup>23</sup>. Higher p100/p52 (*Nfkb2*) expression in lung adenocarcinomas correlated with reduced patient survival ( $p = 3 \times 10^{-7}$ ) (Supplementary Fig. S4B). In contrast, higher expression of *Nfkb1*, which encodes for the NF- $\kappa$ B family member p50, correlated with improved survival ( $p = 0.017$ ) (Supplementary Fig. S4A).

## Discussion

These studies provide evidence of the functional involvement of non-canonical NF- $\kappa$ B signaling in lung carcinogenesis, identifying p52 as a novel mediator of lung cancer progression. In human lung adenocarcinoma tumors, we observed frequent nuclear accumulation of p52, suggesting that p52 activation is common in lung carcinogenesis. Using transgenic mice with inducible expression of p52 in the airway epithelium, we found that p52 over-expression led to more tumors, larger tumors, and more advanced tumors after urethane injection. Our *in vitro* studies and gene expression microarray analysis suggest that p52 enhances tumor burden through the regulation of cell cycle genes that promote proliferation. Expression of p52-associated genes was significantly associated with lung tumors and increased expression of genes from a p52-derived protein-protein interaction network correlated with significantly reduced survival, even for Stage I patients. These findings have important implications for investigations into modulation of p52 or its downstream targets as a novel therapeutic approach in lung cancer.

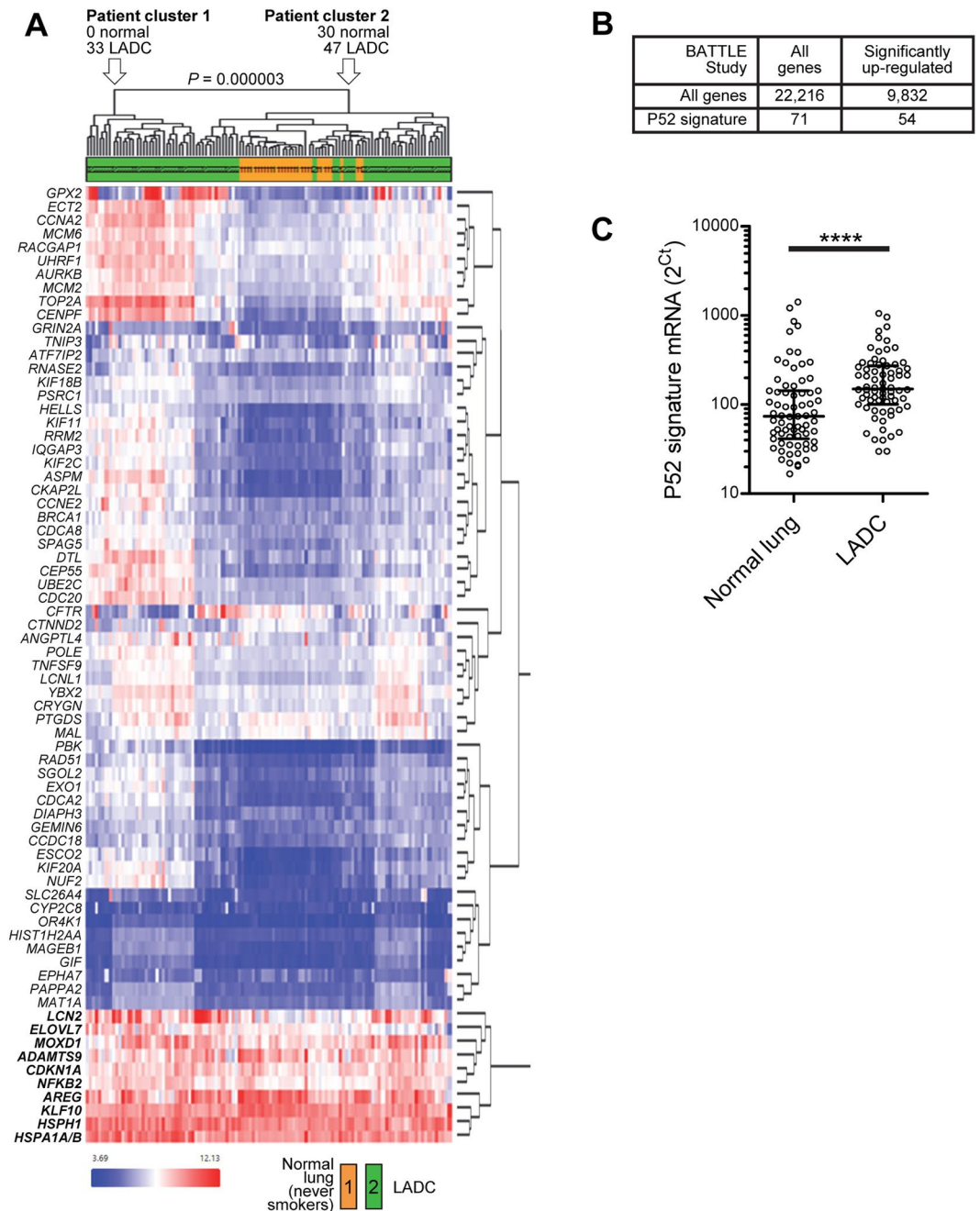
Even though nuclear localization of p52 has been observed in a number of different tumor types, our understanding of the function of p52 activation *in vivo* during carcinogenesis has remained limited due to the constraints of available models. In other studies using a mouse model with global truncation of the *Nfkb2* gene deleting the C-terminal inhibitory domain of p100, mice die from gastric hyperplasia<sup>14</sup>. However, unlike our studies, the individual contribution of inflammatory cells and surrounding stroma with dysregulated p52 activation was not separated from the effects of p52 activation in the epithelial cells of the stomach. Using a transgenic model with expression of the full *Nfkb2* gene under the control of the  $\beta$ -lactoglobulin milk protein promoter, Connelly *et al.* demonstrated ductal thickening and hyperplasia in the mammary gland after repeated pregnancies<sup>24</sup>, but in this model, expression was limited to the periods of milk production during the late stages of pregnancy and lactation. Our studies improved upon these models by using an inducible p52-expressing mouse. The advantages of a mouse model with inducible p52 expression are that p52 can be expressed with any cell type-specific promoter coupled to a reverse tetracycline transactivator, the duration of p52 expression can be manipulated, and p52 expression can be combined with other inducible mouse models, including oncogenic tumor models. Future studies using this p52-inducible model to modulate the duration of expression in the urethane model and to express p52 in other cell types and in combination with other inducible models will be valuable for developing a broader understanding of p52 activation in lung cancer and other contexts.



**Figure 4.** Expression of p52-associated genes correlates with prognosis of lung cancer patients. **(A)** Global and p52-associated genes upregulated in matched tumor and normal samples ( $n = 54$ ) from TCGA dataset. Up-regulated genes were defined as  $\log_{2}FC > 1$  and  $FDR < 0.01$  (hypergeometric test  $p = 6.0 \times 10^{-25}$ ). **(B)** Volcano plot of log fold changes and differential expression p-values of tumor versus normal samples. Black dots represent all genes in the expression data matrix. Red dots represent p52-associated genes. **(C,D)** Kaplan-Meier survival curves of overall **(C)** and Stage I **(D)** patient survival data based on summed expression of p52-associated genes in tumors (log-rank test,  $p = 0.004$  for overall;  $p = 0.036$  for Stage I). **(E)** Protein-protein interaction network generated using 35 p52-associated genes associated with tumors (see Fig. 4A and Supplementary Table S1). Pink circles are genes identified as p52-regulated genes through microarray analysis, and blue circles are imputed genes. **(F)** Kaplan-Meier survival curve of overall patient survival based on summed expression of imputed mediator genes (log-rank test,  $p = 0.047$ ).

Although p100/p52 expression appears to be common in lung adenocarcinomas, how this pathway becomes activated remains unclear. Proteolytic processing of p100 to p52 can be activated by a number of stimuli, including lymphotoxin  $\beta^{10}$ , CD40 ligand<sup>25</sup>, reactive oxygen species<sup>26</sup>, and STAT3 signaling<sup>27</sup>, suggesting that p52 activation can occur in a tumor as a result of oncogenic signaling or microenvironmental stimuli. Additional evidence indicates that crosstalk occurs between the canonical and non-canonical NF- $\kappa$ B pathways, suggesting that factors leading to canonical pathway activation could also activate p52<sup>28</sup>.

Like CCSP-p52 mice, activation of canonical NF- $\kappa$ B signaling in transgenic mice expressing constitutively active IKK $\beta$  in airway epithelium increases tumor burden following urethane treatment<sup>29</sup>. In this model,



**Figure 5.** Increased expression of p52-associated genes in lung adenocarcinoma samples from the BATTLE study. **(A)** Hierarchical clustering based on expression of 71 p52-associated genes in normal lung ( $n = 30$ ) and lung adenocarcinoma (LADC) ( $n = 80$ ) tissues. (Fisher's exact test  $p = 0.000003$ ). **(B)** p52 signature and global genes over-represented in lung adenocarcinoma samples compared with normal lung. Up-regulated genes were defined as  $\log_{2}FC > 1.3$  and  $FDR < 0.01$  (Fischer's exact test  $P = 0.0034$ ). **(C)** mRNA transcript expression levels from 71 p52-associated genes in normal lung tissue and lung adenocarcinoma samples. Raw data points and median with error bars that represent interquartile range are shown. Transcripts were not normally distributed by the Kolmogorov-Smirnov normality test and were hence analyzed by the Wilcoxon matched-pairs signed rank test (\*\*\*\* $p < 0.0001$ ).

canonical NF- $\kappa$ B activation causes a robust immune/inflammatory response, which promotes lung carcinogenesis through paracrine signaling<sup>29,30</sup>. In contrast, our studies indicate that non-canonical NF- $\kappa$ B signaling likely regulates genes that function in a cell autonomous manner to promote carcinogenesis.

Our finding that p52 can promote proliferation of airway epithelial cells during lung carcinogenesis is somewhat unexpected, as several previous studies investigating p52 expression in airway epithelial cells have shown that p52 can promote apoptosis of Club cells *in vivo*<sup>19</sup> and contribute to regulation of inflammatory cytokines *in vitro*<sup>31</sup>. In other studies using *in vitro* and xenograft models, however, p52 activation has been shown to promote



cell proliferation. Nadiminty *et al.* observed that p52 over-expression enhanced proliferation of prostate cancer cells in androgen-deprived conditions by stimulating Cyclin D1 expression<sup>32</sup>. Similarly, increased p100 expression in mammary glands led to hyperplasia and increased Cyclin D1<sup>24</sup>. Although we did not identify Cyclin D1 as an upregulated gene in our microarray analysis, both Cyclin E2 and Cyclin A2 were increased in CCSP-p52 lungs (Supplementary Table S1).

Among NF- $\kappa$ B family members, p52 is structurally most similar to p50. Kravtsova-Ivantsiv and colleagues found that p50 over-expression inhibits tumor growth in xenograft models<sup>33</sup>, and our survival analysis based on p50 expression in lung adenocarcinomas supports their observations. Despite the structural similarity between p50 and p52, we found that p52 over-expression enhances tumorigenesis and leads to reduced survival. These studies suggest that p52 and p50 function differently during tumorigenesis, which may be due to the differential affinities of each for specific promoter sequences or their propensity to interact with other NF- $\kappa$ B family members and transcriptional binding partners<sup>34,35</sup>. While nuclear p52 accumulation is considered indicative of non-canonical NF- $\kappa$ B pathway activation, a study by Zhao, *et al.* suggests that p52 is capable of forming heterodimers with all of the other NF- $\kappa$ B family members<sup>34</sup>. In addition, p52 has been shown to interact with Bcl-3<sup>36–39</sup>, p53<sup>40</sup>, and multiple epigenetic modifiers, including CBP, p300, and HDAC proteins<sup>38,40–43</sup>, and these interactions can be activating or inhibitory depending on the context. Future studies are necessary to understand p52 interactions in the nucleus and the mechanisms by which p52 influences tumor formation and progression.

In these studies, we took a gene list derived from p52 over-expression in a mouse model and demonstrated that these genes predict survival of early stage lung cancer patients. In addition, p52 expression in lung adenocarcinomas was inversely correlated with survival, similar to previous reports with the NF- $\kappa$ B family members p65 and RelB<sup>44,45</sup>. Although differences in gene expression between CCSP-p52 and WT mouse lungs were relatively modest, likely due to the small subset of transgene-expressing lung epithelial cells, we were able to identify genes with significant expression changes as a result of p52 over-expression, and these genes were highly related in function, indicating that p52 regulates expression of a network of related genes. Using gene expression profiling of *Nfkb2* knockdown fibroblasts, lanetti *et al.* described corresponding changes in several of the same genes identified in our study, supporting our identification of a discrete set of p52 target genes<sup>26</sup>. However, future studies comparing p52-induced gene expression changes during different stages of carcinogenesis and in the context of different oncogenic stimuli could be helpful for revealing additional p52 target genes.

Since our data suggest that p52 regulates a network of known cell cycle regulators that have well-established roles in carcinogenesis, p52 modulation may be an attractive therapeutic target for drug development. Although there are no available methods to directly target p52 at present, inhibitors for several of the nodes in our p52-derived protein interaction network have been developed and are being tested in clinical trials, including aurora kinase B, cyclin dependent kinase 1, and polo-like kinase 1, suggesting that treatment with these drugs alone or in combination may be beneficial in patients with p52 activation in their tumors.

## Materials and Methods

**Animal model.** CCSP-tTS/(tet-O)<sub>7</sub>-FLAG-p52 mice on an FVB background were bred to CCSP-rtTA homozygous mice (gift from Dr. J.A. Whitsett, University of Cincinnati, Cincinnati, OH) to generate CCSP-p52 mice, which inducibly express FLAG-tagged p52 in the airway epithelium as previously described<sup>19</sup>. CCSP-p52 mice generated from two separate CCSP-tTS/(tet-O)<sub>7</sub>-FLAG-p52 founder lines were used for these studies. Age- and sex-matched transgenic mice along with genotype-negative littermate controls (called WT in our studies) were used in experiments. Transgene expression was activated by administering doxycycline (dox, Sigma) at a concentration of 2 g/L along with 2% sucrose in drinking water. For urethane experiments, mice were placed on dox for 1 week prior to IP injection of urethane (1 g/kg body weight; Sigma-Aldrich) and remained on dox until euthanized. For 6 month tumorigenesis experiments, mice received weekly IP injections of urethane for 4 weeks. At sacrifice, lungs were lavaged, and total and differential BAL cell counts were determined as previously described<sup>2,29</sup>. The left lung was tied off and frozen, and the right lung was perfused and fixed by inflating with 10% neutral-buffered formalin. All animal studies were approved and conducted according to the guidelines of the Vanderbilt University Medical Center Institutional Animal Care and Use Committee.

**Histology and immunostaining.** After fixation, lungs were embedded in paraffin, sectioned (5  $\mu$ m), and hematoxylin and eosin (H&E)-stained for histological analysis and quantification of tumor burden. Three sections were analyzed per mouse, and each section was separated by 50  $\mu$ m. Tumor area was measured using Image-Pro Plus software (Media Cybernetics), and tumor number was enumerated on each section. AAH lesions and tumor histology were assessed by a pathologist blinded to the experimental groups. Tumors were classified as adenomas (clear border), minimally invasive adenocarcinomas (MIAs; border mostly defined with one invasive edge), and adenocarcinomas (locally invasive with no defined border). To assess p100/p52 expression in human lung adenocarcinomas, immunostaining using a p100/p52 antibody (C-5, Santa Cruz) was performed on a lung tumor tissue microarray (TMA) with 106 adenocarcinoma samples (spots) from 34 patients. Collection of tumor samples for generating the TMA was carried out in accordance with relevant guidelines and regulations, and experimental protocols were approved by the Vanderbilt University Medical Center Institutional Review Board (IRB 000616). Informed consent was obtained from all subjects. p100/p52 staining of tumor spots was scored by two independent readers. Scoring was performed as follows: 0 = no staining; 1 = diffuse cytoplasmic staining, no nuclear staining; 2 = dark granular cytoplasmic staining, no nuclear; 3 = nuclear staining in <25% of tumor cells; 4 = nuclear staining in >25% of tumor cells. Each tumor had at least two spots on the TMA, and the mean score was calculated for each tumor based on the individual score for each spot.

**Western blot analysis.** Nuclear protein was prepared from lung tissue using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific), separated by SDS-PAGE gel, transferred to nitrocellulose membranes, and probed using p100/p52 (4882, Cell Signaling) and nuclear loading control TATA-binding protein (TBP) (N-12, Santa Cruz). Immunodetection was performed using the corresponding AlexaFluor-conjugated antibodies and the Odyssey Infrared Imaging System (LI-COR Biosciences). All images were converted to grayscale.

**In vitro proliferation measurements.** Rat type II alveolar epithelial cell line RLE-6TN was obtained from ATCC and maintained at 37°C 5% CO<sub>2</sub> in DMEM (Invitrogen) with 4.5 g/l glucose and 2 mM L-glutamine, supplemented with 10% fetal bovine serum, 100 units/ml penicillin, and 100 µg/ml streptomycin. Cells were tested periodically for mycoplasma contamination using the Universal Mycoplasma Detection Kit (ATCC). Generation and validation of stable p52 over-expressing (RLE-p52) and control empty vector (RLE-EV) cells are described elsewhere<sup>19</sup>. RLE-EV and RLE-p52 cells were plated in 96 well plates for viability and BrdU incorporation measurements. 48 hours after plating, cells were incubated with BrdU for 4 hours, and BrdU incorporation was measured using the chemiluminescent BrdU Cell Proliferation ELISA (Roche). To ensure plating of equal numbers of cells, viability of RLE-EV and RLE-p52 cells was measured on the same day as a surrogate for cell number using the CellTiter-Glo Luminescent Cell Viability assay (Promega) according to the manufacturer's protocol.

**Microarray gene expression analysis.** The RNeasy Mini kit (Qiagen) was used to isolate whole lung RNA from the lungs of CCSP-p52 and WT mice treated with dox for 1 week and with or without intratracheal LPS administration, which was performed as part of a separate study [described in ref.<sup>19</sup>]. To minimize the effects of mouse-to-mouse-variability, RNA was pooled from 3 mice for each sample, resulting in a total of 8 samples (2 WT dox only, 2 WT with LPS, 2 CCSP-p52 dox only, 2 CCSP-p52 with LPS). RNA quality control, hybridization to the Affymetrix Mouse Gene 1.0 ST array, and array scanning were performed by the Vanderbilt Technologies for Advanced Genomics (VANTAGE) Core. Expression data have been deposited in Gene Expression Omnibus (Accession # GSE71648).

The Robust Multichip Average method<sup>46</sup> implemented in R package “oligo” (v1.28.3)<sup>47</sup> was employed to normalize imported raw data. Probe sets interrogating control, unmapped, or intron sequences were ignored, leaving 79% of probe sets for a differential expression analysis. Linear models and empirical Bayes methods implemented in R package “limma” (v3.20.8)<sup>48</sup> were applied to estimate log fold changes and p-values for the p52-vs-WT effect. Since samples with and without LPS stimulation were processed and microarray profiled at the same time, we assumed the random data variations were similar across all samples and built a linear model across the eight samples accounting for both p52 and LPS experimental factors<sup>49</sup>. The coefficient (estimated log fold change) and p-value associated with the p52 factor were retrieved for the selection of differentially expressed genes. This method enabled us to isolate the specific effects of p52 and LPS, allowing us to focus only on genes changed by p52 over-expression and effectively giving us 4 samples per group (12 mice per group) for our gene expression comparisons. When selecting top-ranking p52-induced genes, only probe sets mapped to known genes with positive expression changes were considered.

**Quantitative PCR.** Quantitative real-time PCR was performed using Sybr Green PCR Master Mix (Applied Biosystems) and the following primer sets: Ccna2 F: AAGAGAATGTCAACCCCGAAAAA; R: ACCCGTCGAGTCTTGAGCTT (PrimerBank ID: 161353443c1), Top2a F: CAACTGGAACATATACTGCTCCG; R: GGGTCCCTTTGTTTGTATCAGC (PrimerBank ID: 6755849a1), Ube2c F: CTCCGCCTTCCCTGAGTCA; R: GGTGCGTTGTAAGGGTAGCC (PrimerBank ID 21312888a1) and Gapdh (F: TGAGGACCAGTTGTCTCTCT R: CCCTGTTGCTGTAGCCGTAT)<sup>50–52</sup>. Expression values were normalized to Gapdh using the  $\Delta$ CT method.

**Statistics.** Unpaired student t-tests were performed for comparisons between two groups. Values are presented as the mean  $\pm$  SEM, and  $p \leq 0.05$  was considered statistically significant. Data were analyzed using GraphPad Prism 5.0 software (GraphPad Software, Inc.). Analysis of transcriptomic data is presented below.

**Human lung cancer expression data analysis.** Raw expression data for 513 lung adenocarcinoma patient samples, generated by the TCGA project, were accessed from the ICGC data portal (<https://dcc.icgc.org/>; release 17). The raw RSEM values<sup>53</sup> were multiplied by 10<sup>6</sup> to normalized values of Transcripts Per Million (TPM). Of these 513 samples, matched tumor samples and paracancerous normal samples were available for 54 patients. A paired t-test was performed for each gene, and the resultant p-values were adjusted to false discovery rates using the Benjamini-Hochberg method<sup>54</sup>.

Of the 513 lung adenocarcinoma samples, 437 primary lung adenocarcinoma samples had accompanying survival outcome data. In each survival analysis, samples were divided into two groups of equal sizes (termed high-expression and low-expression groups) by considering the collective expression of p52-associated genes. R package “survival” (v2.37-7)<sup>55</sup> was employed to conduct survival analyses. In survival analyses, survival time values were truncated to 5 years (1825 days) to reflect 5-year survival rates.

For the BATTLE study analysis, global gene expression profiles of normal lung and lung adenocarcinoma, as assessed by Human Gene 1.0 ST microarrays (Affymetrix, Sta. Clara, CA), were retrieved from Gene Expression Omnibus series GSE43458<sup>21,22</sup>. Raw data were analyzed using Affymetrix Gene Expression and Transcriptome Analysis Consoles.

**Gene functional analysis.** Functions enriched within p52-regulated genes were identified through a hypergeometric test with respect to GO terms<sup>56</sup>, using R package GOSTats (v1.7.4)<sup>57</sup>. For each GO term, genes from the reference set of whole human genome protein-coding genes and the concerned set of p52-regulated genes were respectively identified, and a p-value indicative of enrichment significance was calculated and further adjusted to

the FDR using the Benjamini-Hochberg method<sup>54</sup>. Resultant GO terms were pruned to lowest common ancestors according to the Biological Process semantic topology.

From the Human Protein Reference Database (HPRD, v9)<sup>58</sup>, a connected protein interaction network was derived that comprised 9,218 nodes and 36,728 edges. p52-regulated genes were mapped to this HPRD network, and then the Steiner minimal tree algorithm<sup>59</sup> was implemented to retrieve the most parsimonious subnetwork that joined together the terminal genes.

## References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7–30 (2016).
- Stathopoulos, G. T. *et al.* Epithelial NF- $\kappa$ B activation promotes urethane-induced lung carcinogenesis. *Proc. Natl. Acad. Sci. USA* **104**, 18514–18519 (2007).
- Meylan, E. *et al.* Requirement for NF- $\kappa$ B signalling in a mouse model of lung adenocarcinoma. *Nature* **462**, 104–107 (2009).
- Bassères, D. S., Ebbs, A., Levantini, E. & Baldwin, A. S. Requirement of the NF- $\kappa$ B subunit p65/RelA for K-Ras-induced lung tumorigenesis. *Cancer Res* **70**, 3537–3546 (2010).
- Tang, X. *et al.* Nuclear factor- $\kappa$ B (NF- $\kappa$ B) is frequently expressed in lung cancer and preneoplastic lesions. *Cancer* **107**, 2637–46 (2006).
- Dimitrakopoulos, F. I. D. *et al.* NSCLC and the alternative pathway of NF- $\kappa$ B: uncovering an unknown relation. *Virchows Arch.* **460**, 515–523 (2012).
- Giopanou, I. *et al.* Comprehensive evaluation of nuclear factor- $\kappa$ B expression patterns in non-small cell lung cancer. *PLoS One* **10**, e0132527 (2015).
- Franzoso, G. *et al.* Mice deficient in nuclear factor (NF)- $\kappa$ B/p52 present with defects in humoral responses, germinal center reactions, and splenic microarchitecture. *J. Exp. Med.* **187**, 147–159 (1998).
- Caamaño, J. H. *et al.* Nuclear Factor (NF)- $\kappa$ B2 (p100/p52) is required for normal splenic microarchitecture and B cell-mediated immune responses. *J. Exp. Med.* **187**, 185–196 (1998).
- Dejardin, E. *et al.* The lymphotoxin- $\beta$  receptor induces different patterns of gene expression via two NF- $\kappa$ B pathways. *Immunity* **17**, 525–535 (2002).
- Migliazza, A. *et al.* Heterogeneous chromosomal aberrations generate 3' truncations of the NFKB2/lyt-10 gene in lymphoid malignancies. *Blood* **84**, 3850–60 (1994).
- Neri, A. *et al.* B cell lymphoma-associated chromosomal translocation involves candidate oncogene lyt-10, homologous to NF- $\kappa$ B p50. *Cell* **67**, 1075–1087 (1991).
- Neri, A. *et al.* Molecular analysis of cutaneous B- and T-cell lymphomas. *Blood* **86**, 3160–3172 (1995).
- Ishikawa, H., Carrasco, D., Claudio, E., Ryseck, R.-P. & Bravo, R. Gastric hyperplasia and increased proliferative responses of lymphocytes in mice lacking the COOH-terminal ankyrin domain of NF- $\kappa$ B2. *J. Exp. Med.* **186**, 999–1014 (1997).
- Cogswell, P. C., Guttridge, D. C., Funkhouser, W. K. & Baldwin, A. S. Selective activation of NF- $\kappa$ B subunits in human breast cancer: potential roles for NF- $\kappa$ B2/p52 and for Bcl-3. *Oncogene* **19**, 1123–1131 (2000).
- Lessard, L., Bégin, L. R., Gleave, M. E., Mes-Masson, A.-M. & Saad, F. Nuclear localisation of nuclear factor-kappaB transcription factors in prostate cancer: an immunohistochemical study. *Br. J. Cancer* **93**, 1019–1023 (2005).
- Seo, S. I. *et al.* Immunohistochemical analysis of NF- $\kappa$ B signaling proteins IKK $\epsilon$ , p50/p105, p52/p100 and RelA in prostate cancers. *APMIS* **117**, 623–628 (2009).
- Wharry, C. E., Haines, K. M., Carroll, R. G. & May, M. J. Constitutive non-canonical NF $\kappa$ B signaling in pancreatic cancer cells. *Cancer Biol. Ther.* **8**, 1567–1576 (2009).
- Saxon, J. A. *et al.* p52 overexpression increases epithelial apoptosis, enhances lung injury, and reduces survival after lipopolysaccharide treatment. *J. Immunol.* **196**, 1891–1899 (2016).
- You, M., Candrian, U., Maronpot, R. R., Stoner, G. D. & Anderson, M. W. Activation of the Ki-ras protooncogene in spontaneously occurring and chemically induced lung tumors of the strain A mouse. *Proc. Natl. Acad. Sci. USA* **86**, 3070–3074 (1989).
- Kabbout, M. *et al.* ETS2 mediated tumor suppressive function and met oncogene inhibition in human non-small cell lung cancer. *Clin. Cancer Res.* **19**, 3383–3395 (2013).
- Kim, E. S. *et al.* The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov.* **1**, 44–53 (2011).
- Györfy, B., Surowiak, P., Budczies, J. & Lánczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* **8**, e82241 (2013).
- Connelly, L. *et al.* A transgenic model reveals important roles for the NF- $\kappa$ B alternative pathway (p100/p52) in mammary development and links to tumorigenesis. *J. Biol. Chem.* **282**, 10028–10035 (2007).
- Coope, H. J. *et al.* CD40 regulates the processing of NF- $\kappa$ B2 p100 to p52. *EMBO J.* **21**, 5375–5385 (2002).
- Iannetti, A. *et al.* Regulation of p53 and Rb links the alternative NF- $\kappa$ B pathway to EZH2 expression and cell senescence. *PLoS Genet.* **10**, e1004642 (2014).
- Nadiminty, N. *et al.* Stat3 activation of NF- $\kappa$ B p100 processing involves CBP/p300-mediated acetylation. *Proc. Natl. Acad. Sci. USA* **103**, 7264–7269 (2006).
- Liptay, S., Schmid, R. M., Nabel, E. G. & Nabel, G. J. Transcriptional regulation of NF- $\kappa$ B2: evidence for  $\kappa$ B-mediated positive and negative autoregulation. *Mol. Cell. Biol.* **14**, 7695–7703 (1994).
- Zaynagetdinov, R. *et al.* Epithelial nuclear factor- $\kappa$ B signaling promotes lung carcinogenesis via recruitment of regulatory T lymphocytes. *Oncogene* **31**, 3164–3176 (2011).
- Cheng, D.-S. *et al.* Airway epithelium controls lung inflammation and injury through the through the NF- $\kappa$ B pathway. *J. Immunol.* **178**, 6504–6513 (2007).
- Tully, J. E. *et al.* Cooperation between classical and alternative NF- $\kappa$ B pathways regulates proinflammatory responses in epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **47**, 497–508 (2012).
- Nadiminty, N., Chun, J. Y., Lou, W., Lin, X. & Gao, A. C. NF- $\kappa$ B2/p52 enhances androgen-independent growth of human LNCaP cells via protection from apoptotic cell death and cell cycle arrest induced by androgen-deprivation. *Prostate* **68**, 1725–1733 (2008).
- Kravtsova-Ivantsiv, Y. *et al.* KPC1-mediated ubiquitination and proteasomal processing of NF- $\kappa$ B1 p105 to p50 restricts tumor growth. *Cell* **161**, 333–347 (2015).
- Zhao, B. *et al.* The NF- $\kappa$ B genomic landscape in lymphoblastoid B cells. *Cell Rep.* **8**, 1595–1606 (2014).
- Siggers, T. *et al.* Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF- $\kappa$ B family DNA binding. *Nat. Immunol.* **13**, 95–102 (2011).
- Bours, V. *et al.* The oncoprotein Bcl-3 directly transactivates through  $\kappa$ B motifs via association with DNA-binding p50B homodimers. *Cell* **72**, 729–739 (1993).
- Bundy, D. L. & McKeithan, T. W. Diverse effects of BCL3 phosphorylation on its modulation of NF- $\kappa$ B p52 homodimer binding to DNA. *J. Biol. Chem.* **272**, 33132–33139 (1997).
- Rocha, S., Martin, A. M., Meek, D. W. & Perkins, N. D. p53 represses cyclin D1 transcription through down regulation of Bcl-3 and inducing increased association of the p52 NF- $\kappa$ B subunit with histone deacetylase 1. *Mol. Cell. Biol.* **23**, 4713–4727 (2003).

39. Wang, F., Shi, Y., Yadav, S. & Wang, H. p52-Bcl3 complex promotes cyclin D1 expression in BEAS-2B cells in response to low concentration arsenite. *Toxicology* **273**, 12–18 (2010).
40. Schumm, K., Rocha, S., Caamano, J. & Perkins, N. D. Regulation of p53 tumour suppressor target gene expression by the p52 NF- $\kappa$ B subunit. *EMBO J.* **25**, 4820–4832 (2006).
41. Stefano, V. D., Wang, B., Parobchak, N., Roche, N. & Rosen, T. RelB/p52-mediated NF- $\kappa$ B signaling alters histone acetylation to increase the abundance of corticotropin-releasing hormone in human placenta. *Sci. Signal.* **8**, ra85 (2015).
42. Vallabhapurapu, S. D. *et al.* Transcriptional repression by the HDAC4–RelB–p52 complex regulates multiple myeloma survival and growth. *Nat. Commun.* **6**, 8428 (2015).
43. Nadiminty, N. *et al.* Aberrant activation of the androgen receptor by NF- $\kappa$ B2/p52 in prostate cancer cells. *Cancer Res.* **70**, 3309–3319 (2010).
44. Zhang, Z., Ma, J., Li, N., Sun, N. & Wang, C. Expression of nuclear factor- $\kappa$ B and its clinical significance in nonsmall-cell lung cancer. *Ann. Thorac. Surg.* **82**, 243–248 (2006).
45. Qin, H. *et al.* Prognostic significance of RelB overexpression in non-small cell lung cancer patients. *Thorac. Cancer* **7**, 415–421 (2016).
46. Bolstad, B. M., Irizarry, R. A., Åstrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
47. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).
48. Smyth, G. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Gentleman, R., Carey, V., Dudoit, R. & Irizarry, W.) 397–420 (Springer, 2005).
49. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
50. Spandidos, A., Wang, X., Wang, H. & Seed, B. PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Res.* **38**, D792–799 (2010).
51. Spandidos, A. *et al.* A comprehensive collection of experimentally validated primers for polymerase chain reaction quantitation of murine transcript abundance. *BMC Genomics* **9**, 633 (2008).
52. Wang, X. & Seed, B. A PCR primer bank for quantitative gene expression analysis. *Nucleic Acids Res.* **31**, e154 (2003).
53. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
54. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
55. Therneau, T. & Grambsch, P. *Modeling Survival Data: Extending the Cox Model*. (Springer, 2000).
56. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
57. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
58. Keshava Prasad, T. S. *et al.* Human Protein Reference Database - 2009 update. *Nucleic Acids Res.* **37**, D767–772 (2009).
59. Klein, P. & Ravi, R. A nearly best-possible approximation algorithm for node-weighted steiner trees. *J. Algorithms* **19**, 104–115 (1995).

## Acknowledgements

This work was funded by the Department of Veterans Affairs Merit Review Award 1101BX002378. GTS was supported by European Research Council 2010 Starting Independent Investigator and 2015 Proof of Concept Grants (grant numbers # 260524 and # 679345, respectively). The Vanderbilt VANTAGE Core provided technical assistance for this work. VANTAGE is supported in part by CTSA Grant (5UL1 RR024975-03), the Vanderbilt Ingram Cancer Center (P30 CA68485), the Vanderbilt Vision Center (P30 EY08126), and NIH/NCRR (G20 RR030956). We thank Dr. Kang Tu for sharing his Steiner tree script for our use and Dr. Chen-Ching Lin for illuminating discussion about TCGA data analysis. Partial results are based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

## Author Contributions

Conception/design of work: J.S., T.B. Data collection: J.S., V.P., L.G., A.M. Data analysis and interpretation: J.S., H.Y., G.T.S. Drafting the article: J.S., H.Y., T.B. Critical revision of the article: J.S., H.Y., V.P., G.T.S., L.G., A.M., P.M., F.Y., Z.Z., T.B.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-24488-8>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018