



# Network Biology of Tumor Stem-like Cells Identified a Regulatory Role of *CBX5* in Lung Cancer

Yau-Hua Yu<sup>1,14</sup>, Guang-Yuh Chiou<sup>2,5</sup>, Pin-I Huang<sup>2,5</sup>, Wen-Liang Lo<sup>3,4</sup>, Chien-Ying Wang<sup>2,6</sup>, Kai-Hsi Lu<sup>7</sup>, Cheng-Chia Yu<sup>1,8</sup>, Gil Alterovitz<sup>9,10</sup>, Wen-Chien Huang<sup>11</sup>, Jeng-Fan Lo<sup>1,3</sup>, Han-Shui Hsu<sup>12,13</sup> & Shih-Hwa Chiou<sup>1,2,5</sup>

SUBJECT AREAS:

ONCOGENESIS

COMPUTATIONAL BIOLOGY

CANCER STEM CELLS

SYSTEMS BIOLOGY

Received  
21 May 2012

Accepted  
1 August 2012

Published  
16 August 2012

Correspondence and requests for materials should be addressed to Y.-H.Y. (yauhuayu@gmail.com) or S.-H.C. (shchiou@vghtpe.gov.tw)

<sup>1</sup>Institute of Oral Biology, National Yang-Ming University, Taipei, Taiwan, <sup>2</sup>Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan, <sup>3</sup>Faculty of Dentistry, National Yang-Ming University, Taipei, Taiwan, <sup>4</sup>Department of Stomatology, Taipei Veterans General Hospital, Taipei, Taiwan, <sup>5</sup>Department of Medical Research and Education, Taipei Veterans General Hospital, Taipei, Taiwan, <sup>6</sup>Department of Emergency, Taipei Veterans General Hospital, Taipei, Taiwan, <sup>7</sup>Department of Medical Research and Education, Cheng Hsin General Hospital, Taipei, Taiwan, <sup>8</sup>Institute of Oral Biology and Biomaterial Science, Chung-Shan Medical School, Taichung, Taiwan, <sup>9</sup>Children's Hospital Informatics Program at the Harvard/MIT Health Sciences and Technology Division, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Boston, United States, <sup>10</sup>Partners HealthCare Center for Personalized Genetic Medicine, Boston, United States, <sup>11</sup>Division of Thoracic Surgery, Department of Surgery, Mackay Memorial Hospital, Taipei, Taiwan, <sup>12</sup>Division of Thoracic Surgery and Department of Surgery, Taipei Veterans General Hospital, Taiwan, <sup>13</sup>Institute of Emergency and Critical Care Medicine, National Yang-Ming University, Taipei, Taiwan, <sup>14</sup>Harvard School of Dental Medicine, Boston, United States.

**Mounting evidence links cancers possessing stem-like properties with worse prognosis. Network biology with signal processing mechanics was explored here using expression profiles of a panel of tumor stem-like cells (TSLCs). The profiles were compared to their parental tumor cells (PTCs) and the human embryonic stem cells (hESCs), for the identification of gene chromobox homolog 5, *CBX5*, as a potential target for lung cancer. *CBX5* was found to regulate the stem-like properties of lung TSLCs and was predictive of lung cancer prognosis. The investigation was facilitated by finding target genes based on modeling epistatic signaling mechanics via a predictive and scalable network-based survival model. Topologically-weighted measurements of *CBX5* were synchronized with those of *BIRC5*, *DNMT1*, *E2F1*, *ESR1*, *MLH1*, *MSH2*, *RB1*, *SMAD1* and *TAF5*. We validated our findings in another Taiwanese lung cancer cohort, as well as in knockdown experiments using sh-*CBX5* RNAi both *in vitro* and *in vivo*.**

It has been long understood that cancer results from sequentially evolving genetic events. In solid tumors, malignancies are viewed as a collection of diseases that are heterogeneous in nature in: genomics, transcriptomic variations, and clinical outcomes. Lately, evidence has supported the claim that cancers possessing stem-like properties typically have a worse prognosis<sup>1,2</sup>. Other studies showed that overexpression of epithelial-mesenchymal transition transcription factors enhanced stem-like properties and increased the aggressiveness of tumor cells<sup>3-5</sup>. We established several panels of tumor stem-like cells (TSLCs) in head and neck<sup>2,6</sup>, brain<sup>7</sup>, and breast<sup>8</sup>. Here, we focus on lung adenocarcinomas (LACs), for which we have established a panel of lung TSLCs previously as well<sup>9</sup>. Indeed, lung cancer is one of the leading causes of cancer-related deaths worldwide<sup>10</sup>. Its highly invasive and metastatic phenotypes are the major reasons for treatment failure and poor prognosis<sup>11</sup>.

The study aim is to identify a critical target regulating both lung cancer survival and the stem-like properties of lung TSLCs. Lately, epigenetic regulators such as chromatin modifiers and polycomb group proteins were shown to be important players in cell fate decisions and reprogramming<sup>12</sup>. Nuclear perturbation was also known to play an important role in cancer biology<sup>13</sup>. Given the noisy and scarce nature of TSLCs, we first try to consolidate a consensus gene signature of low variation and consistent gene activities across the panel of TSLCs of different tissue of origins. Such gene signature is important that it could distinguish TSLCs from the parental tumor cells (PTCs) and from the human embryonic stem cells (hESCs). In this study, enriched signaling pathways of DNA methylation and establishment and/or maintenance of chromatin architecture were found in the consensus TSLC networks generated by the consensus gene signature. Base on the lung TSLC-specific gene signature, we further built the lung TSLC network model for survival prediction. *CBX5*, a chromatin regulator in the polycomb group,



was identified as a significant target in lung cancer survival analysis through a scalable network-based target identification process. Importantly, we verified that *CBX5* was also essential for the maintenance of aggressiveness and stem-like properties in lung TSLCs. We believe that our work supports a stronger claim of the epigenetic roadmap for the future understanding of lung cancer and lung TSLCs.

Our network models are derived from gene expression signature given the tumor stem-like states. Topologically-weighted signal mechanics incorporated in the network model are designed to dissect a possible role of functional noise. Several lines of evidence showed that stochastic fluctuations in gene expression were observed in embryonic stem cells leading to different lineages and cell fates<sup>14–16</sup>. Researchers also tried integration of electrical potentials in neurons and the brain functional magnetic resonance images<sup>17</sup> in relational network-based models to understand noisy signals. Therefore, we propose such network-based topologically-weighted signal model to estimate individual cancer survival time.

In summary, network-based models based on the TSLC panels were developed in this work to help understand the underlying biological perturbation leading to the variable survival time of lung cancer patients. It was expected that such network-based models could further elucidate the regulatory mechanisms leading to tumor invasion and metastasis. Data is available at GEO GSE35603. The R codes for all analysis could be accessed at <https://sites.google.com/site/nwtopsignalincancer/>.

## Results

From our previous experiences working on TSLCs, we have found that the panels of TSLCs were quite heterogeneous depending on the experimental cultivation procedures and/or the original tumor samples. In addition, due to the scarce nature of TSLCs, it was difficult to have a comprehensive transcriptome of TSLCs within a single tumor type. In this study, we therefore first aimed at establishing the soundness and importance of commonality across the panel of TSLCs. We then proceeded to develop an application model for lung cancer survivals based on the common and consistent gene expression profiles in lung CD133<sup>+</sup>-TSLCs.

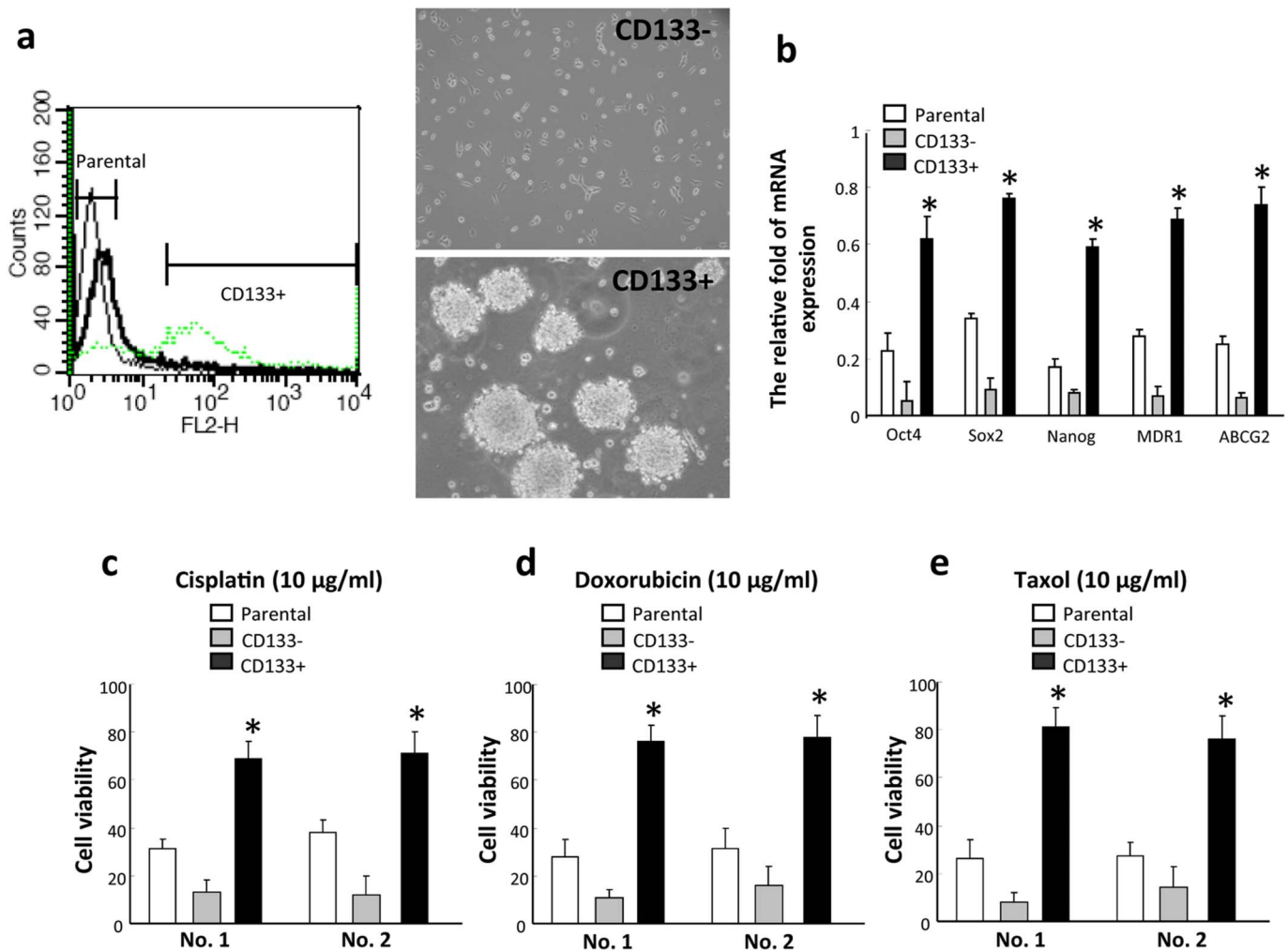
**Characterization of lung TSLCs.** Recent studies showed that expression of CD133 in lung cancers represents high tumorigenicity and resistance to cytotoxic therapy<sup>18,19</sup>. We previously reported greater chemoradioresistance of CD133<sup>+</sup>-TSLCs isolated from non-small cell lung cancers (NSCLCs) compared to CD133<sup>-</sup>-NSCLCs<sup>9</sup>. Here, we isolated CD133<sup>+</sup>-TSLCs from 7 NSCLCs (Fig. 1a). Isolated lung CD133<sup>+</sup>-TSLCs could form floating spheroid-like bodies in serum-free medium more easily than CD133<sup>-</sup>-NSCLCs. Quantitative RT-PCR results showed a higher level of transcripts of stemness genes (*Oct4*, *Sox2*, and *Nanog*) and drug resistant genes (*MDR1*, *ABCG2*) in lung CD133<sup>+</sup>-TSLCs (Fig. 1b). Lung CD133<sup>+</sup>-TSLCs displayed not only higher invasion activity, as well as enhanced foci formation, but also resistance to cisplatin, doxorubicin, and taxol (Fig. 1c–e). *In vivo*, transplants of lung CD133<sup>+</sup>-TSLCs exhibited more aggressiveness tumorigenicity in the lungs (Table 1).

**Distinct transcriptional patterns of the inter-modular hubs of the consensus TSLC networks.** First, we performed differential expression analysis between panels of TSLCs vs. PTCs for each tumor tissue type or experimental technique to come up with eight gene lists. These lists are of similar length, setting the 87.5% of absolute value of fold changes as the filtering threshold, merged as geneset A (Fig. S1, Table S1). Second, we ranked the top 500 probes with minimal transcriptional variability within TSLCs. A nonredundant geneset B consisting of 459 probes of low transcriptional variability in TSLCs as well as geneset A was summarized. Third, we filtered out gene signatures with inconsistent activities of TSLCs

comparing to PTCs to be the concordant geneset C. We identified a consensus gene-list of 64 probes characterized with low variation and commonality (*Iv\_com*). They were found not only having concordant gene activities in at least two tumor types/experimental conditions, but also either differentially expressed in at least two tumor types/conditions (n=18) or with minimal transcriptional variation in TSLCs (n=46). In clustering analysis, consensus gene signature of *Iv\_com* demonstrated its capability of best separation across the three panels of cells. Hierarchical clustering and non-metric multi-dimensional scaling of TSLCs, PTCs, and hESCs based on *Iv\_com* were displayed in Fig S2. By using *Iv\_com* as inputs in the Ingenuity Pathways Analysis (IPA), DNA methylation and transcriptional repression signaling pathways were found significantly enriched (Fisher exact test,  $-\log(P\text{-value})=4.17$ ). The output literature networks were merged with human protein-protein interactions (PPIs). We consolidated links in the merged networks by setting threshold of gene-gene co-expression within TSLCs. There are 77 links co-expressed among all TSLCs with abs(Pearson Correlation Coefficients; PCCs) >0.4 out of the total 161 links in the merged networks (Fig. S3; Table S2&3). PCCs of 77 links were mostly positive with the maximum of 0.92 between *HNRNPD* and *ILF3*. There are 49 genes, topologically categorized as 12 inter-modular hubs, 22 intra-modular hubs, and 15 periphery genes, in the consensus TSLCs networks. Of note, in the group of inter-modular hubs, averaged gene activities (*Exprs*) and SNR of TSLCs were statistically different from those of hESCs or those of PTCs (Fig. S2). *DNMT3A* was the only gene that distinguished three panels (Fig. S4). Gene Ontology (GO) functional annotation revealed differences in the biological processes characterized by these two different kinds of hubs (Table 2). Intra-modular hubs were all membrane bound intracellular organelles and 43% of them participated in the establishment and/or maintenance of chromatin architecture. All of the inter-modular hubs had molecular function of protein binding.

**Network signaling in lung-TSLC networks.** There are 96 genes - 25 inter-modular hubs, 44 intra-modular hubs, and 27 periphery genes - and 144 links in the lung-TSLC networks (Fig. S5; Table S4&5). Network-based survival analyses were conducted using two different sets of member genes, i.e. hub genes only ( $N_j=69$ ) or all genes in the lung-TSLC networks ( $N_j=96$ ), as well as using different combination of weights and weighting genes ( $N_i$ ): (1) intra-modular hubs weighted by degrees; (2) inter-modular hubs weighted by degrees; (3) inter-modular hubs weighted by focality; (4) intra- and inter-modular hubs weighted by degrees; and (5) intra-modular hubs weighted by degrees plus inter-modular hubs weighted by focality. We calculated the measurements of *Exprs*, *wt.Exprs*, *Mag*, *Spec*, and SNR derived from each combinatory network model ( $N_i$  vs.  $N_j$ ) and tested them in the survival analyses. By grouping lung cancer patients into quartiles given the network-based measurements, we found that at least one type of measurement could significantly rank patients into 2 to 4 risk groups (Table S6). To eliminate the possibility of the sample size in each dataset being too small, we further conducted meta-analyses with the pooled metastasis-free survivals (MFS; n=374) and overall survivals (OS; n=828). *Exprs* values of inter-modular hubs were consistently significant predictors of OS and MFS. With regard to the MFS, measurements of *Exprs*, *wt.Exprs*, *Spec*, and SNR of both the intra- and inter-modular hubs demonstrated trend-like significance (Table S6). We speculated that the lung-TSLC network model might be more sensitive to tumor progression.

**Identifying a regulatory role of *CBX5* in lung-TSLC networks modulating variation of lung cancer survivals.** To identify potential targets in the lung-TSLC networks, we tried out each single gene as the weighting gene in the survival analyses ( $N_i=1$ ;  $N_j=69$ ). Network-based measurements were calculated and first



**Figure 1 | Characterization of lung TSLCs.** (a) Lung CD133<sup>+</sup>-TSLCs were sorted and characterized by FACS assay and cultured in bFGF and EGF with DMEM serum-free medium. (b) The relative mRNA levels of *Oct4*, *Sox2*, *Nanog*, *MDR1*, and *ABCG2* were measured for the PTCs, CD133<sup>-</sup>-NSCLCs, and lung CD133<sup>+</sup>-TSLCs. (c–e) Evaluation of the cell viability under treatment of cisplatin, doxorubicin, and taxol for the PTCs, CD133<sup>-</sup>-NSCLCs, and lung CD133<sup>+</sup>-TSLCs. (\**P* < 0.05) All results shown are means of three independent experiments ± SD. Samples were isolated from No.1 patient listed in Table 1.

tested by survival analyses based on quartiles, and then tested by linear model fit with survival times. Genes showing statistical significance in multiple tests were identified into 3 groups: (1) OS-related; *MLH1* and *SMAD1*; (2) MFS-related; *CBX5*, *CPSF1*, *DNMT1*, *HNF1*, *IRS1*, *KPNA2*, *MSH2*, and *RASA1*; and (3) OS/MFS-related; *CDC2*, *COL18A1*, *RACGAP1*, and *SHC1*.

*CBX5* was chosen as a target for further validation for its potential role in lung cancer survival as well as in lung TSLCs based on the MFS analyses using the public lung cancer transcriptome. Foremost, levels of *Exprs*, *Spec*, and *SNR* of *CBX5* in quartiles were all significantly demonstrating dosage-like effects. Moreover, levels of *Exprs* and *SNR* of *CBX5* were significantly correlated with the MFS survival time in the metastasis-free group (Fig. 2a–c). In addition, a general linear model fit existed between *Spec* of *CBX5* and the reciprocal of MFS time in the metastasis group (Fig. 2b). It is worthy to note that *CBX5* was originally found differentially expressed in TSLCs of atypical teratoid/rhabdoid tumor (AT/RT-TSLCs) and consistently induced in lung-TSLCs. The topological characteristics of *CBX5* in the network models might provide a possible explanation of its role in TSLCs. *CBX5* was an intra-modular hub connected with *RB1* and *E2F1* in the lung-TSLC networks as well as an inter-modular hub connected with *DNMT3A* in the consensus TSLC networks.

In order to determine whether *CBX5* participated in lung tumorigenesis, we examined the levels of *CBX5* in 20 pairs of LAC samples (T) vs. the corresponding controls (N) by qRT-PCR analysis. RNA transcripts of *CBX5* were significantly higher in the tumor samples as well as in the metastatic lesions (Fig. 2d). We further collected another Taiwanese validation cohort of LAC patients for immunohistochemical staining (Table S7). The results supported that the *CBX5*-positive LAC cases were associated with worse overall survivals (Fig. 2e).

**Validation of *CBX5* in regulating self-renewal of lung TSLCs.** We tried to verify the significance of *CBX5* in the tumorigenicity and invasiveness of lung cancers by sh-RNAi knockdown of *CBX5* in lung TSLCs (Fig. 3a). We showed that the capabilities of sphere formation, colony formation, and migration/invasiveness of CD133<sup>+</sup>-TSLCs treated by sh-*CBX5* RNAi were indeed significantly inhibited (Fig. 3b–d). Additionally, the percentages of CD133<sup>+</sup>-TSLCs and side population (SP) cells treated by sh-*CBX5* RNAi were dramatically decreased (Fig. 3e–g).

In order to understand the gene-gene interplays of *CBX5* in lung-TSLC networks, we calculated the pair-wise correlations between levels of *Exprs*, *Mag*, *Spec*, and *SNR* of *CBX5* with those of the survival significant genes using the lung cancer transcriptome. We


**Table 1 | Characterization of tumorigenicity and effects of sh-CBX5 RNAi in lung CD133<sup>+</sup>-TSLCs from 7 non-small cell lung cancer (NSCLC) patient**

NSCLC Case	Age/Sex/Type/Stage	CD133 <sup>+</sup> (%)	Spheres Formation	Number of cells injected (Number of animals having tumor/Total of 3 animals in each experiment)					
				CD133 <sup>+</sup>	CD133 <sup>+</sup> sh-RNA vector	CD133 <sup>+</sup> sh-CBX5	CD133 <sup>+</sup> IR-4 Gy	CD133 <sup>+</sup> sh-CBX5 & IR 4 Gy	CD133 <sup>-</sup>
1	82/M/AD/IIIa	7.1	Yes	1,000 (3/3)	1,000 (2/3)	1,000 (0/3)	1,000 (2/3)	1,000 (0/3)	1,000 (0/3)
				3,000 (3/3)	3,000 (3/3)	3,000 (0/3)	3,000 (2/3)	3,000 (0/3)	3,000 (0/3)
				10,000 (3/3)	10,000 (3/3)	10,000 (1/3)	10,000 (2/3)	10,000 (0/3)	10,000 (0/3)
2	59/F/AD/IIb	4.8	Yes	1,000 (2/3)	1,000 (1/3)	1,000 (0/3)	1,000 (1/3)	1,000 (0/3)	1,000 (0/3)
				3,000 (1/3)	3,000 (1/3)	3,000 (0/3)	3,000 (2/3)	3,000 (0/3)	3,000 (0/3)
				10,000 (3/3)	10,000 (3/3)	10,000 (2/3)	10,000 (2/3)	10,000 (1/3)	10,000 (0/3)
3	63/F/AD/IIb	3.6	Yes	1,000 (1/3)	1,000 (1/3)	1,000 (0/3)	1,000 (1/3)	1,000 (0/3)	1,000 (0/3)
				3,000 (2/3)	3,000 (1/3)	3,000 (0/3)	3,000 (1/3)	3,000 (0/3)	3,000 (0/3)
				10,000 (2/3)	10,000 (2/3)	10,000 (1/3)	10,000 (2/3)	10,000 (0/3)	10,000 (0/3)
4	70/M/SC/IIb	11.2	Yes	1,000 (3/3)	1,000 (3/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)
				3,000 (3/3)	3,000 (3/3)	3,000 (1/3)	3,000 (0/3)	3,000 (0/3)	3,000 (0/3)
				10,000 (3/3)	10,000 (3/3)	10,000 (3/3)	10,000 (2/3)	10,000 (0/3)	10,000 (0/3)
5	75/M/AD/IIa	3.9	Yes	1,000 (1/3)	1,000 (2/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)
				3,000 (2/3)	3,000 (1/3)	3,000 (0/3)	3,000 (1/3)	3,000 (0/3)	3,000 (0/3)
				10,000 (2/3)	10,000 (2/3)	10,000 (1/3)	10,000 (2/3)	10,000 (0/3)	10,000 (0/3)
6	67/M/AD/IIb	15.5	Yes	1,000 (3/3)	1,000 (3/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)
				3,000 (3/3)	3,000 (2/3)	3,000 (1/3)	3,000 (3/3)	3,000 (0/3)	3,000 (0/3)
				10,000 (3/3)	10,000 (3/3)	10,000 (2/3)	10,000 (3/3)	10,000 (1/3)	10,000 (1/3)
7	68/M/SC/IIb	2.3	Yes	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)	1,000 (0/3)
				3,000 (1/3)	3,000 (1/3)	3,000 (0/3)	3,000 (1/3)	3,000 (0/3)	3,000 (0/3)
				10,000 (2/3)	10,000 (1/3)	10,000 (1/3)	10,000 (2/3)	10,000 (0/3)	10,000 (0/3)

NSCLC: non-small cell lung carcinoma. Tumor types- AD: lung adenocarcinoma; SC: lung squamous cell carcinoma. Tumor grade/stage: TMN staging. sh-CBX5 & IR 4 Gy: combined sh-CBX5 RNAi with IR (4Gy) treatment. Cells were transplanted into NOD-SCID mice through the tail vein. After 8 weeks of transplantation, the tumorigenic ability of tumor-bearing NOD-SCID mice was measured by histological survey in the whole lung.

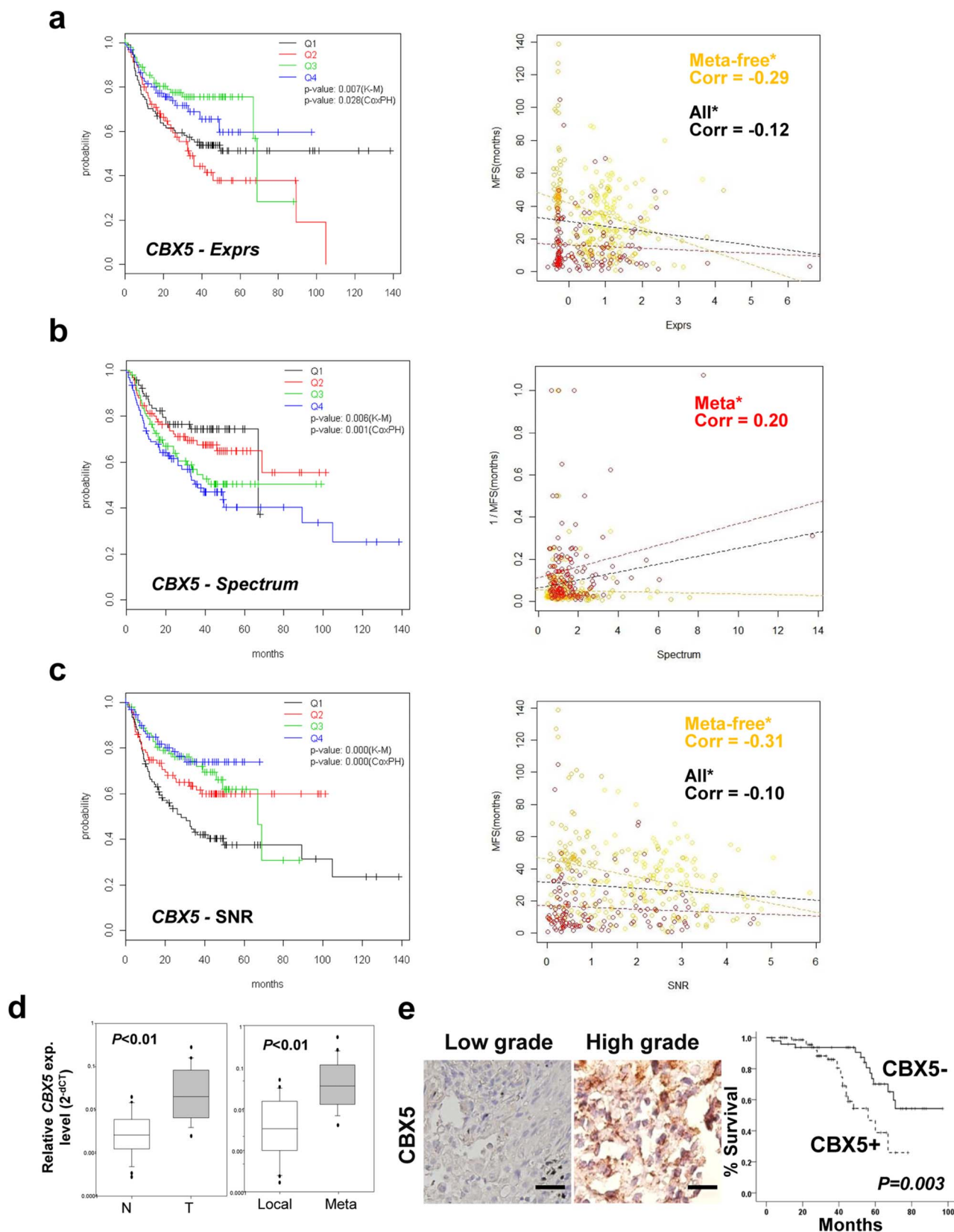
found that *CBX5* was significantly correlated with the survival significant genes such as *BIRC5* and *DNMT1* among the metastasis-free patients. The correlations between the *Spec* or SNR levels were higher than those of the *Exprs* or *Mag*. These results indicated that gene-gene regulatory controls were indeed synchronized under such a network-based model, especially when taking into account gene membership, network topology, as well as signal stochasticity. We further tried to experimentally validate the identified correlated synchronization between *CBX5* vs. the survival significant genes using *in vitro* sh-CBX5 RNAi inhibition in lung CD133<sup>+</sup>-TSLCs. Using RT-PCR, we detected decreased mRNA expression levels of *BIRC5*, *DNMT1*, *E2F1*, *ESR1*, *MLH1*, *MSH2*, *RBI*, *SMAD1*, *SIN3A* and *TAF5* and in the lung CD133<sup>+</sup>-TSLCs treated with sh-CBX5 RNAi (Fig. 3h). Our results suggested that knockdown of *CBX5* in lung-TSLCs could simultaneously inhibit these correlated survival significant genes.

**Validation of *CBX5* in modulating tumorigenicity and aggressiveness of lung carcinoma *in vivo*.** *In vivo* models were utilized to further examine the effect of sh-CBX5 RNAi knockdown. By injecting 2×10<sup>5</sup> sh-CBX5 RNAi treated lung CD133<sup>+</sup>-TSLCs vs. sh-Luc controls through tail vein after 8 weeks, we demonstrated that the tumorigenic engraftment, tumor growth rate (Fig. 4a), and metastatic tendency to lung by lung CD133<sup>+</sup>-TSLCs (Fig. 4b&c) were prominently blocked by sh-CBX5 RNAi knockdown. For lung cancers, surgery is the current standard of care treatment. However, for locally advanced lung tumors (stage 3b or above) that cannot be surgically removed, treatment with combined radiation and chemotherapy would be given to improve survivals. Therefore, given the aggressive nature of lung CD133<sup>+</sup>-TSLCs, we further tested to show that treatment of sh-CBX5 RNAi significantly increased the radiosensitivity of CD133<sup>+</sup>-TSLCs *in vitro* as well (Fig. 4d). Mice

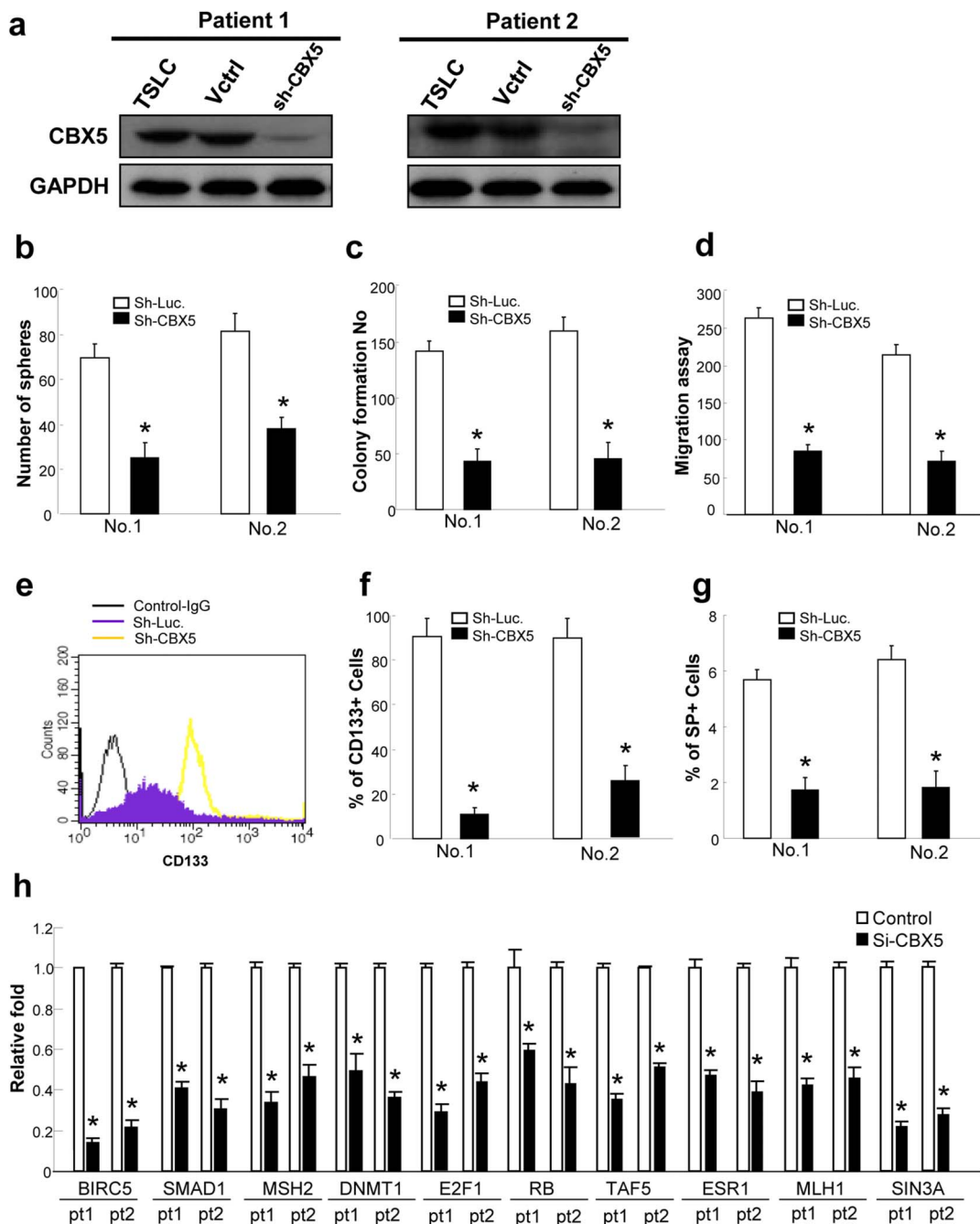
**Table 2 | Gene Ontology Functional Enrichment Annotation of consensus TSLC network genes with different gene grouping according to the topological characteristics**

Gene grouping	GO Category	GO Term	%	P-Value
All genes	BP	GO:0006355~regulation of transcription, DNA-dependent	46%	1.8E-06
	BP	GO:0042981~regulation of apoptosis	27%	4.0E-08
	CC	GO:0031981~nuclear lumen	25%	2.9E-06
	BP	GO:0006325~establishment and/or maintenance of chromatin architecture	21%	3.2E-07
Inter-modular hubs	MF	GO:0005515~protein binding	100%	8.1E-05
	BP	GO:0043170~macromolecule metabolic process	83%	0.02
	BP	GO:0048731~system development	42%	0.03
Intra-modular hubs	BP	GO:0006468~protein amino acid phosphorylation	33%	0.01
	CC	GO:0043231~intracellular membrane-bound organelle	100%	1.3E-07
	BP	GO:0043170~macromolecule metabolic process	95%	6.4E-06
	BP	GO:0050794~regulation of cellular process	90%	1.8E-08
	BP	GO:0006355~regulation of transcription, DNA-dependent	71%	9.2E-08
	BP	GO:0006325~establishment and/or maintenance of chromatin architecture	43%	2.9E-09





**Figure 2** | Network-topologically-based measurements of *CBX5* in lung-TSLC network model modulated variability of lung cancer survivals. (a–c) MFS analyses of quartile groups of *Exprs*, *Spec*, and *SNR* of *CBX5*. Xy plots with general linear model (GLM) fits for *Exprs* and *SNR* of *CBX5* vs. MFS time. Xy plots with GLM fit for *Spec* of *CBX5* vs. the reciprocal of MFS time. Metastasis patients colored red and metastasis-free yellow. (\* $P < 0.05$ ) (d) Levels of *CBX5* mRNA by quantitative real-time PCR from 20 pairs of primary LAC vs. adjacent non-tumorous lung tissues. Levels of *CBX5* mRNA between local lung vs. metastatic lesions of 10 patient-pairs were also shown. Results are means of 3 independent experiments  $\pm$  SD. (e) Representative results of immunohistochemical staining for *CBX5* in LAC patients at different grades (left, low-grade; right, high-grade). Overall survival analysis according to the *CBX5* expression levels in 125 Taiwanese LAC patients.



**Figure 3 | Knockdown of *CBX5* with sh-RNAi *in vitro*.** (a) Western blot of knockdown of *CBX5* in lung CD133<sup>+</sup>-TSLCs derived from two patients (No.1&2). The abilities of (b) sphere formation, (c) colony formation, and (d) migration in CD133<sup>+</sup>-TSLCs treated with sh-*CBX5* RNAi were decreased. The percentages of (e-f) CD133<sup>+</sup>-TSLCs and (g) SP cells were significantly reduced. (h) Using RT-PCR, we measured suppression of expression levels of *BIRC5*, *SMAD1*, *MSH2*, *DNMT1*, *E2F1*, *RB1*, *TAF5*, *ESR1*, *MLH1*, and *SIN3A* in the sh-*CBX5* RNAi treated lung CD133<sup>+</sup>-TSLCs. These genes were identified by statistically significant correlation with *CBX5* in lung cancer survival analysis. All data shown are the mean  $\pm$  SD of 3 experiments. (\* $P < 0.05$ )

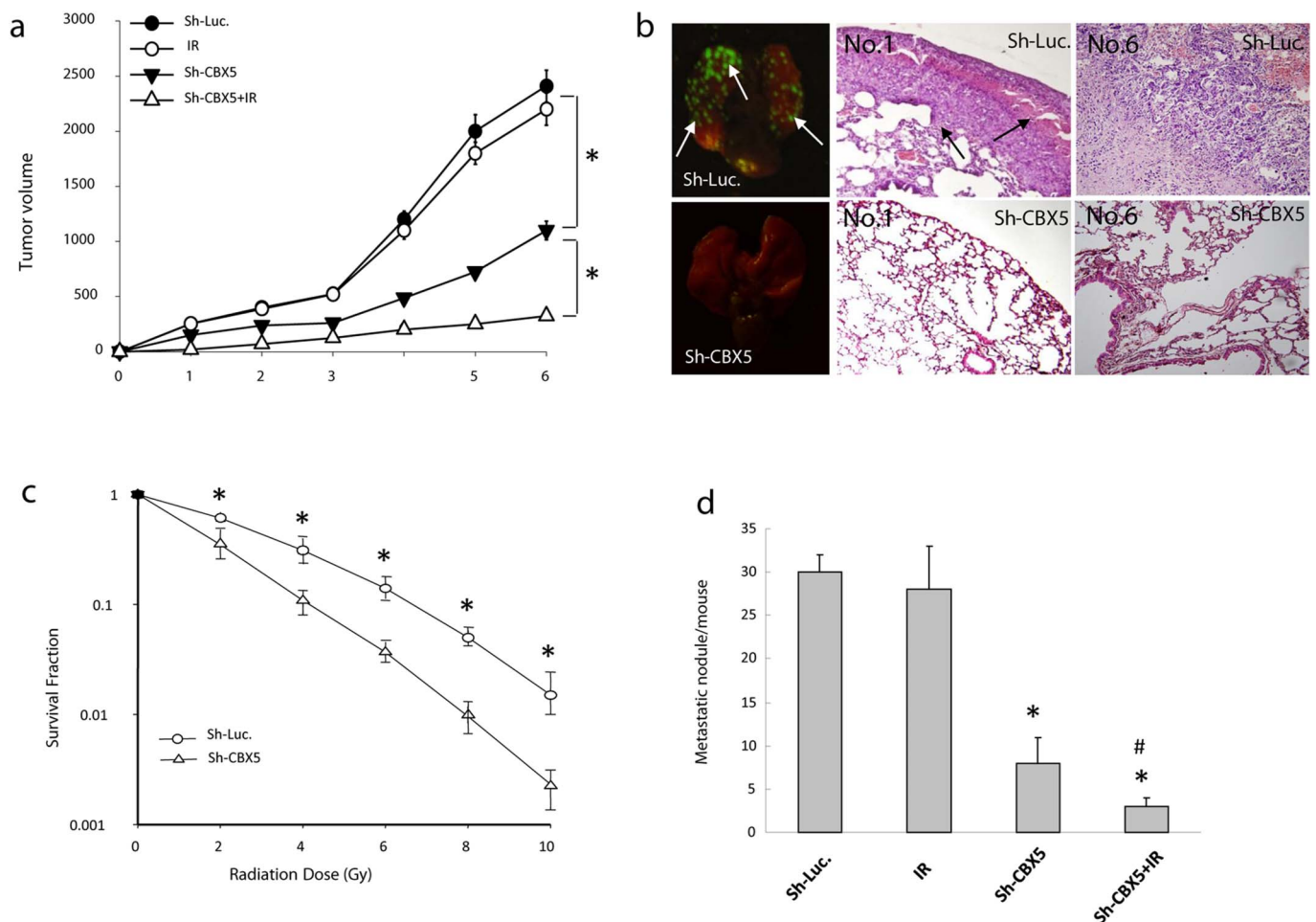
transplanted with the sh-*CBX5* RNAi-treated lung-TSLCs had significantly prolonged survivals as well (data not shown).

## Discussion

We have identified *CBX5* as a potential target regulating lung cancer survivals and the stem-like properties of lung CD133<sup>+</sup>-TSLCs. Moreover, interplays of *CBX5* with other genes in the lung-TSLC network model were statistically tested and experimentally validated. Lung cancer patients of higher *CBX5* gene activities were of poorer prognosis and the knockdown of *CBX5* with sh-RNAi in lung CD133<sup>+</sup>-TSLCs demonstrated lessened aggressiveness *in vivo*. We demonstrated that a scalable and predictable target identification

approach was feasible, given the context of network topology and signaling mechanics.

*CBX5*, a highly conserved nonhistone protein containing chromatin organization modifier domain, i.e. chromodomain, belongs to the heterochromatin protein family<sup>20</sup>. In rodent and *D. melanogaster* cells, *CBX5* was found to interact with H3K9me3 or colocalized with H3K9me to the heterochromatin regions<sup>21</sup>. The role of heterochromatin in transcriptional gene silencing and long-range chromatin interactions has been well-established. However, in mammalian cells, *CBX5* and H3K9me were found to associate with coding regions of activated genes, although the possible mechanism was unclear. Evidence also showed that *CBX5* served as a common gene



**Figure 4** | Inhibition of *CBX5* lessened tumorigenicity and aggressiveness of lung TSLCs *in vivo*.  $2 \times 10^5$  lung CD133<sup>+</sup>-TSLCs treated with sh-*CBX5* RNAi or sh-Luc controls were injected through tail vein of NOD-SCID mice (n=6 per group). The animals were sacrificed and examined 8 weeks after injection. (a) Tumor volume and (b) the number of metastatic foci (arrows) in the lungs were analyzed by *ex vivo* GFP imaging and histological examination. (c) The effects of IR of 2, 4, 6, 8, and 10 Gy were evaluated in CD133<sup>+</sup>-TSLCs treated with sh-Luc or sh-*CBX5* RNAi. Data shown are the mean  $\pm$  SD of 3 experiments. (d) Knockdown of *CBX5* in lung CD133<sup>+</sup>-TSLCs effectively reduced the number of lung metastases in transplanted mice. Treatment of sh-*CBX5* RNAi combined with 4 Gy IR further enhanced the anti-tumor efficacy (\* $P < 0.05$ ).

expression signature shared by human mature oocytes and embryonic stem cells<sup>22</sup>. Recently, Wong and colleagues identified that *ATRX* working together with H3.3 and *CBX5* might be a key regulator of ES-cell telomere chromatin<sup>23</sup>. Here, *CBX5* was identified in both the consensus TSLC and the lung-TSLC network models. Our findings further supported a recent finding<sup>24</sup> that *CBX5* was essential in the maintenance of leukemia stem cells (LSCs). To date, we are the first to report *CBX5* playing an essential role in the regulatory control of lung-TSLCs, as well as in malignant lung carcinomas. Survival significant genes identified from our analysis, specifically the identified target gene *CBX5*, again highlighted the importance of epigenetic regulatory controls. Thus, the lung-TSLC network model provided a link between experimentally cultivated lung-TSLCs and clinical lung cancer survival times, with statistical significance and mechanistic understandings.

Recent works by Bröske and colleagues demonstrated the indispensability of *DNMT1* for the cell-autonomous survival of hematopoietic stem cells (HSCs) and LSCs<sup>25</sup>. *De novo* methylation by *DNMT3A* and *DNMT3B* was also shown essential for HSCs renewal but not for differentiation<sup>26</sup>. Our findings of *DNMT3A* in the TSLC-consensus network, and *DNMT1* synchronized with *CBX5* in the lung-TSLC networks, were compatible to the above-mentioned reports. We recognized that the network models were built on known gene interactions and knowledge. Nevertheless, high co-expression

in TSLCs lent a better support of their validity. PCC demonstrating co-expression of *DNMT1* with *E2F1* was 0.98 and with *BIRC5* 0.87; and PCC of *DNMT3A* with *CBX5* was 0.72, with *EED* 0.62, and with *MYC* 0.48, respectively. Collectively, we supported and extended the importance of epigenetic regulations of TSLCs. However, it remains an open question in fully understanding the underlying regulation.

Network-based survival models have been developed for breast cancer<sup>27</sup> and glioblastoma<sup>28</sup>. We are the first to address the stochastic gene expression activities embedded in biological networks by summarizing them in the noise-like *Spec*, as well as SNR, in malignant lung carcinomas. This approach provides each patient a unique estimated profile to summarize the variable transcriptional signature within the same set of genes in a network model. In conclusion, we demonstrated that stochastic element of transcriptional profiles of lung cancers, given the relational model based on the lung-TSLC networks, could be useful in estimating the prognostic survival time. Last, the methodology is generic and future exploitation in other research areas will establish the validity of its robustness and applicability.

## Methods

**Microarray data.** (1) PTCs and TSLCs: The cultivated TSLCs were of six tissue of origins: breast, lung, colon, head and neck, glioblastoma, and AT/RT, whose parental cells were MCF7, A549, SW480 and HT29, FaDu and SAS, PT1 (primary culture) and





U87, ATRT, BT, BT6, and BT12, respectively. TSLCs were cultivated by methods described elsewhere<sup>2,3,6–9</sup>. For lung TSLCs, we isolated CD133<sup>+</sup>-TSLCs from tissue samples of NSCLC patients using the magnetic bead method of FACS assay. We extracted and purified total RNA according to procedures published elsewhere<sup>2</sup>. RNA was hybridized on microarrays of Affymetrix GeneChip HG-U133 plus 2.0 at the genomic core facilities at the National Yang-Ming University Genome Research Center. (2) hESCs: Oocytes (GSE12034, N=3) and human embryonic stem cells (hESCs) microarrays were downloaded from the Gene Expression Omnibus (GEO) of NCBI (GSE7879: VUB01, N=3; SA01, N=3; and Sheff4 hESCs, N=3. GSE9440: T3ES, N=3. H1 hESCs, N=4, from GSE9196 and GSE9510. H9 hESCs, N=6, from GSE9196, GSE9510, and GSE9940). Please see Table S1 for details. Gene expression data of all three panels could be downloaded from GEO GSE35603. (3) **Lung cancer transcriptome**: Eight datasets with lung cancer survivals were downloaded from GEO and supplements as described<sup>29–32</sup>.

**A precompiled gene set.** A gene list was compiled (5812 GeneIDs) to incorporate properties of migration<sup>33,34</sup>, stemness<sup>13,35</sup>, calcium-related processes<sup>36</sup>, and cancer-specific transcript variants<sup>37,38</sup> for gene expression analysis.

**Gene expression analysis.** Please see Fig. S1 for the analysis flow chart. All CEL files were pre-processed and standardized with mean of zero and SD of 1. We used R/Bioconductor software for the analysis. Differential gene expression analysis was controlled for FDR < 0.05<sup>39</sup>, and set with threshold of fold changes of TSLCs vs. PTCs, to generate geneset A. Coefficient of variance<sup>40</sup> was calculated to rank the top 500 probes of lowest transcriptional variability in TSLCs. We combined geneset A and low variability gene signatures (as geneset B) and further excluded those gene signatures with inconsistent activities of TSLCs comparing to PTCs (as geneset C). Among gene signatures with concordant gene activities in at least two tumor types/experimental conditions, we identified the first list of 64 probes, either differentially expressed in at least two tumor types/experimental conditions or from the low variability genes in TSLCs, denoted as a consensus gene-list characterized with low variation and commonality (*lv\_com*). From geneset C, we found the second list of 145 probes (including 14 probes from *lv\_com*) with consistent gene activities in lung TSLCs comparing to those of lung PTCs.

**Network construction from literature knowledge base, human protein-protein interactions (PPIs), and co-expression profiles of cultivated TSLCs.** Literature networks (svg files) using *lv\_com* and the lung-TSLCs concordant gene signatures as inputs in the IPA were extracted, parsed, and compiled<sup>41</sup>. Please visit the supplementary website for the Perl script and R code. Correlated output genes generated from the IPA as well as the input genes were further mapped onto the human PPIs downloaded from the NCBI (HPRD, BioGrid, and BIND). PPIs would be retrieved if and only if both of the reactants were queried. Then, IPA generated networks were merged with the mapped human PPIs. To consolidate the network models, we calculated the co-expression Pearson correlation coefficients (PCCs) of every gene-gene interactions in the merged networks using all TSLCs and the lung TSLC only. Absolute values of PCCs of co-expression were calculated using all TSLCs and 0.4 was set as the cut-off threshold for the consensus TSLC networks. For the lung-TSLC networks, we set the cut-off threshold of abs(PCCs in lung TSLCs) > 0.8. The thresholds were determined such that the number of nodes in the final networks would be less than 100. Functional annotation clustering of genes in the TSLC-consensus networks was analyzed by DAVID (Database for Annotation Visualization and Integrated Discovery, NIH)<sup>42</sup>.

**Network topological analysis and predictive measurements derived from signal processing mechanics.** Network topological analyses and classification of genes were performed according to methods previously published<sup>43</sup>. We developed five measurements to describe the network signal processing mechanics: expression level (*Exprs*); topologically weighted expression level (*wt.Expr*s); the 0-order magnitude (*Mag*), i.e. amplitude of the transcriptional signal; the 1-order property spectrum (*Spec*), i.e. the pair-wise relative transcriptional noise; and the signal-to-noise ratio (SNR). In the network model, there would be  $N_i$  weighting genes as well as  $N_j$  member genes of scalable sizes. For each gene,  $g$ , we assigned a measure of topological property,  $w_t g$ . Importantly,  $w_t g$  would be different according to the topological grouping: that is, zero for the periphery genes; degrees (number of nodes connected) for the intra-modular hubs; and either degrees or the estimated effects of perturbation (focality)<sup>41</sup> for the inter-modular hubs. Then, for a single weighting gene,  $g$ , in the model with  $N_j$  member genes, the expression value (*Exprs*) was  $\theta_g$ ; *wt.Expr*s was the value of  $w_t g * \theta_g$ ; *Mag* would be  $m(g) = w_t g * \text{abs}(\theta_g)$ ; *Spec* would be calculated as

$$s(g) = \frac{1}{2(N_j - 1)} \left( \sum_{\substack{g=1 \\ g \neq j}}^{N_i} w_t g * \text{abs}(\theta_g) * \ln \frac{\text{abs}(\theta_g)}{\text{abs}(\theta_j)} + \sum_{\substack{j=1 \\ j \neq g}}^{N_i} w_t j * \text{abs}(\theta_j) * \ln \frac{\text{abs}(\theta_j)}{\text{abs}(\theta_g)} \right); \text{ and}$$

SNR would be defined as  $x(g) = m(g)/s(g)$ . For a group of  $N_i$  weighting genes, *Exprs*, *wt.Expr*s, and *Mag* would be the averaged values. *Spec* would be calculated as

$$s(g) = \frac{1}{2N_i(N_j - 1)} \left( \sum_{\substack{g=1 \\ g \neq j}}^{N_i} \sum_{j=1}^{N_i} w_t g * \text{abs}(\theta_g) * \ln \frac{\text{abs}(\theta_g)}{\text{abs}(\theta_j)} + \sum_{\substack{j=1 \\ j \neq g}}^{N_i} \sum_{g=1}^{N_i} w_t j * \text{abs}(\theta_j) * \ln \frac{\text{abs}(\theta_j)}{\text{abs}(\theta_g)} \right)$$

and SNR unchanged.

**Statistical and survival analyses.** Student t test and bootstrap Kolmogorov-Smirnov test were used to determine the statistical significance of means or distributions. Kaplan-Meier survival curves based on quartiles of network-based predictive measurements were tested by log-rank tests. We also fitted Cox proportional hazard regression model and used Wald test statistics to determine a trend of gene dosage effects. Statistical significance was set at  $P < 0.05$ . Please visit the supplementary website for R codes used in the survival analysis.

**Clonogenic assay.** For a clonogenic assay, cells were exposed to different chemotherapeutic agents (cisplatin, doxorubicin, and taxol)(10  $\mu\text{g}/\text{ml}$ ). After incubation for 10 days, colonies (>50 cells per colony) were fixed and stained for 20 min with a solution containing crystal violet and methanol. Cell survival was determined by a colony formation assay. The plating efficiency (PE) and survival fraction (SF) were calculated as follows: PE = (colony number/number of inoculated cells)  $\times$  100%. SF = colonies counted/(cells seeded  $\times$  (PE/100)).

**Western blot assay.** Fifteen microliters of sample were boiled at 95 °C for 5 min and separated by 10% SDS-PAGE. The proteins were transferred to Hybond-ECL nitrocellulose paper (Amersham) by a wet-transfer system. The primary antibodies used was antibody rabbit anti-human CBX5 (Cell Signaling Technology). The reactive protein bands were detected by the ECL detection system (Amersham).

**In vitro cell invasion analysis and soft agar assay.** The 24-well plate Transwell® system with a polycarbonate filter membrane was used (8  $\mu\text{m}$  pore size; Corning, United Kingdom). Cell suspensions were seeded in the upper compartment of the Transwell chamber at a density of  $1 \times 10^5$  cells in 100  $\mu\text{L}$  of serum-free medium. The opposite surface of the filter membrane facing the lower chamber was stained with Hoechst33342 for 3 min, and migrating cells were visualized under an inverted microscope. For the soft agar assay, the bottom of each well (35 mm) of a six-well culture dish was coated with 2 mL of an agar mixture (DMEM, 10%(v/v) FCS, 0.6% (w/v) agar). After the bottom layer solidified, 2 mL of a top agar-medium mixture (DMEM, 10%(v/v) FCS, 0.3%(w/v) agar) containing  $2 \times 10^4$  cells was added and incubated at 37 °C for 4 weeks. The plates were stained with crystal violet. The number of colonies was counted using a dissecting microscope.

**Quantitative real-time RT-PCR and patients and tissue samples.** Lung adenocarcinomas (LACs) and adjacent noncancerous tissues were obtained at the time of surgery from 20 patients in Taipei Veterans General Hospital. All patients gave their informed consent under institution's approval. Total RNA was extracted from tissue samples using TRIzol according to the manufacturer's protocol (Invitrogen, Carlsbad, CA). The amplification and PCR reaction were carried out (Roche, Alameda, CA). Standard curves (cycle threshold values versus template concentration) were prepared for each target gene and for the endogenous reference (GAPDH) in each sample.

**Patient subjects and immunohistochemistry (IHC).** Between 1996 and 2009, 125 patients with operable LAC, without histories of radiation or chemotherapy, underwent surgery at Taipei Veterans General Hospital (Table S6). All samples were obtained after informed consent according to the tenets of the Declaration of Helsinki. Tissue samples were spotted on glass slides for IHC staining, deparaffinised, rehydrated, processed with antigen retrieval by 1X Trilogy diluted in H<sub>2</sub>O (Biogenics), immersed in 3% H<sub>2</sub>O<sub>2</sub> for 10 min, and washed with PBS 3 times. The tissue sections were then blocked with serum (Vectastain Elite ABC kit, Vector Laboratories, Burlingame, CA) for 30 min, followed by incubating with the primary antibody rabbit anti-human CBX5 (Cell Signaling Technology) in PBS solution at room temperature for 2 hr, washed with PBS 3 times, incubated with biotin-labeled secondary antibody for 30 min, incubated with streptavidin-horse radish peroxidase conjugates for 30 min, washed with PBS 3 times, and immersed with chromogen 3-3'-diaminobenzidine plus H<sub>2</sub>O<sub>2</sub> substrate solution (Vector® DBA/Ni substrate kit, SK-4100, Vector Laboratories, Burlingame, CA) for 10 min. Hematoxylin was applied for counter-staining (Sigma Chemical Co., USA). Study pathologists, blinded to the clinical data, examined and scored the IHC staining. The interpretation was done in five high-power views for each slide, and 100 cells per view were counted for analysis.

**Xenograft tumorigenicity assay.** All procedures involving animals were in accordance with the institutional animal welfare guideline and the experiment was approved by Taipei Veterans General Hospital. Virus-infected lung TSLCs were harvested, washed with PBS, and re-suspended in normal culture medium. Lung TSLC cells ( $2 \times 10^5$ ) infected with sh-CBX5 RNAi or control vector were injected through tail vein of 8-week-old male NOD-SCID mice. All mice were anesthetized and killed on day 56 (8 weeks) after injection. The number of tumor nodules and tumor volume in lung of the transplanted mice were measured by *ex vivo* and H&E survey. Ionizing radiation (IR) was delivered by a cobalt unit (Theatron International, Inc., Ottawa, Canada) at a dose rate of 1.1 Gy/min (source-to-surface distance = 57.5 cm). Lung CD133<sup>+</sup>-TSLCs treated with sh-CBX5 RNAi were exposed to the radiation doses of 2, 4, 6, 8, and 10 Gy.

1. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**, 499–507 (2008).





2. Chiou, S. H. *et al.* Positive correlations of Oct-4 and Nanog in oral cancer stem-like cells and high-grade oral squamous cell carcinoma. *Clin Cancer Res* **14**, 4085–4095 (2008).
3. Yang, M. H. *et al.* Direct regulation of TWIST by HIF-1 $\alpha$  promotes metastasis. *Nat Cell Biol* **10**, 295–305 (2008).
4. Polyak, K. & Weinberg, R. A. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* **9**, 265–273 (2009).
5. Lo, J. F. *et al.* The Epithelial-Mesenchymal Transition Mediator S100A4 Maintains Cancer Initiating Cells in Head and Neck Cancers. *Cancer Res* **71**, 1912–1923 (2011).
6. Chen, Y. C. *et al.* Aldehyde dehydrogenase 1 is a putative marker for cancer stem cells in head and neck squamous cancer. *Biochem Biophys Res Commun* **385**, 307–313 (2009).
7. Chiou, S. H. *et al.* Identification of CD133-positive radioresistant cells in atypical teratoid/rhabdoid tumor. *PLoS One* **3**, e2090 (2008).
8. Lien, H. C. *et al.* Molecular signatures of metaplastic carcinoma of the breast by large-scale transcriptional profiling: identification of genes potentially related to epithelial-mesenchymal transition. *Oncogene* **26**, 7859–7871 (2007).
9. Chen, Y. C. *et al.* Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells. *PLoS One* **3**, e2637 (2008).
10. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J Clin* **61**, 69–90.
11. Lam, W. K. & Watkins, D. N. Lung cancer: future directions. *Respirology* **12**, 471–477 (2007).
12. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954.
13. Zaidi, S. K. *et al.* Nuclear microenvironments in biological control and cancer. *Nat Rev Cancer* **7**, 454–463 (2007).
14. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173.
15. Macarthur, B. D., Ma'ayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* **10**, 672–681 (2009).
16. Kalmar, T. *et al.* Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* **7**, e1000149 (2009).
17. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* **10**, 186–198 (2009).
18. Bertolini, G. *et al.* Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. *Proc Natl Acad Sci U S A* **106**, 16281–16286 (2009).
19. Salnikow, A. V. *et al.* CD133 is indicative for a resistance phenotype but does not represent a prognostic marker for survival of non-small cell lung cancer patients. *Int J Cancer* **126**, 950–958 (2010).
20. Maison, C. & Almouzni, G. HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* **5**, 296–304 (2004).
21. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**, 1025–1040 (2007).
22. Dinant, C. & Luijsterburg, M. S. The emerging role of HP1 in the DNA damage response. *Mol Cell Biol* **29**, 6335–6340 (2009).
23. Wong, L. H. *et al.* ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res* **20**, 351–360.
24. Somervaille, T. C. *et al.* Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell* **4**, 129–140 (2009).
25. Broske, A. M. *et al.* DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet* **41**, 1207–1215 (2009).
26. Tadokoro, Y., Ema, H., Okano, M., Li, E. & Nakauchi, H. De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *J Exp Med* **204**, 715–722 (2007).
27. Taylor, I. W. *et al.* Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol* **27**, 199–204 (2009).
28. Bredel, M. *et al.* A network model of a cooperative genetic landscape in brain tumors. *Jama* **302**, 261–275 (2009).
29. Shedden, K. *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822–827 (2008).
30. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).
31. Raponi, M. *et al.* Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* **66**, 7466–7472 (2006).
32. Nguyen, D. X. *et al.* WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. *Cell* **138**, 51–62 (2009).
33. Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R. & Geiger, B. Functional atlas of the integrin adhesome. *Nat Cell Biol* **9**, 858–867 (2007).
34. Simpson, K. J. *et al.* Identification of genes that regulate epithelial cell migration using a siRNA screening approach. *Nat Cell Biol* **10**, 1027–1038 (2008).
35. Muller, F. J. *et al.* Regulatory networks define phenotypic classes of human stem cell lines. *Nature* **455**, 401–405 (2008).
36. Monteith, G. R., McAndrew, D., Faddy, H. M. & Roberts-Thomson, S. J. Calcium and cancer: targeting Ca<sup>2+</sup> transport. *Nat Rev Cancer* **7**, 519–530 (2007).
37. He, C., Zhou, F., Zuo, Z., Cheng, H. & Zhou, R. A global view of cancer-specific transcript variants by subtractive transcriptome-wide analysis. *PLoS One* **4**, e4732 (2009).
38. Network, T. C. G. A. R. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
39. Strimmer, K. A unified approach to false discovery rate estimation. *BMC bioinformatics* **9**, 303 (2008).
40. Yu, K. *et al.* A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. *PLoS Genet* **4**, e1000129 (2008).
41. Yu, Y. H., Kuo, H. K. & Chang, K. W. The evolving transcriptome of head and neck squamous cell carcinoma: a systematic review. *PLoS One* **3**, e3215 (2008).
42. Dennis, G., Jr. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4**, P3 (2003).

## Acknowledgements

Dr. Yau-Hua Yu would like to express the appreciation to her thesis advisor, Dr. Marco Ramoni, who unexpectedly passed away in 2010, for the guidance which lead to the initial framework of such network-topologically-based signal processing survival prediction model. The authors would like to thank Dr. Yeu Su for the consultation in study design as well as to thank Dr. Hsu-Ko Kuo for the help with IPA analysis. The work was done during the period when he was still at the Dept. of Geriatrics and Gerontology, National Taiwan University Hospital, Taiwan.

This study was supported by research grants from the National Science Council (NSC 100-2325-B-010-010-MY3/98-2314-B-010-024-MY2/97-3111-B075-001-MY3/96-2314-075-056-MY3); National Yang-Ming University (Ministry of Education, Aim for the Top University Plan: 96ADD122, 96ADD125, 96ADT191, 97ACD113, 97ACT302, 98ACT302, 98ACD107, and 98ACT192; and Brain Research Center-3T-MRI project); Taipei Veterans General Hospital (98-C1-099/E1-003/ER3-001); the Joint Projects of VGHUST (98-G6-6/98-P1-01/99-P6-39); Chi-Mei Medical Center (CMYM9801); Yen-Tjing-Ling Medical Foundation (96/97/98); Taipei City Hospital (96-002-62-092); and Technology Development Program for Academia (TDPA: 98-EC-17-A-19-S2-0107), Department of Industrial Technology, Ministry of Economic Affairs; all in Taiwan. NSC 101-2325-B-010-009; The Department of Health Cancer Center Research of Excellence (DOH101-TD-C-111-007).

## Author contributions

Design of study: Yu YH and Chiou SH; Experiments: Chiou GY, Lo WL, Yu CC, Huang PI, Lo JF, and Liu KH; Data analysis: Yu YH, Chiou GY, Liu KH, and Chiou SH; Clinical collaborators: Wang CY, Huang WC, and Hsu HS; Manuscript write-up: Yu YH, Alterovitz G, and Chiou SH.

## Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

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

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

**How to cite this article:** Yu, Y. *et al.* Network Biology of Tumor Stem-like Cells Identified a Regulatory Role of CBX5 in Lung Cancer. *Sci. Rep.* **2**, 584; DOI: 10.1038/srep00584 (2012).