Rewiring of human lung cell lineage and mitotic networks in lung adenocarcinomas

Il-Jin Kim1,2, David Quigley2, Minh D. To1,2, Patrick Pham1, Kevin Lin2, Brian Jo2, Kuang-Yu Jen2, Dan Raz1, Jae Kim1, Jian-Hua Mao3, David Jablons1,2 & Allan Balmain2,4

Analysis of gene expression patterns in normal tissues and their perturbations in tumours can help to identify the functional roles of oncogenes or tumour suppressors and identify potential new therapeutic targets. Here, gene expression correlation networks were derived from 92 normal human lung samples and patient-matched adenocarcinomas. The networks from normal lung show that NKX2-1 is linked to the alveolar type 2 lineage, and identify PEBP4 as a novel marker expressed in alveolar type 2 cells. Differential correlation analysis shows that the NKX2-1 network in tumours includes pathways associated with glutamate metabolism, and identifies Vaccinia-related kinase (VRK1) as a potential drug target in a tumour-specific mitotic network. We show that VRK1 inhibition cooperates with inhibition of poly (ADP-ribose) polymerase signalling to inhibit growth of lung tumour cells. Targeting of genes that are recruited into tumour mitotic networks may provide a wider therapeutic window than that seen by inhibition of known mitotic genes.
lung cancer remains the leading cause of cancer death in women and men in the United States, taking more lives each year than breast, prostate and colorectal cancers combined. Despite major advances in our knowledge of lung cancer genetics, the 5-year survival rate for non-small cell lung cancer (NSCLC) patients is around 16% (ref. 2). Lung cancer treatment is therefore moving rapidly towards an era of personalized medicine, where the molecular characteristics of a patient’s tumour will dictate the optimal treatment modalities. NSCLC patients with EGFR mutations show significantly improved responses to treatment with tyrosine kinase inhibitors, for example, gefitinib or erlotinib that target this receptor kinase. However, almost all of these patients eventually relapse due to rewiring of the genetic architecture of the tumours, either through secondary EGFR mutations, or other unknown genetic alterations. This paradigm of tumour rewiring in the face of targeted treatment is evident from many clinical trials that have been reported or are in progress, suggesting that novel approaches to the identification of drug targets and/or methods for analysis of the rewiring process to find complementary targets will be necessary to improve survival and the probability of long-term cures.

We have previously shown that genetic analysis of gene expression can provide a network view of the molecular architecture of normal mouse skin, and of the changes in this architecture associated with transformation to carcinoma. This approach allowed us to identify networks that control normal skin cellular components, such as the hair follicles, muscle, and inflammatory cells, as well as the changes in these networks that accompany development of carcinomas. We have applied the same strategy to analysis of normal human lung, and matched lung adenocarcinomas from the same patients. The networks generated from a panel of 92 matched normal-tumour lung samples contain motifs representing major cell structural and functional components of the lung, as well as changes in mitotic and inflammation-associated networks in carcinomas. The availability of matched adenocarcinoma tissue from each patient enabled us to generate a coordinated network of genes that are up- or downregulated in tumours compared with the normal tissue from the same individuals. This ‘progression network’ revealed important contributions from stromal and infiltrating inflammatory cells due to influx of host cells into the tumour microenvironment, but also revealed tumour-specific changes in the cell cycle and mitotic networks that were exploited to identify VRK1 (Vaccinia-related kinase 1) as a mitotic target for the tumour cell cycle.

**Results**

Gene expression networks in normal lung. Sets of genes that exhibit correlated expression patterns in normal tissue from a genetically heterogeneous population often share a common function or are part of the same physical structure. We first carried out gene expression network analysis of 92 samples of normal unaffected human lung samples from lung cancer patients. Although previous studies of normal mouse skin clearly identified both structural and functional components of this tissue, application of the same approach to normal human lung identified a large number of significant correlations that were not readily resolved into similar motifs (Supplementary Fig. S1).

Adenocarcinomas share many characteristics with cells of the normal alveolar type 2 lineage, for which the best characterized markers are the genes encoding the surfactant proteins such as SFTPBP. We therefore constructed a targeted gene expression network for SFTA2 and SFTP8 and the transcription factor NKX2-1 from the normal lung samples, and noted that several surfactant proteins (SFTPA1B, SFTPA2B, SFTA3) and other alveolar type 2 markers (for example, LAMP3) are represented in this correlation network (Fig. 1a).

NKX2-1 has previously been reported to be a lineage-specific oncogene that is amplified in a subset of NSCLC, in particular in adenocarcinomas. Others have demonstrated by immunohistochemistry that the NKX2-1 protein is also mainly expressed in alveolar type 2 cells. A functional association of NKX2-1 with surfactant protein genes is supported by the observation that mutation of NKX2-1 results in interstitial lung disease in Brain-Lung-Thyroid Syndrome by surfactant protein promoter dysregulation. Abnormal function of NKX-2-1 also affects respiratory distress syndrome of the newborn. Taken together, these data support a role for NKX2-1 in type 2 pneumocyte lineage determination both in normal lung and in lung adenocarcinomas.

While many of the genes in these networks are known markers of their respective lineages, this approach has the potential to identify completely novel markers for specific cell types within the lung and other tissues. We validated the network approach for identification of markers for alveolar type 2 cells using antibodies to PEBP4, a RAF kinase inhibitory protein family member and a scaffold protein controlling myoblast and muscle cell differentiation through MEK and ERK signalling. This protein has not previously been associated with alveolar type 2 cells, but was significantly coexpressed with surfactant protein genes in normal human lungs. Immunohistochemical staining was carried out on normal human lung sections using antibodies specific for PEBP4 and two other type 2 pneumocyte markers, SFTP8 and SFTPD. The results shown in Fig. 1b demonstrate that PEBP4 protein is localized in the same cell population as the known type 2 markers, thus providing support for the predictions made by the gene expression networks.

Rewiring of gene expression networks in lung tumours. Correlation analysis of lung tumours from the same patients revealed several major motifs representing cell growth and mitosis, inflammatory responses and stromal cells (Fig. 2a). In contrast to the normal lung network, the tumour gene expression correlations identified clear modules representing structural (Clara cells) and functional (inflammation, mitosis, translation) components in lung adenocarcinomas. We then produced a molecular view of the perturbations in gene expression between these normal lung tissues and matched adenocarcinomas by combining differential expression and correlation to identify functionally related sets of genes that are coordinately modified during tumorigenesis. This ‘progression network’ (Fig. 2b) was produced by subtracting gene expression results in patient’s normal samples from the corresponding adenocarcinoma samples to produce a fold-change value for each gene (see Methods). We then identified gene pairs, which were significantly correlated in their fold-change and significantly up- or downregulated.

The progression network identified numerous structural and functional features of the normal lung that were altered either positively or negatively in matched lung tumours from the same patients. Motifs significantly enriched for genes with roles in mitosis, translation, tumour–stromal interactions and immune responses were significantly upregulated in matched tumours compared with normal tissue. The collagen and stromal gene motif contained genes, such as Fibroblast activation protein (FAP) and fibronectin type III domain containing 1 (FNDC1), which are associated with epithelial tumour growth. Motifs associated with negative regulation of cell proliferation, cell adhesion and focal adhesion assembly were coordinately downregulated. Consistent with the morphological changes that occur during...
tumorigenesis, the networks corresponding to the alveolar type 2 cells (surfactant protein genes) and Clara cells (secretoglobins CC-10 or SCG1A1) formed distinct cliques that were significantly downregulated in cancers.

**Type 2 pneumocyte network rewiring in lung adenocarcinomas.** NKX2-1 is significantly correlated with surfactant protein gene expression in tumours (Fig. 3), supporting a model whereby NKX2-1 maintains a role in determination of the alveolar type 2 cell lineage. However, the NKX2-1 tumour correlation network provides evidence for significant rewiring of the NKX2-1 gene network associated with transformation. Although several components of the normal NKX2-1 network including surfactant protein genes were also significantly correlated with NKX2-1 in tumours, the tumour network included many genes significantly correlated in tumours that showed minimal association with NKX2-1 expression levels in normal lungs (see Methods and Table 1). Five of the 23 genes in this set are involved in the glutamine biosynthetic process (glutaminase 2: GLS2; gamma-glutamyl carboxylase: GGCX; gamma-glutamyltransferase 1: GGT1; gamma-glutamyltransferase 2: GGT2; Gamma-glutamyltransferase light chain 1: GGTLC1). This gene set is significantly enriched for genes with roles in glutathione metabolism and glutamate secretion (corrected p-value < 2.4 x 10^{-3}, hypergeometric test). Glutamine metabolism is increased in cancer cells and has been linked to the Warburg effect associated with Myc deregulation^{18}. These data indicate that the NKX2-1 transcription factor pathway is substantially rewired in lung adenocarcinoma cells, and is now associated with pathways that are required for rapid cell growth and the metabolic requirements of cancer cells. The availability of small-molecule inhibitors of GLS2 (ref. 19) may provide a possible therapeutic approach to exploitation of this rewired network for treatment of adenocarcinomas that overexpress NKX2-1.

Several additional genes not associated with normal type 2 pneumocytes were also linked to NKX2-1 in this network (Fig. 3). One of these genes, CLPTM1L (CLPTM1-like) has been identified from genome-wide association studies as a candidate susceptibility gene for lung cancer^{20–22}. CLPTM1L is physically located on chromosome 5p15 in close proximity to TERT, which is another strong candidate susceptibility gene in the region. Although the known functions of TERT in control of telomere length, as well as its other functions in growth control^{23,24}, make it a prime candidate, the strong coregulation of CLPTM1L with a gene known to have an important role in adenocarcinoma lineage commitment suggests that it may also be involved in lung cancer susceptibility.

**Figure 1 | Gene expression network for alveolar type II markers in normal lung.** (a) Gene coexpression networks associated with surfactant protein markers and NKX2-1 in normal human lung tissues. Gene pairs that are significantly correlated are drawn as blue boxes connected by a line; PEBP4 is shown in yellow. Other type 2 pneumocyte markers are shown in orange. (b) Adjacent sections of three normal slides from different lung cancer patients (1101N, 1116N and 1137N) were stained for expression of SFTPD, SFTPB, PEBP4 protein and a negative control (−) (microscope magnification: ×100). Type 2 pneumocytes were stained for known markers SFTPD and SFTPB as well as PEBP4, which was identified as a candidate marker for type 2 pneumocytes in gene expression networks. These slides indicate that the same cells express both surfactant proteins and PEBP4 protein.
**Figure 2 | Human lung adenocarcinoma tissue gene expression network.** (A) Significantly correlated gene pairs that were part of a three-gene clique were plotted ($r \geq +/− 0.75$). This analysis identified network motifs associated with mitosis, protein translation, stromal cells/invasion, inflammation and Clara cells. (B) Progression network of lung adenocarcinoma. Gene pairs with significantly correlated change ($r \geq +/− 0.7$) and significant change in mRNA gene expression level between normal lung and adenocarcinomas are drawn as nodes (false discovery rate < 5%, fold-change ≥ 2). Red nodes indicate increased expression and green nodes indicate decreased expression, with darker colour indicating more extreme change. Grey lines connect genes with significant direct correlated change and red lines indicate inverse correlation. a, Clara cell and collagen motifs; b, cell cycle motif; c, type II pneumocyte motif.
Inhibitors of several of these kinases are presently being investigated as drug targets for cancer therapy, including the Aurora kinases that are attractive drug targets for cancer therapy, including the Aurora-A (AURKA) and Aurora-B (AURKB) kinases.

The most strongly upregulated motif in the adenocarcinoma network in tumours shows coexpression with genes relevant to tumour susceptibility (for example, CLPTM1L and tumour metabolism (for example, GLS2, GGT2, GGTLC1, GGCX, GPD1L) that correlated at lower levels in the comparable normal lung network (coloured in pink). Grey lines connect genes with significant direct correlation and red lines indicate inverse correlation.

VRK1 as a drug target in adenocarcinoma mitotic networks. The strongly upregulated motif in the adenocarcinoma progression network was significantly enriched for genes active in the M-phase of the mitotic cycle (corrected \( P = 6.4 \times 10^{-15} \)) (Fig. 2b). This network comprises many genes that are attractive drug targets for cancer therapy, including the kinases AURKA, AURKB, CCNB2, TTK, MELK and PLK4 (refs 25–30). Inhibitors of several of these kinases are presently being tested in clinical trials, however their essential roles in normal mitosis may preclude development as cancer therapeutics because of the toxicity that would be expected owing to inhibition of dividing cells in the bone marrow, colon and other proliferative tissues. A major goal of this study was to use network analysis to identify ‘druggable’ target genes that may function selectively in tumour as compared with normal lung tissue. As the cell cycle has been a rich source of cancer drug targets, we carried out network analysis of AURKA and CCNB2, both well-known cell cycle genes, in both normal lung and tumour samples. Although a number of kinases exhibited significant correlations (\( r > 0.6 \)) in gene expression with AURKA and CCNB2 in both normal and tumour tissue, CDC2 and VRK1 showed a clear cell cycle network that was specific for lung cancers. (Supplementary Fig. S2).

We decided to focus our functional studies on one of these targets, VRK1, which is a serine/threonine protein kinase originally characterized through homology to a Vaccinia-related kinase\(^{31}\). Subsequent work determined that VRK1 phosphorylates the MDM2-binding site of TP53 (ref. 32) and is in turn downregulated by wild-type TP53\(^{33}\) but not mutant TP53\(^{34}\). Ablation of VRK1 in vitro blocks cell cycle progression at G1\(^{35}\). In normal lung tissue, VRK1 expression is significantly correlated with a wide spectrum of different genes, possibly reflecting the diverse roles of this kinase in normal tissues, but this list is not significantly enriched in genes that control the cell cycle. In contrast, in the lung adenocarcinoma network shown in Fig. 2a, VRK1 is significantly correlated with the mitotic cell cycle gene network (corrected \( P \)-value for enrichment: \( 5 \times 10^{-15} \); Fig. 4a). In contrast, no association with mitosis was found for VRK1 in normal lung gene expression networks (Supplementary Fig. S1). Although VRK1 gene expression is frequently upregulated in tumours compared with matched normal lung samples, the mean change in gene expression elevation at the RNA level was relatively modest (less than twofold) and substantially less than that seen for other known mitotic genes, such as Aurora kinases and Cyclin B2 (Fig. 4b). For this reason, VRK1 did not appear in the progression network shown in Fig. 2b, which only shows genes that show the strongest upregulation in tumours compared with matched normal tissue.

**Table 1 | Differential correlation analysis of NKX2-1.**

| Symbol | Probe | rho_normal | rho_tumor | Difference |
|--------|-------|------------|-----------|------------|
| BCAM   | ILMN_1790455 | 0.10 | 0.64 | 0.74 |
| INTS5  | ILMN_1796968 | 0.02 | 0.72 | 0.74 |
| ZNF323 | ILMN_1655748 | 0.07 | 0.64 | 0.71 |
| GGCX   | ILMN_1758232 | 0.06 | 0.72 | 0.66 |
| GLS2   | ILMN_1709771 | 0.08 | 0.73 | 0.65 |
| ZNF323 | ILMN_1692795 | 0.01 | 0.64 | 0.63 |
| GGTL1C | ILMN_1692188 | 0.07 | 0.68 | 0.61 |
| CDX6   | ILMN_1902615 | 0.11 | 0.70 | 0.59 |
| GGTL2  | ILMN_1737387 | 0.09 | 0.68 | 0.59 |
| GGTL1  | ILMN_1753340 | 0.05 | 0.65 | 0.59 |
| CD163L1| ILMN_1661905 | 0.05 | 0.64 | 0.59 |
| STXB1  | ILMN_1728747 | 0.11 | 0.69 | 0.58 |
| APLF2  | ILMN_1710482 | 0.08 | 0.65 | 0.58 |
| PRUNE  | ILMN_1789114 | 0.08 | 0.64 | 0.56 |
| CLPTM1L| ILMN_1752802 | 0.08 | 0.64 | 0.56 |
| HOXD1  | ILMN_1717381 | 0.10 | 0.65 | 0.56 |
| GPD1L  | ILMN_1694106 | 0.15 | 0.69 | 0.54 |
| EXPH5  | ILMN_1712066 | 0.18 | 0.72 | 0.54 |
| GPR116 | ILMN_1712878 | 0.14 | 0.68 | 0.54 |
| ZDHHC9 | ILMN_1734428 | 0.15 | 0.67 | 0.52 |
| SALL2  | ILMN_1740842 | 0.11 | 0.64 | 0.53 |
| UBTD1  | ILMN_1794914 | 0.15 | 0.67 | 0.52 |
| TMEM50 | ILMN_1765877 | 0.17 | 0.68 | 0.51 |

Listing of genes that fit the following criteria: i) significantly correlated with NKX2-1 in lung adenocarcinomas; ii) correlation of the same pair is 0.5 lower in normal lung tissue. This list is enriched for genes with roles in glutamine biosynthetic process (highlighted in bold).
One possible explanation for the lack of association to mitosis in normal lung is that relatively few cells in normal lung are actually cycling. We therefore investigated the correlations for Vrk1 in an independent data set derived from mouse skin, which is a proliferative self-renewing tissue\(^5\). No significant enrichment for Vrk1 correlations with mitosis genes was found in these normal tissues by BiNGO analysis (Supplementary Fig. S3a), although previous analysis has clearly shown the presence of a strong normal mitotic network\(^5\). In contrast, significant numbers of mitosis-related genes correlating with VRK1 were seen in tumour gene expression data sets from mouse carcinomas (Supplementary Fig. S3b), and in an independent set of human lung tumours\(^36\) and two human breast tumour data sets\(^37,38\) (data not shown).

As VRK1 has previously been implicated in p53 signalling and DNA damage responses\(^33,34\), we calculated the correlation between VRK1 and all other genes separately in adenocarcinomas with a wild-type TP53 gene (\(N = 54\)) and those with mutant TP53 (\(N = 23\)). We identified genes significantly correlated with VRK1 in TP53 wild-type (WT) but not TP53 mutant tumours, requiring a change in \(r\) statistic of at least 0.5. Of the nine genes matching these criteria, several are involved in DNA damage responses, including X-ray repair cross-complementing protein 3 (XRCC3), a RAD51 family member involved in double-strand break repair, FKBP3, an MDM2-interacting protein that contributes to TP53 regulation\(^39\), and CDC25A, a key cell cycle checkpoint gene inhibited when TP53 is activated during the DNA damage response\(^40\). This provides supporting evidence that VRK1 influences the TP53 DNA damage response pathway, but this influence is ablated in tumours where TP53 has been mutated. We also performed EGFR and K-ras mutation screening and divided samples into mutant and wildtype for gene expression network analysis. There was no significantly different network based on either EGFR or K-ras mutation status, due to the small numbers of samples in each subset and the relative heterogeneity of tumours within a subset.

VRK1 colocalizes with cell cycle markers. Three lung cancer cell lines (A549, H1299 and H460) were transfected with vectors encoding VRK1 short hairpin RNA (shRNA), and immunofluorescence (IF) staining was used to verify correlation of VRK1 expression with cell cycle markers such as Ki-67 and CDC2. Ki-67 is a well-known proliferation marker and CDC2 (CDK1) is a critical kinase involved in the G2/M phase of cell cycle progression. VRK1 and Ki-67 showed mainly nuclear staining in two lung cancer cell lines tested (A549 and H460) (Fig. 5a,b). Down-regulation of VRK1 using shRNA in both cell lines led to depletion of nuclear staining for both Ki-67 and VRK1, with residual staining mainly in the cytoplasm or nuclear membrane. CDC2, which was correlated in gene expression with VRK1 in the tumour networks, showed both cytoplasmic and nuclear staining in A549 and H1299 and mainly nuclear staining in H460 (Fig. 5c–e). CDC2 expression was colocalized with VRK1 and was reduced in VRK1 shRNA-transfected cells (Fig. 5c–e). These data show that VRK1 correlates with CDC2 at both mRNA and protein levels.

We also checked the protein expression levels of VRK1 and CDC2 in normal lung and matched tumour tissue slides by immunohistochemistry (IHC) (Supplementary Fig. S4). VRK1 showed weak staining in mainly bronchioles and was very weak in other regions of normal lung, while it exhibited very strong staining in certain regions of tumours (Supplementary Fig. S4). Although VRK1 mRNA expression increased by only \(1.5–2\) fold in lung tumours compared with normal lung (Fig. 4), VRK1 protein levels were strongly increased in lung tumours (Supplementary Fig. S4). Staining patterns observed for CDC2 were similar to patterns for VRK1 in both normal and matched tumour slides (Supplementary Fig. S4). While bronchiole regions showed only weak but recognizable staining in normal lung slides, most of the tumours were stained by antibodies against CDC2.

Combined VRK1 and PARP inhibition alters lung cancer growth. We designed shRNA probes to downregulate VRK1 expression and transfected them into lung cancer cell lines. After

---

**Figure 4 | VRK1 correlation network in adenocarcinomas.** (a) Gene pairs that are significantly correlated with VRK1 (highlighted in yellow) are drawn as in Fig. 4. This network is significantly enriched for genes that function in the M-phase of mitosis. (b) mRNA gene expression of VRK1, AURKA and CCNB2 in normal lung and matched tumour tissues. Box plot displaying fold-change values for matched lung adenocarcinoma and normal lung tissue for three genes: VRK1, Aurora Kinase A (AURKA) and Cyclin B 2 (CCNB2). The mean increase in gene expression between normal and matched tumour samples is smaller in VRK1 than in AURKA and CCNB2.
verifying that the shRNAs induced a significant VRK1 expression reduction (Fig. 6a), flow cytometry/PI staining was done in vector and VRK1 shRNA-transfected lung cancer cells. G1 arrest was observed in VRK1 shRNA-transfected H460 lung cancer cell line, whereas normal cell cycle progression was observed in vector-transfected H460 (Student’s t-test, \( P < 0.0001 \), Fig. 6b). Similar results were obtained with another lung cancer cell line H1299 (Fig. 6b). VRK1 has previously been shown to have a TP53-mediated role in DNA damage repair, and VRK1 gene expression is positively correlated with BRCA1 in lung tumours (correlation \( r = 0.59 \)). Poly (ADP-ribose) polymerase (PARP) inhibitors are novel anti-cancer drugs that are presently in clinical trials for treatment of breast cancers carrying BRCA1 mutations41. PARP inhibitors have been used for ovarian, breast and prostate cancer but not in lung cancers. Thus, we checked whether VRK1 shRNA-transfected cells may be sensitized to treatment with PARP inhibitors (ABT-888). Significantly increased sensitivity to PARP inhibitor was observed only in VRK1 shRNA cells compared with vector-transfected cells (Fig. 6c). Around a 40% decrease of cell viability was observed after ABT-888 treatment of control cells compared with dimethylsulphoxide (DMSO) treatment, but a 70% decrease in cell viability was shown after ABT-888 treatment of the VRK1 shRNA cell line (Student’s t-test, \( P = 0.001 \); Fig. 6c). This suggests that VRK1 could be a viable therapeutic target in lung cancer patients, especially in combination with PARP inhibitors. A colony-formation assay was done in H1299 lung cancer cell line with VRK1 shRNA and a control vector transfection. A significant reduction of colony number was identified in VRK1 shRNA-transfected cells after PARP inhibitor (ABT-888) treatment (Fig. 6d). ABT-888 treatment results in two additional lung cancer cell lines (H2170 and A549) with VRK1 shRNA are shown in Supplementary Fig. S5.

To check the effect of VRK1 in normal lung cell lines, we transfected VRK1 shRNA and a control vector into MRC-5 (normal human fetal lung fibroblasts) and BEAS-2B (normal human lung epithelial cells immortalized with an adenovirus 12-SV40 virus hybrid (Ad12SV40)). There was no effect of VRK1 downregulation on growth of the MRC-5 cells (Student’s t-test, \( P = 0.64 \)), although treatment with ABT-888 (PARP inhibitor, 20 \( \mu M \)) in both VRK1 shRNA and control vector-transfected cells reduced cell growth. There was no synergistic effect of VRK1 downregulation and PARP inhibition (ABT-888 20 \( \mu M \)) in both MRC-5 (Student’s t-test, \( P = 0.45 \)) and BEAS-2B (Student’s t-test, \( P = 0.142 \)) cells, which supports our tumour-specific VRK1 mitotic network data. We selected two genes, Gli1 (control) and AURKA (a known mitotic cycle gene), to test the possibility that downregulation of any cell cycle gene may act together with PARP inhibition to reduce cell viability. Short interfering RNAs (siRNAs) for Gli1, AURKA, VRK1 or a scrambled control siRNA (Qiagen) were transfected into H2170 lung cancer cell line. ABT-888 (20 \( \mu M \)) was added to the cell line 24 h after siRNA transfection and measured for cell viability using CellTiter-Glo Luminescent Cell Viability Assay kit (Promega). No significant additive or synergistic effect of PARP inhibition and

Figure 5 | IF staining of VRK1 in lung cancer cell lines. (a,b) VRK1 and K-67 was also costained with different fluorescent dyes (VRK1, green; Ki-67, red) in vector and VRK1 shRNA-transfected cells. After shRNA knockdown of VRK1 in cell lines (A549 and H460), both VRK1 and Ki-67 expression was significantly reduced. (c–e) VRK1 (green) and CDC2 (red) were costained in three lung cancer cell lines (A549, H460 and H1299). VRK1 and CDC2 were found to be mainly colocalized in VRK1 shRNA cell lines. Mouse monoclonal VRK1 antibody (1F6) was used for panels a–e. Scale bar, 50 \( \mu m \).
siRNA-mediated downregulation of Gli1 or AURKA was observed, while VRK1 siRNA treatment showed a significant cell viability decrease with PARP inhibition (Supplementary Fig. S6).

**Discussion**

Lung cancer is projected to increase substantially in the future due to the continued expansion of smoking habits particularly in Asia. Substantial advances have been made in development of targeted treatments for lung cancers, in particular through the availability of inhibitors of EGFR (Cetuximab, Tarceva). These drugs have been used successfully for treatment of lung cancers carrying activating mutations in the EGFR kinase, but a consistent effect on long-term survival has remained elusive, and most patients relapse due to development of resistance through various mechanisms. This paradigm, by which targeted therapies initially show success in subsets of patients followed by development of resistance, has been seen also for many other novel cancer drugs including Imatinib (Gleevec), and a series of BRAF kinase inhibitors. There is therefore a major requirement for approaches that allow us to visualize the signalling networks that control both normal and tumour cell behaviour, to identify signalling hubs that may allow development of synergistic combinations of inhibitors and chemotherapeutic drugs.

Network analysis of genetic control of gene expression has been previously carried out in lower organisms such as yeast and has been applied to tissues from heterogeneous populations of mice. We have applied this network analysis approach to matched normal and tumour tissue from patients with lung adenocarcinomas. Analysis of the coordinated changes in gene expression between normal lungs and matched tumours resulted in a ‘progression network’ that displays the main structural and functional components of the lung that show global changes in gene expression during adenocarcinoma development. Lung tumours contain invading immune inflammatory cells, tumour-associated stroma and vascular cell populations. While we believe that motifs representing these cell types are relevant to the overall biology of lung adenocarcinomas, the proportion of epithelial to non-epithelial tissue in each sample is variable, and this cellular heterogeneity will have an impact on the normal and tumour networks. This approach allowed us to identify motifs associated...
with the major type 2 pneumocyte cell lineage, and the changes in these motifs that accompany transformation to adenocarcinomas. The network identified structural components of the cells in the type 2 pneumocyte lineage (for example, the genes encoding the surfactant proteins) and the transcription factors such as NKX2-1 that are linked to cell type-specific morphogenesis and lineage determination. The NKX2-1 transcription factor is a classical marker for adenocarcinomas and although the surfactant protein gene expression motif is downregulated in the tumours, NKX2-1 remains linked to this network. However, both the progression network and the differential correlation analysis that measures changes in the edges linking genes independently of alteration in actual levels, demonstrated rewiring of NKX2-1 networks in tumours. Novel genes associated with NKX2-1 only in tumours included several genes linked to glutamine and glutathione homeostasis that are critical for growth and survival of tumour cells. Several of these genes are enzymes that may be suitable for drug development to find combinations of small-molecule inhibitors that coordinately affect this tumour-specific network.

Elevated expression of NKX2-1 may for example be a biomarker for tumours suitable for treatment with inhibitors of GLS2 that have already been identified. NKX2-1 was originally implicated as an oncogene, showing amplification and overexpression in a subset of lung cancers,

but more recent data from mouse models suggests that NKX2-1 may act as a tumour suppressor. High expression of NKX2-1 is linked to tumour metabolism and growth, but the tumours in this category also express high levels of differentiation markers linked to good prognosis. NKX2-1 may therefore be neither an oncogene nor a tumour suppressor in the classical sense, but act as a mediator of alveolar type 2 lineage determination, loss of which allows development of less well-differentiated tumours expressing new markers of progression. These data may help to rationalize recent controversy regarding the proposed disparate roles for NKX2-1 in lung cancer.

The testing of all possible drug targets or their combinations that can be identified by this network approach is not possible within the scope of the present study. We therefore sought to validate the network predictions, both in normal lungs and in lung tumours, by carrying out immunohistochemical studies on normal lung, primary tumours and cell lines to detect coexpression of network components at the protein level. For normal lung, this analysis identified PEBP4 as a novel marker within the type 2 pneumocyte lineage in the lung. Although this gene is quite widely expressed and is therefore not uniquely expressed in lung-derived cells, it is nevertheless specifically expressed in type 2 pneumocytes both in normal lung and adenocarcinomas, and may therefore be a useful marker for tumours of this lineage. These data, therefore, demonstrate the value of whole-tissue gene expression correlation analysis for identification of markers expressed only or preferentially in specific cell types within the tissue.

To identify potentially novel drug targets in lung tumours, we focused initially on the network motif associated with mitosis, which historically has proven to be a rich source of ‘druggable’ genes for cancer therapy. Inhibitors of several mitotic kinases, including AURKA and AURKB, PLK1-4 and MELK, have been developed and several are presently in clinical trials. Initial results using some of these inhibitors have shown that their utility may be limited by toxicity due to inhibition of the cell cycle in completely normal dividing cells. We therefore searched for kinases that are integral components of the cell cycle network in tumours, but are not linked to the same markers in normal tissue such as human lung or mouse skin, which is an example of a proliferative normal tissue. VRKI was a suitable candidate, as it was not part of the cell cycle network in normal tissue, but was strongly associated with cell cycle and DNA damage response markers in a range of cancer types, including lung (Fig. 4 and Supplementary Fig. S2). BRCA1 mutations are known to induce sensitivity to the effects of PARP inhibition. We checked the ‘BRCA1’ network in normal lung and matched tumour samples and found that VRKI and BRCA1 showed a significant correlation only in tumours, and not in normal samples. Moreover, the BRCA1 network showed significant enrichment for genes involved in the cell cycle, DNA repair and DNA damage (cell cycle: $P = 1.1 \times 10^{-50}$, DNA repair: $P = 3.3 \times 10^{-19}$ and a response to DNA damage: $P = 4.8 \times 10^{-21}$ (P-values from BINGO), which is consistent with known functions of BRCA1 and also confirms the robustness of our network analysis. The strong similarity between the BRCA1 and VRKI networks suggested that downregulation of VRKI may be a good strategy for sensitization to PARP inhibition. Downregulation of VRKI using shRNA resulted in significant growth arrest in lung cancer cells, particularly in combination with a small-molecule inhibitor of PARP. While additional studies will be required to assess VRKI as a drug target in a range of cancer types, we believe that this study validates the network approach both for identification of coordinated changes in gene expression motifs that are linked to cancer development, as well as for identification of genes within these motifs that may be suitable for drug development. The application of these bioinformatics tools to matched normal and cancer samples from human patients has the potential to elucidate mechanisms of network rewiring in cancer cells that are not dependent on changes in absolute levels of gene expression in tumours—traditionally a prerequisite for identification of a candidate drug target.

### Methods

#### Lung cancer patients

Paired normal/lung tumour (adenocarcinoma) RNA and DNA samples were obtained from patients undergoing surgery at the University of California, San Francisco. All tumour samples in this study were specifically selected to allow a focused analysis of adenocarcinomas, and they contain at least 50% of tumour cells. All samples were obtained from patients with informed consent under the IRB approval of the University of California, San Francisco.

#### Microarray analysis

RNA was extracted with Trizol followed by DNase I treatment and purification (RNase A purification kit, Qiagen). The purified RNA was quantified with Nanodrop and checked by Bioanalyzer for the quality. Gene expression was measured with the Illumina Human Whole Genome 6 2.0 array (Illumina Inc., San Diego, CA, USA). Sample preparations, hybridization and scanning were performed according to manufacturer’s instructions. Briefly, 200 ng of total RNA was prepared to make cRNA by using Illumina TotalPrep RNA Amplification Kit (Ambion). First-strand complementary DNA was generated with T7 oligo (dT) primer and ArrayScript and then second cDNA was also made with DNA polymerase.

Biotin-NTP mix with T7 enzyme mix was used to make biotinylated cRNAs. The labelled cRNAs were purified, quantified and checked again with Bioanalyzer. The labelled cRNA target was used for hybridization and scanning according to Illumina Human Whole Genome 6 2.0 array protocol. Probes with present/absent call $\geq 0.05$ assessed by Illumina Beadstudio software were marked absent. Probes were re-annotated using data from ref. 57.

#### Statistical analysis

Raw microarray data were quantile normalized and log2-transformed. Correlation analysis was performed using probes present in $\geq 90\%$ of samples using Spearman rank correlation. Correlation was defined as significant at the 5% alpha level using the experiment-wise Genome-Wide Error Rate permutation method as described in ref. 58. To calculate tumour progression networks, lung and adenocarcinoma microarrays were normalized together. Change in expression levels between normal and tumour tissue was assessed using a paired t-test in SAM, retaining genes with false discovery rate $\leq 5\%$ and greater than the twofold change. The progression network was plotted using all differentially expressed gene pairs where correlation in change from normal to tumour was $\geq 0.7$ and where each gene was a member of a clique with size $\geq 3$. The decision to use networks of degree 3 was made to enrich for sets of genes, which were correlated with each other without eliminating all but the largest and densest sub-networks. We are using clique membership as a way to increase the likelihood of finding a biologically relevant signal. Gene expression networks were plotted in Cytoscape (http://www.cytoscape.org). Gene Ontology enrichment was calculated using BINGOGO. Meaningful decreases in correlation was defined as correlation at the
GWER alpha-level for the target condition (for example, tumour tissue) and a decrease in correlation magnitude of ≥0.5 in the other condition (for example, normal tissue). To facilitate complete replication of the statistical findings presented in this manuscript from the raw data (Supplementary Data 1–5), a software script is available as Supplementary Software.

**Mutation screening.** PCR reactions were generally carried out in a volume of 25 μl containing 10 ng genomic DNA, 10 pmol of each primer, 250 units of polymerase and the reaction buffer provided by the supplier (Qiagen, Hilden, Germany). Samples were denatured for 5 min at 94 °C in a PCR machine (MJ Research, Waltham, USA), and then amplified by 35 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min, with a final elongation of 10 min at 72 °C. Hot-spot areas of K-ras (codons 12 and 13) and TP53 (exons 4–8) were screened by bidirectionally sequencing using the Taq dye terminator cycle sequencing kit and an ABI 3730 DNA sequencer (Applied Biosystems, Foster City, CA). EGFR mutation was screened by Sequenom OncoCarta panel (Sequenom).

**VRK1 shRNA constructs.** VRK1 shRNA sequences were designed by BLOCK-iT RNAi Designer (Invitrogen). A shRNA was cloned into pSuper retrovirus cassettes (OligoEngine). The shRNA sequence for VRK1 is 5′-GCAGA-TATGCAAGGCTCTAAT-3′ (VRK1_sh526). The lung cancer cell lines (A549, NCI-H460 and NCI-H1299) were infected with retroviral stocks produced by the transfection of 293T Phoenix cells using lipofectamine 2000 (Invitrogen). After infection with supernatant medium of VRK1 shRNA-transfected cell, cells were selected with 1–5 μM of puromycin. A pluper vector plasmid without shRNA sequences was used as a control. Cell culture was done with RPMI medium 1640 and 10% fetal bovine serum (FBS).

**FACS analysis.** Cells were trypsinized in a 5-ml Petri plate with 1 ml of trypsin. After washing with PBS three times, 3 ml of 70% cold EtOH were added to cells. After washing with PBS three times , cells were then resuspended in 50 μl PBS with 50 μl RNaseA (Sigma, 5 mg ml−1) and incubated at 37 °C for 30 min. A volume of 500 μl of propidium iodide (PI, Calbiochem; 50 μg ml−1) was added to the cells and incubated at 4 °C for 30 min in the dark. The prepared cells were analysed with BD FACSCalibur Flow Cytometer and CellQuest Pro software. The data was analysed with FLOWJO software (Tree Star, Inc, OR).

**IF staining.** Cells were seeded and incubated for 24 h in a chamber slide (4-Well Lab-Tek Chamber Slide, BioExpress) and then washed with PBS after removing a medium. A solution with acetone/methanol (1:1) was used for fixing cells for 20 min in −20 °C followed by PBS washing three times. Slide-blocking process was done with solution (10% normal goat serum plus 3% Triton X-100) for 30 min at room temperature. Primary antibodies in a blocking solution were incubated with slides for 2 h at room temperature (dilution 1:500). VRK1 antibodies were used from either Sigma (Rabbit polyclonal, HPA000660) or 1F6 (mouse monoclonal, sc-25324, Santa Cruz Biotechnology), SFTPD (sc-25324, Santa Cruz Biotechnology) and EGFR was selected with 1–5 μM of antibodies. A volume of 500 μl of PI was added to the cells and incubated for 30 min in a dark, mild vortexing was done. These cells were stored at least overnight in a dark, −20 °C.

**Immunohistochemistry staining.** The same antibodies that were used for IF staining (VRK1 and CDC2) were used for IHC staining. Antibodies for SFTPBP (sc-133143, Santa Cruz Biotechnology), SFTPFD (sc-25324, Santa Cruz Biotechnology) and PEP4B (HPA025864, Sigma) were used in IHC staining. Sections (5 μm in thickness) were hydrated in xylene and varying concentrations of ethanol and then steamed in citrate from Biogenex for 20 min. The slides were then incubated at room temperature with a blocking solution (10% goat serum in TBS and 1% BSA) for 1 h. A blocking solution was removed and the primary antibody was applied. The slides were incubated at 4 °C overnight. Following the overnight incubation, the slides were rinsed in 1 × TBS containing 0.025% Triton, 1 × TBS containing 0.3% H2O2, and 1 × TBS. The slides were incubated using reagents from the Invitrogen Histostain Plus Broad Spectrum Kit (85-9643) according to the protocol provided by Invitrogen. The slides were rinsed in running water for 5 min and then dipped in Gills no. 2 hematoxylin stain for 1 min. The slides were rinsed in running water for 1 min and dehydrated in xylene and varying concentrations of ethanol. The slides were then mounted using organic mounting media and stored at 4 °C.

**PARP inhibitor treatment.** ABT-888 (20 μM; Chemie Tek) was treated to cells in a serum-free condition. After 48 h of ABT-888 treatment, cell viability was measured using Promega CellTiter-Glo Luminescent Cell Viability Assay kit and Synergy HT Multi-Mode Microplate Reader (Biotek).
