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

Yau-Hua Yu1,14, Guang-Yuh Chiou2,5, Pin-I Huang2,5, Wen-Liang Lo3,4, Chien-Ying Wang2,6, Kai-Hsi Lu7, Cheng-Chia Yu1,8, Gil Alterovitz9,10, Wen-Chien Huang11, Jeng-Fan Lo1,3, Han-Shui Hsu12,13 & Shih-Hwa Chiou1,2,5

1Institute of Oral Biology, National Yang-Ming University, Taipei, Taiwan, 2Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan, 3Faculty of Dentistry, National Yang-Ming University, Taipei, Taiwan, 4Department of Stomatology, Taipei Veterans General Hospital, Taipei, Taiwan, 5Department of Medical Research and Education, Taipei Veterans General Hospital, Taipei, Taiwan, 6Department of Emergency, Taipei Veterans General Hospital, Taipei, Taiwan, 7Department of Medical Research and Education, Cheng Hsin General Hospital, Taipei, Taiwan, 8Institute of Oral Biology and Biomaterial Science, Chung-Shan Medical School, Taichung, Taiwan, 9Children’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, 10Partners HealthCare Center for Personalized Genetic Medicine, Boston, United States, 11Division of Thoracic Surgery, Department of Surgery, Mackay Memorial Hospital, Taipei, Taiwan, 12Division of Thoracic Surgery and Department of Surgery, Taipei Veterans General Hospital, Taipei, Taiwan, 13Institute of Emergency and Critical Care Medicine, National Yang-Ming University, Taipei, Taiwan, 14Harvard 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, ESRI, 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 prognosis1-3. Other studies showed that overexpression of epithelial-mesenchymal transition transcription factors enhanced stem-like properties and increased the aggressiveness of tumor cells4-12. We established several panels of tumor stem-like cells (TSLCs) in head and neck2,6, brain3,6, brain7, and breast4. Here, we focus on lung adenocarcinomas (LACs), for which we have established a panel of lung TSLCs previously as well13. Indeed, lung cancer is one of the leading causes of cancer-related deaths worldwide14. Its highly invasive and metastatic phenotypes are the major reasons for treatment failure and poor prognosis15.

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 reprogramming16. Nuclear perturbation was also known to play an important role in cancer biology13,17. 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,
CD133 expression of CD133 in lung cancers represents high tumorigenicity.

Characterization of 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\(^\text{14-16}\). Researchers also tried integration of electrical potentials in neurons and the brain functional magnetic resonance images\(^\text{17}\) 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/nwtoposignalcancer/.

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 sound- 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\(^+\) TSLCs.

Characterization of lung TSLCs

Recent studies showed that expression of CD133 in lung cancers represents high tumorigenicity and resistance to cytotoxic therapy\(^\text{18}\). We previously reported greater chemoradioresistance of CD133\(^+\) TSLCs isolated from non-small cell lung cancers (NSCLCs) compared to CD133\(^+\) NSCLCs\(^2\). Here, we isolated CD133\(^+\) TSLCs from 7 NSCLCs (Fig. 1a). Isolated lung CD133\(^+\) TSLCs could form floating spheroid-like bodies in serum-free medium more easily than CD133\(^+\) 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\(^+\) TSLCs (Fig. 1b). Lung CD133\(^+\) 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\(^+\) TSLCs exhibited more aggressiveness tumorigenity 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 (\(lv\_com\)). 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 \(lv\_com\) 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 \(lv\_com\) were displayed in Fig S2. By using \(lv\_com\) 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&S3). 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 peripheral 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&S5). 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, wtExprs, Mag, Spec, and SNR derived from each combinatorial 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, wtExprs, Mag, 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
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.

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+TSLCs treated by sh-CBX5 RNAi were indeed significantly inhibited (Fig. 3b–d). Additionally, the percentages of CD133+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
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 Exps 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 network topology, as well as signal stochasticity. We further tried to experimentally validate the identified correlated synchronization network topology, as well as signal stochasticity. The correlations between the Spec or SNR levels were higher than those of the Exps 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.

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 2×10⁵ sh-CBX5 RNAi treated lung CD133⁺-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⁺-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⁺-TSLCs, we further tested to show that treatment of sh-CBX5 RNAi significantly increased the radiosensitivity of CD133⁺-TSLCs in vitro as well (Fig. 4d). Mice

| NSCLC Case | Age/Sex/Type/Stage | CD133⁺ (%) | Spheres Formation | CD133⁺ sh-RNA vector | CD133⁺ sh-CBX5 | CD133⁺ IR-4 Gy | CD133⁺ & IR 4 Gy | CD133⁺ |
|------------|--------------------|------------|-------------------|---------------------|----------------|----------------|----------------|---------|
| 1          | 82/M/AD/Illa       | 7.1        | Yes               | 1,000 (3/3)         | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3) |
| 2          | 59/F/AD/Ilb        | 4.8        | Yes               | 1,000 (0/3)         | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3) |
| 3          | 63/F/AD/Ilb        | 3.6        | Yes               | 1,000 (0/3)         | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3) |
| 4          | 70/M/SC/Ib         | 11.2       | Yes               | 1,000 (0/3)         | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3) |
| 5          | 75/M/AD/Ila        | 3.9        | Yes               | 1,000 (0/3)         | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3) |
| 6          | 67/M/AD/Ilb        | 15.5       | Yes               | 1,000 (0/3)         | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3) |
| 7          | 68/M/SC/Ib         | 2.3        | Yes               | 1,000 (0/3)         | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3)    | 1,000 (0/3) |

NSCLC: non-small cell lung carcinoma. Tumor types: AD: lung adenocarcinoma; SC: lung squamous cell carcinoma. Age/Sex/Type/Stage: MN staging. Cells were transplanted into NOD-SCID mice through the tail vein. After 8 weeks of transplantation, the tumorigenicity of tumor-bearing NOD-SCID mice was measured by histological survey in the whole lung.

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—macromolecular metabolic process | 83% | 0.02 |
|               | BP          | GO:0048731—system development | 42% | 0.03 |
|               | BP          | GO:0006466—protein amino acid phosphorylation | 33% | 0.01 |
| Intra-modular hubs | CC          | GO:0043231—intracellular membrane-bound organelle | 100% | 1.3E-07 |
|               | BP          | GO:0043170—macromolecular 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 |

www.nature.com/scientificreports
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 ± 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.
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\(^{+}\)-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\(^{+}\)-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. In rodent and *D. melanogaster* cells, CBX5 was found to interact with H3K9me3 or colocalized with H3K9me to the heterochromatin regions. 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 3 | Knockdown of CBX5 with sh-RNAi in vitro.** (a) Western blot of knockdown of CBX5 in lung CD133\(^{+}\)-TSLCs derived from two patients (No.1&2). The abilities of (b) sphere formation, (c) colony formation, and (d) migration in CD133\(^{+}\)-TSLCs treated with sh-CBX5 RNAi were decreased. The percentages of (e–f) CD133\(^{+}\)-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*, *TAES*, *ESR1*, *MLH1*, and *SIN3A* in the sh-CBX5 RNAi treated lung CD133\(^{+}\)-TSLCs. These genes were identified by statistically significant correlation with CBX5 in lung cancer survival analysis. All data shown are the mean ± SD of 3 experiments. (*P < 0.05*)
expression signature shared by human mature oocytes and embryonic stem cells. Recently, Wong and colleagues identified that ATRX working together with H3.3 and CBX5 might be a key regulator of ES-cell telomere chromatin. Here, CBX5 was identified in both the consensus TSLC and the lung-TSLC network models. Our findings further supported a recent finding 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 Bröskie and colleagues demonstrated the indispensability of DNMT1 for the cell-autonomous survival of hematopoietic stem cells (HSCs) and LSCs. De novo methylation by DNMT3A and DNMT3B was also shown essential for HSCs renewal but not for differentiation. 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 and glioblastoma. 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, RT, BT6, and BT12, respectively. TSLCs were cultivated by methods described elsewhere. For lung TSLCs, we isolated CD133+ U87, ATRT, BT, BT6, and BT12, respectively. TSLCs were cultivated by methods previously published. We developed five interactions (PPIs), and co-expression profiles of cultivated TSLCs were the starting point for the construction of the model with a polycarbonate filter membrane was used (8 μm pore size; Corning, USA). Cell suspensions were seeded in the upper compartment of the Transwell chamber at a density of 1 × 10^5 cells in 100 μL 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) was coated with 2% (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 2×10^5 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.

**Network construction from literature knowledge base, human protein-protein interactions (PPIs), and co-expression profiles of cultivated TSLCs.** Literature networks (svg files) using IV_com and the lung-TSLCs concordant gene signatures as inputs in the IPA were extracted, parsed, and compiled. Please visit the supplementary website for the Perl script and R code. Correlated output genes generated from the IPA and the TSLC PPIs 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 human PPIs. To consolidate the network models, we calculated the co-expression Pearson correlation coefficients (PCCs) of every gene-gene interaction 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 10,000. Functional annotation clustering of genes in the TSLC consensus networks was analyzed by DAVID (Database for Annotation Visualization and Integrated Discovery, NIH).

**Network topological analysis and predictive measurements derived from signal processing mechanisms.** Network topological analyses and classification of genes were performed according to methods previously published. We developed five measurements to describe the network signal processing mechanisms: expression level (Exp); topologically weighted expression level (wt.Exprs); 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 weighting genes as well as N member genes of each available size. For each gene, g, we determined the measure of transcriptional activity (weight for the property, wt). Importantly, wt would be different according to the topological grouping: that is, zero for the periphery genes; degrees (number of nodes connected) for the intra-modal hubs; and either degrees or the estimated effects of perturbation (f(Toy)) for the inter-modal hubs. Then, for a single weighting gene, g, in the model with N member genes, the expression value of g was (Exp). wt.Exprs was the value of wt*Exp. Mag would be (Mag) = wt * abs(Exp); Spec would be calculated as: 

\[ S(g) = \frac{1}{2(N-1)} \left( \sum_{j \neq g} w_{ij} * abs(Exp) + \sum w_{ij} * abs(Exp) - \sum w_{ij} * abs(Exp) + \sum w_{ij} * abs(Exp) \right) \] 

SNR would be defined as \( S(g) = M(g)/S(g) \). For a group of N weighting genes, Exp, wt.Exprs, and Mag would be the averaged values. Spec would be calculated as: 

\[ S(g) = \frac{1}{2(N-1)} \sum_{j \neq g} w_{ij} * abs(Exp) + \sum w_{ij} * abs(Exp) - \sum w_{ij} * abs(Exp) + \sum w_{ij} * abs(Exp) \] 

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 see 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 μg/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 colony formation assay. The plating efficiency (PE) and survival fraction (SF) were calculated as follows: PE = ( colony number/number of inoculated cells ) × 100%. SF = colonies counted/(cells seeded x (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 membrane (Amersham) with a wet-transfer system. The primary antibodies used was antibody rabbit anti-human CBX5 (Cell Signaling Technology). The reaction protein bands were detected by the ECL detection system (Amershams).
3. Yang, M. H.

2. Chiou, S. H.

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 SI00A4 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 metastatic 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 (2001).

11. Lam, W. K. & Watkins, D. N. Lung cancer: future directions. *Nat Rev Cancer* **7**, 519–530 (2007).

12. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).

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 (2010).

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**, e1001149 (2009).

17. Bullmore, E. & Sporns, O. Complex networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* **10**, 186–198 (2009).

18. Bertolino, G. et al. Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by clastatin treatment. *Proc Natl Acad Sci U S A* **106**, 16281–16286 (2009).

19. Salnikov, 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. & Almenouzi, 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 leukaemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell* **4**, 129–140 (2009).

25. Brouske, A. M. et al. DNA methyltransferase 1 protects hematopoietic stem cell multipotency from myeloid-restriction deprivation. *Nat Genet* **41**, 1207–1215 (2009).

26. Tadokoro, Y., Ema, H., Okano, M., Li, E. & Nakajichi, 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 adeno-carcinoma: 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., Izkovitz, 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 an 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. Montefith, G. R., McAndrew, D., Faddy, H. M. & Roberts-Thomson, S. J. Calcium and cancer: targeting Ca2+ 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**, e2315 (2008).

42. Dennis, G., Jr. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* **4**, P3 (2003).