Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas

Wu, Kui; Zhang, Xin; Li, Fuqiang; Xiao, Dakai; Hou, Yong; Zhu, Shida; Liu, Dongbing; Ye, Xiaofei; Ye, Mingzhi; Yang, Jie; Shao, Libin; Pan, Hui; Lu, Na; Yu, Yuan; Liu, Liping; Li, Jin; Huang, Liyan; Tang, Hailing; Deng, Qiuhua; Zheng, Yue; Peng, Lihua; Liu, Geng; Gu, Xia; He, Ping; Gu, Yingying; Lin, Weixuan; He, Huiming; Xie, Guoyun; Liang, Han; An, Na; Wang, Hui; Teixeira, Manuel; Vieira, Joana; Liang, Wenhua; Zhao, Xin; Peng, Zhiyu; Mu, Feng; Zhang, Xiqing; Xu, Xun; Yang, Huanming; Kristiansen, Karsten; Zhong, Nanshan; Wang, Jun; Pan-Hammarström, Qiang; He, Jianxing; Wang, Jian

Published in:
Nature Communications

DOI:
10.1038/ncomms10131

Publication date:
2015

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Wu, K., Zhang, X., Li, F., Xiao, D., Hou, Y., Zhu, S., Liu, D., Ye, X., Ye, M., Yang, J., Shao, L., Pan, H., Lu, N., Yu, Y., Liu, L., Li, J., Huang, L., Tang, H., Deng, Q., ... Wang, J. (2015). Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas. Nature Communications, 6, [10131]. https://doi.org/10.1038/ncomms10131
Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas

Kui Wu1,2,3,4,5,*, Xin Zhang2,3,4,*, Fuqiang Li1,*, Dakai Xiao2,3,6,*, Yong Hou1,5,*, Shida Zhu1,5,*, Dongbing Liu1, Xiaofei Ye1,7, Mingzhi Ye1,8, Jie Yang1, Libin Shao1, Hui Pan2,3,6, Na Lu1, Yuan Yu1, Liping Liu2,3,6, Jin Li2,3,6, Liyan Huang2,3, Hailing Tang2,3, Qiuhua Deng2,3,6, Yue Zheng1, Lihua Peng1, Geng Liu1, Xia Gu9, Ping He9, Yingying Gu3,9, Weixuan Lin6, Huiming He6, Guoyun Xie1, Han Liang1, Na An1, Hui Wang1, Manuel Teixeira10, Joana Vieira10, Wenhua Liang2,3,4, Xin Zhao1, Zhiyu Peng1,8, Feng Mu1,11, Xiuling Zhang1,8, Xun Xu1, Huanming Yang1,12, Karsten Kristiansen1,2, Jian Wang1,12, Nanshan Zhong3,4, Jun Wang1,5,**, Qiang Pan-Hammarström1,7,** & Jianxing He2,3,4,**

The landscape of genetic alterations in lung adenocarcinoma derived from Asian patients is largely uncharacterized. Here we present an integrated genomic and transcriptomic analysis of 335 primary lung adenocarcinomas and 35 corresponding lymph node metastases from Chinese patients. Altogether 13 significantly mutated genes are identified, including the most commonly mutated gene TP53 and novel mutation targets such as RHPN2, GLI3 and MRC2. TP53 mutations are furthermore significantly enriched in tumours from patients harbouring metastases. Genes regulating cytoskeleton remodelling processes are also frequently altered, especially in metastatic samples, of which the high expression level of IQGAP3 is identified as a marker for poor prognosis. Our study represents the first large-scale sequencing effort on lung adenocarcinoma in Asian patients and provides a comprehensive mutational landscape for both primary and metastatic tumours. This may thus form a basis for personalized medical care and shed light on the molecular pathogenesis of metastatic lung adenocarcinoma.
lung cancer is the leading cause of cancerous deaths worldwide\textsuperscript{1,2} with two major types: non-small-cell lung cancer (NSCLC) and small cell lung cancer (SCLC), accounting for \textasciitilde 85\% and 15\% of all diagnosed lung cancers, respectively\textsuperscript{3}. Lung adenocarcinoma is the most common histological type of NSCLC, resulting in \textasciitilde 500,000 deaths globally every year\textsuperscript{4}. Despite advances in surgery, molecular subtyping and targeted therapy, prognosis of lung adenocarcinoma remains poor and the reasons for this could be due to: (1) Diagnosis was often made already at a late stage when localized malignant tumours spread to regional and distant tissues\textsuperscript{5}; (2) Lack of known targetable driver genes in approximately half of the diagnosed patients\textsuperscript{6,7}; (3) Complexity of inter- and intra-tumour heterogeneity\textsuperscript{8,9}; and (4) Poor understanding on the mechanism of metastasis development, as well as lack of corresponding treatment.

Previous studies have characterized the genomic landscape of lung adenocarcinomas and identified many potential cancer driver genes\textsuperscript{4,9–11}, of which targeting therapies have been developed for several activated oncogenes such as \textit{EGFR}, \textit{ERBB2} and \textit{BRAF}\textsuperscript{12–14} and translocations or fusions involving \textit{ALK}, \textit{ROS1} and \textit{RET}\textsuperscript{15–17}. Most of these studies, however, mainly focused on tumour samples obtained from European or North American patients, and the majority of specimens were collected at early disease stages. Undoubtedly, it is important to have a comprehensive genomic analysis on lung adenocarcinomas from Eastern Asian population when considering its rapidly increasing incidence rate and potential genetic heterogeneity between different ethnic populations. Moreover, genetic characterization of advanced lung adenocarcinomas especially those harbouring corresponding metastases will not only expand the spectrum of potential cancer driver genes involved in lung carcinogenesis, but also improve our knowledge on metastasis formation and further guide diagnosis and therapies for metastatic lung adenocarcinomas. A recent study demonstrated high concordance of recurrent somatic alterations between primary tumours and matched metastases in NSCLCs\textsuperscript{18}. This initial survey was limited though by focusing on targeted sequencing of 189 cancer-related genes.

Here we performed a comprehensive genetic analysis of 101 Chinese lung adenocarcinomas, as well as 35 corresponding lymph node metastases through multiplatform sequencing. Two hundred and thirty-four primary tumours were furthermore included into this study as an independent validation cohort. In an addition to a number of previously reported lung cancer driver genes, we have identified several novel, potentially oncogenic genes that are significantly mutated in our cohort. Integrative analyses through genomic data furthermore highlight pathways that may play a critical role in driving tumour metastasis. These results provide new insights on the pathogenesis of lung adenocarcinomas and also form a basis for further improvement of clinical management of our patients in the precision medicine era.

Results

Sample description and sequencing statistics. Paired tumours and normal adjacent tissues were obtained from 335 patients that provided written informed consent to carry out genomic studies in accordance with local Institutional Review Boards. All tumour specimens were reviewed by independent pathologists to determine the histological subtype, TNM stage and tumour cellularity (Supplementary Fig. 1). Detailed clinical features were summarized in Table 1 and Supplementary Data 1.

Whole-genome, transcriptome sequencing data were obtained from primary tissues and corresponding lymph node metastases of 24 Chinese lung adenocarcinoma patients. Genomes were sequenced to a mean depth of 49.6 \times (range: 42.0–57.8 \times) for primary tumours and 51.2 \times (44.4–66.3 \times) for metastatic specimens, while it was 31.9 \times (23.6–34.9 \times) for adjacent normal tissues. On average 93M clean reads (73–110M) were generated from whole-transcriptome sequencing. Whole-exome sequencing (WES) was performed on both primary and matched lymph node metastases from an additional 11 cases, and from primary tumour samples from a further group of 66 patients. RNA-seq analysis was furthermore performed in 32 cases from the latter group. Exome sequencing reached median fold coverage of 81.4 \times (37.6–293.7 \times) and 37.3 M reads (20.0–53.4 M) were generated from RNA-seq (Supplementary Data 2 and 3). These 101 cases constituted the discovery cohort.

For verification, an independent hybrid-recapture and ultra-deep DNA sequencing were performed on a set of 51 selected genes with a mean fold coverage of 140 \times (48.2–455.1 \times) on custom target regions in 98 cases of the discovery cohort described above and additional 234 primary lung adenocarcinomas (the validation cohort) (Supplementary Data 4). The cohort description with different data sets was summarized in Supplementary Data 5.

Somatic DNA alterations and verification. Somatic point variations and small insertions/deletions (indels; <50 bp) were detected using MuTect\textsuperscript{19} and Platypus\textsuperscript{20}, respectively. A mean of 9.7 somatic mutations per Mb (1.7–64.4) were identified from the first 101 primary lung adenocarcinomas, while the mean somatic...

| Table 1 | Clinical feature summary of 335 sequenced lung adenocarcinomas. |
|---------|----------------------|----------------------|----------------------|----------------------|
|         | Discovery cohort     | Validation cohort    | Total                | P value              |
| Age at surgery, years | 59.2 (25.2–81.6) | 58.6 (32.4–84.9) | 58.8 (25.2–84.9) | 0.4918$^\dagger$ |
| Gender  |                       |                      |                      |                      |
| Male    | 62 (61.4\%)           | 121 (51.7\%)         | 183 (54.6\%)         | 0.1303$^\ddagger$   |
| Female  | 39 (38.6\%)           | 113 (48.3\%)         | 152 (45.4\%)         |                      |
| Smoking status |                       |                      |                      | 0.3373$^\dagger$     |
| Smoker  | 37 (36.6\%)           | 68 (29.1\%)          | 105 (31.3\%)         |                      |
| Non-smoker | 58 (57.4\%)          | 141 (60.3\%)         | 199 (59.4\%)         |                      |
| NA      | 6 (5.9\%)             | 25 (10.7\%)          | 31 (9.3\%)           |                      |
| Follow-up, months | 22 (4–80)            | 37 (1–77)            | 36 (1–80)            | 0.1042$^\dagger$      |
| Tumour stage |                       |                      |                      |                      |
| I       | 19 (18.8\%)           | 63 (26.9\%)          | 82 (24.5\%)          | 0.1447$^\dagger$     |
| II      | 18 (17.8\%)           | 51 (21.8\%)          | 69 (20.6\%)          |                      |
| III     | 56 (55.4\%)           | 98 (41.9\%)          | 154 (46\%)           |                      |
| IV      | 8 (7.9\%)             | 21 (9\%)             | 29 (8.7\%)           |                      |
| NA      | 0 (0\%)               | 1 (0.4\%)            | 1 (0.3\%)            |                      |
| Metastasis |                       |                      |                      | 0.1087$^\dagger$     |
| Negative | 25 (24.8\%)          | 80 (34.2\%)          | 105 (31.3\%)         |                      |
| Positive | 76 (75.2\%)           | 153 (65.4\%)         | 229 (68.4\%)         |                      |
| NA      | 0 (0\%)               | 1 (0.4\%)            | 1 (0.3\%)            |                      |
| Survival status |                   |                      |                      | 0.3612$^\dagger$     |
| Alive   | 67 (66.3\%)           | 160 (68.4\%)         | 227 (67.8\%)         |                      |
| Dead    | 33 (32.7\%)           | 60 (25.6\%)          | 93 (27.8\%)          |                      |
| NA      | 1 (1\%)               | 14 (6\%)             | 15 (4.5\%)           |                      |

NA, not applicable. \(^{\dagger}\) Wilcoxon rank sum test. \(^{\ddagger}\) Pearson’s \(r\)-test.

ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/ncomms10131
The mutational signature detection in our discovery cohort reveals a consistently higher mutation rate in lung cancer than the other tumour types except malignant melanoma, which links to exposure to ultraviolet light \(^ {21} \). Mutation spectrum analysis revealed that the most common somatic substitutions in our discovery cohort were C→G transversions and C→T transitions as previously reported for lung adenocarcinomas \(^ {4} \) (Supplementary Fig. 2). Mutational signatures of 24 cases with whole-genome sequencing (WGS) data were subsequently characterized on both primary and metastatic tumours based on the 96 possible mutation types and were compared with the WGS data derived from previously reported European lung adenocarcinomas \(^ {4} \). Four highly confident mutational signatures were extracted from each sample set and showed no apparent difference between primary and metastatic samples (signature 1, 2, 3, 4) (Fig. 1). When compared with the four signatures extracted from European lung adenocarcinomas, highest correlation (Pearson Correlation >0.99) was detected in signatures that were predominated by C→A mutations (Chinese signature 1 and European signature 1) and associated with cigarette smoking. The second highly correlated signature was Chinese signature 3 to European signature 3 (Pearson correlation >0.95), which attributed to over-activated members of the APOBEC family of cytidine deaminase. Indeed the expression levels of APOBEC1, APOBEC3B, APOBEC3C and APOBEC3F were found to be significantly higher in both primary and metastatic tumours as compared with normal tissues (\( P<0.01 \), Supplementary Fig. 3). Similar analysis was also applied to 10 previously reported stage I Chinese lung adenocarcinomas with WGS data \(^ {22} \). Three high confident signatures from these tumours were identified, which show high correlation with signature 1, signature 3 and signature 4 that were extracted from 24 Chinese lung adenocarcinomas at late stage and 22 European lung adenocarcinomas. This observation indicated the significant association of these three signatures with lung adenocarcinomas, at both early and late stage of the disease, and they are shared within different ethnic populations. It is notable that signature 2, with transitions and transversions that predominantly occur at ApNpA and TpNpT trinucleotides (Fig. 1), is specific to our cohort that was not found in any of the 30 cancer types reported before \(^ {21} \) or in early stage lung adenocarcinoma. Furthermore, the metastatic tumours seem to more often (46%, 11/24) carry dominant proportion of ‘signature 2’ (with >20% of contribution) as compared with the primary tumours (25%, 6/24). The contribution of ‘signature 2’ is highly correlated with the proportion of indel mutations, as tumours carried signatures with higher contribution of ‘signature 2’ also harboured significantly higher proportion of indel mutations in all somatic mutations detected (Supplementary Fig. 4a). In contrast, the other signatures do not show such linear association with indel mutation. Since the mutational signature analysis was carried out only using base substitutions, the association of indel mutation of specific signature should not be attributed to incorporation of indel mutations to the analysis. We then verified our finding in the mutation data of 22 European lung adenocarcinomas. We found that the contribution of ‘signature 2’ in these 22 European lung adenocarcinomas is only 4% on average, significantly lower than that in primary tumours (18%) and metastatic tumours (22%) in our sequenced cohort (\( P<0.001 \), Wilcoxin Signed Ranks Test). This explained that why this signature was not evidently identified from the previous analysis in the European sample set. However, we confirmed that in the European tumours, the contribution of ‘signature 2’ is also highly correlated with the proportion of indel mutations (Supplementary Fig. 4b). No germline or somatic mutations in DNA repair genes including various mismatch repair genes, BRCA1 and BRCA2 could explain the occurrence of ‘signature 2’. Although this signature seems to be due to inclusion of large number of late stage of lung adenocarcinoma patients with metastasis in our study, WGS data in a larger cohort of patients will be required to confirm this association and to further understand the underlying mutagenesis mechanism.

Distinct combinations of mutational signatures were observed in each individual cancer, which indicated inter-tumour heterogeneity (Supplementary Fig. 5a). It is interesting that though we identified similar landscape of mutational signatures between primary and metastatic sample sets, about 33% (8/24) cases showed different contribution of signatures between the paired primary and metastatic tumours, which may be attributed to intra-tumour heterogeneity or mutational selection during metastases (Supplementary Fig. 5b). No significant correlation was observed between the mutation signature distribution and patient age, gender, smoking status or tumour purity (Supplementary Data 10), which may be partially due to the limited sample size.

**Significantly mutated genes.** Given the high level of background mutation rate in lung cancer, onco-driver gene event analysis was carried out through a modified pipeline described previously \(^ {23} \), \(^ {24} \), which considered the mutation prevalence in the context of the background mutation rate and gene sequence length, as well as evaluation of functional impact. The MutSig \(^ {25} \) algorithm was subsequently applied to identify significantly mutated genes and the mutant frequency was calculated by integrating the discovery and validation cohorts, for altogether 335 primary tumours and 35 metastasis specimens. These analyses revealed 13 statistically significant mutated genes (\( q<0.1 \), Fig. 2a, Supplementary Fig. 6, Supplementary Data 11). The most frequently mutated genes in primary tumours are TP53 (44%), EGF (39%), LR1P1 (19%) and KRAS (11%), indicating the major contribution of these genes in lung carcinogenesis. The other frequently mutated genes include well-known tumour suppressor genes: PTEN (7%), STK11 (4%) and SMAD2 (2%), and oncogenes PIK3CA (5%), BRAF (4%) and FLTI (3%). Consistent with the previous studies, KRAS mutations are mutually exclusive with those of EGF, and are more commonly observed in smokers \(^ {1} , ^ {10} , ^ {26} \). We also confirmed a higher mutation rate of EGF than KRAS in our cohort, which is in contrast to that in the Caucasian populations. Notably, 54% (71/132) of the EGF mutations are Leu858Arg, and another 29% (39/132) EGF mutations are exon 10 deletions, which are
GLI3 was significantly associated with lymph node metastasis and a study on NSCLC demonstrated that overexpression of truncated GLI3 modulates the sonic hedgehog (SHH) pathway, and one recent study on Chinese lung adenocarcinoma patients confirms the importance of testing these specific mutations for sensitive targets of tyrosine kinase inhibitor therapies. This confirms the importance of testing these specific mutations for Chinese lung adenocarcinoma patients.

Mutations in RHPN2 (5%), GLI3 (4%) and MRC2 (2%) have not been reported previously as driver genes in lung adenocarcinomas but were recurrently observed in our cohort and are nominated to be significantly mutated. RHPN2, which encodes a member of the rhophilin family of Ras-homologous (Rho) GTPase-binding proteins, was first identified to interact with RhoA as a downstream effector molecule to regulate the actin cytoskeleton, a process that is involved in cancer cell migration and invasion. A recent study further demonstrated that RHPN2 could drive mesenchymal transformation in malignant glioma via RhoA activation. Notably, a novel but recurrent mutation V73M was identified for RHPN2 (Supplementary Fig. 6b), which affecting the Rho binding domain, implying that this site could be a potentially functional important hotspot mutation. GLI3 encodes a zinc-finger transcription factor that modulates the sonic hedgehog (SHH) pathway, and one recent study on NSCLC demonstrated that overexpression of truncated GLI3 was significantly associated with lymph node metastasis and poor survival. Another investigation also showed that high expression of GLI3, together with GLI1, is associated with tumour progression in advanced lung adenocarcinoma. The exact mechanism of GLI3 mutations in tumorigenesis and whether the potentially affected SHH signalling could be targeted have not been determined. MRC2 (also known as uPARAP, Endo180 or CD280), a member of mannose receptor family, is found to be involved in extracellular matrix remodelling by mediating collagen degradation. Functional experiments and clinical surveys have demonstrated that high expression of MRC2 can promote tumour growth and drive metastasis, results in significantly worse prognosis in several cancers.

Additional recurrently mutated genes that did not reach a statistical significance by MutSig, but may functionally impact carcinogenesis, were also identified. These include APC (5%), KEAP1 (4%), ATF7IP (4%), ITIH5 (3%), IQGAP3 (3%), MET (3%), ERBB2 (2%) and TERT (2%) (Fig. 2a, Supplementary Fig. 6 and Supplementary Data 11), of which KEAP1 was mutually exclusively mutated with EGFR. The mutation frequencies of all the mutated genes described above in metastatic specimens are also shown in Fig. 2a. It is interesting to note that though KRAS was frequently mutated in primary tumours, no mutations were called for this gene in the 35 sequenced metastases, including those 5 samples with paired, KRAS-mutated primary tumours. The five metastatic tumours with KRAS mutations ‘loss’ were first confirmed as genetically identical with the corresponding primary tumours, by comparing both the pre-sequencing mass spectrometric fingerprint genotyping (Supplementary Data 12) and the genome-wide SNPs called from sequencing data. Visual review of the sequencing reads in these samples subsequently confirmed the sufficient coverage of all mutated sites for high quality mutation calling (Supplementary Data 13). Visual inspection furthermore confirmed that none of the normal tissues carried any reads of the mutated KRAS allele. At most two or three reads with mutation can be visualized, however, from metastatic specimens, thus suggesting a significantly lower frequency (0.01 on average; range: 0–0.02) in the metastases as compared with the corresponding primary tumours (0.30 on average; range: 0.18–0.46). Sanger sequencing further showed evident mutant genotype in primary tumours but no peak of altered signal was detected either in normal tissues or metastatic tumours.

The mutation patterns were further correlated with other clinical features such as gender, smoking status, age and...
tumour stage. The EGFR mutations were more frequently observed in females, while mutations in LRP1B, KRAS, PTPRD, APC, STK11, KEAP1, ITIH5 and FLT1 were significantly enriched in males ($P<0.05$, Fisher’s exact test). While EGFR mutations were enriched in non-smokers, mutations of LRP1B, KRAS, PTPRD, KEAP1 and GLI3 were significantly enriched in smokers ($P<0.05$, Fisher’s exact test). Notably, TP53, GLI3 and ITIH5 were significantly more mutated in patients aged $>60$ years ($P<0.05$, Fisher’s exact test) but none of the recurrently mutated genes was associated with tumour stage (Fig. 2b).

Metastasis is one of the most critical issues for disease progression in lung adenocarcinoma, however, whether the mutation patterns are different between tumours with or without metastasis is yet to be studied. All the samples were thus classified into two groups, those who had metastases in adjacent lymph nodes or distant organs on diagnosis or surgery (PM +, $n = 229$), and those who were metastasis free at diagnosis (PM −, $n = 105$). This classification was not associated with gender ($P = 0.64$), smoking status ($P = 0.79$) or age ($P = 0.08$). Fisher’s exact test indicated that TP53 was the only gene that was significantly ($P<0.05$) enriched in patients harbouring metastases (Fig. 2b), suggesting an important role of this tumour suppressor gene not only in primary tumorigenesis but also in driving the metastatic process.

Kaplan–Meier survival analysis was performed in all patients ($n = 335$) and validation cohort ($n = 234$), respectively, to explore the potential association between the recurrently mutated genes and individual outcome. Individuals harbouring somatic mutations in TP53, LRP1B, STK11, KEAP1, BRAF, MET and MRC2 had significantly shorter survival time than individuals with wild-type genotypes (Supplementary Fig. 7), which suggested that alterations in these genes could be used as prognostic markers in clinical practice.
Somatic CNVs and mRNA expression profiling. Somatic copy number alterations (SCNAs) were profiled in the 101 primary tumours and 35 metastases from the discovery cohort. The results revealed frequent abnormalities of chromosomal arms involving gain of 1q, 3q, 5p, 7p/q, 8q, 14q, 16p, 17q and 20q, as well as loss of 3p, 4q, 6q, 8p, 9p, 12q, 13q, 15q, 17p and 18q (Supplementary Fig. 8), most of which were consistent with previous results reported in lung adenocarcinomas. GSITIC\(^7\) was applied to identify statistically significant recurrent focal copy number variants (CNVs) (Fig. 3a, Supplementary Data 14). Significant somatic amplifications of NKX2-1, TERT, EGF, CCND1, MDM2, CDK4, MET, MYC and MECOM were observed, as well as deletion of TP53, PTPRD and CDKN2A/2B that have been widely reported in lung adenocarcinoma. Significantly amplified regions that encompassed genes involved in cytoskeleton organization or focal adhesion were also identified, including IQGAP3, TRIO, FSCN1/FSCN2 (fascin homologue 1/2, actin-bundling protein), RAC1/RAC3 and ITGB4/ITGB8. Recurrent, novel focal deletion regions were furthermore detected affecting tumour suppressor genes SMAD2 and SMAD4. Profiling of CNAs in metastatic tumours additionally identified amplification of SOX3 and deleted regions affecting STK11, KEAP1 and MGA. It is interesting that CNAs of several genes with function in histone or chromatin modification, such as KDM5B, CREBBP, SETD2, SMARCA4 and MECP2 were also found either in primary or metastatic tumours, suggesting a potential role of epigenetic regulation in lung adenocarcinoma. Profiling of mRNA expression status of 56 tumour/adjacent normal pairs confirmed correlated higher expression level in genes with copy number amplification than those with deleted CNAs (Supplementary Fig. 9).

Unsupervised clustering analysis of gene expression data from the 56 primary tumours identified three subgroups (cluster 1, cluster 2 and cluster 3) (Fig. 3b): Kaplan–Meier survival analysis indicated that cluster 3, which has higher frequencies of KRAS,

---

Figure 3 | Genomic copy number alterations and mRNA expression profiling. (a) Landscape of genomic copy number alterations in Chinese lung adenocarcinomas. Amplifications and deletions across chromosome 1-22 and X were shown with y-axis presenting G-score altitude. CNV profiles in primary and metastatic tumours were shown with different colours. Putative cancer driver genes were marked in locations with peaks across the genome. (b) Cluster classification of 56 tumours indicated three clusters with different gene expression pattern, 464 representative genes are included. (c) Gene expression clusters integrated with genomic mutations. Tumours were ordered as three clusters shown in b. Alterations of selected genes were shown across clusters, revealing mutations (Mut.) of KRAS, KEAP1, FLT1 as well as copy number amplification (Amp.) of CEPI72 were enriched in cluster 3. Cluster 3 was also characterized as having exceeded expression (Expr.) status in genes participating in PI3K-Akt pathway or cytoskeleton remodelling process.
FLT1, KEAP1 mutations and amplified CEP72, demonstrated a significantly shorter survival time than patients in the other two clusters (Supplementary Fig. 10). We also found that several genes participate in the PI3K-Akt pathway or cytoskeleton remodelling process, such as PIK3CA, PKN2, PTK2, DOCK1, SOS1 and MAPK1 had elevated expression status in cluster 3 as compared with cluster 1 and cluster 2, implying the activation of these processes may direct to disease progression (Fig. 3c).

**Identification of structural alterations.** Taking advantage of high depth WGS (50× coverage on tumour samples and 30× coverage for adjacent normal tissues) and corresponding transcriptome sequencing of 24 cases with paired primary and metastasis tumours sequenced, the structural alterations in lung adenocarcinoma were characterized. A total of 2,287 and 1,910 somatic structural aberrations were detected from primary and metastatic tumours, respectively (Supplementary Data 15 and Supplementary Fig. 11). Among the translocation events in primary tumours, 1,404 were inter-chromosomal (CTX), 201 were intra-chromosomal (ITX), 295 were inversions (INV) and 387 were deletions (DEL). The numbers of aberrant events in metastases were 1,467 CTXs, 127 ITXs, 107 INVs and 211 DELs. Well-known fusion genes EML4-ALK were identified in 2 out of these 24 cases in both primary and metastatic tumours with supportive evidence from both DNA and RNA data. Further screening through targeted sequencing of all 335 cases revealed 7 additional tumours harbouring this gene–gene fusion, giving an overall 3% alteration rate. Concurrent and mutual exclusion analysis indicated that EML4-ALK fusion was mutually exclusive with EGFR mutations (Supplementary Fig. 12), consistent with a recent large-scale clinical screening of NSCLC for EGFR, KRAS mutations and ALK rearrangements. Previously reported rearrangements of ROS1 and RET genes in NSCLC patients were not observed in our patient cohorts.

We further identified additional 13 putative gene–gene fusion events in these 24 metastatic cases using a combination of the whole-genome and transcriptome sequencing data (Supplementary Data 16, Supplementary Fig. 13). It is notable that among these fusions, several genes that function in cytoskeleton regulation were involved, including IQGAP3, EPB41, CDC42, PARD6G, PTK2B and KALRN, suggesting that these alterations may promote metastasis development. Other fused genes with putative role in tumorigenesis or metastasis formation include CHEK2, FGFR2, MIIIP, CPB1, EIF4A2, MECOM and USP32.

**Analysis of the IQGAP3 locus.** Integrated analysis of somatic mutations, CNAs, structural aberrations and gene–gene fusions using multiplatform studies highlighted IQGAP3 was altered in a notable number of cases. This gene, a member of the IQGAP family that has both IQ motifs and Ras GTPase-activating protein-related domain (GRD) was altered in 10% of lung adenocarcinomas (discovery cohort, n = 101), by somatic mutations (three cases), copy number amplification (seven cases) and translocation (one case). Targeted regional sequencing of the validation cohort (n = 234) further identified eight additional cases harbouring somatic mutations on IQGAP3.

Analysis of mRNA expression of 56 RNA sequenced primary tumours and 24 RNA sequenced metastases revealed that IQGAP3 was expressed at a significantly higher level in tumour specimens than adjacent normal tissues (Supplementary Fig. 14). The IQGAP3 expression in 132 primary lung adenocarcinomas (61 from discovery cohort and 71 from validation cohort) with available RNA within the 335 sequenced cases was subsequently measured by quantitative reverse transcription–PCR (RT–PCR; Supplementary Data 17), and confirmed the higher expression level of IQGAP3 in these tumours (Fig. 4a). An elevated level of IQGAP3 protein was also confirmed in a subset of primary tumours by immunohistochemical staining (n = 42, P < 0.001, Wilcoxon Signed Ranks Test, Supplementary Fig. 15). Kaplan–Meier analysis indicated that patients with a higher level of IQGAP3 mRNA had significantly shorter overall survival and disease-free survival (Fig. 4a, Supplementary Fig. 16). Furthermore, Multivariate Cox proportional hazard analysis showed that
a higher level of IQGAP3 was an independent risk factor for poor prognosis of the lung adenocarcinoma patients (Supplementary Data 18 and 19, hazard ratio = 2.28, 95% confidence interval = 1.25–4.14, P = 0.007).

The association of IQGAP3 genetic alterations and mRNA expression was subsequently investigated. No significant difference was observed in IQGAP3 expression between tumours and normal tissues in the eight cases carrying somatic mutations, including four mutated sites located in the Ras GTPase-Activating Protein (RasGAP) domain (Q1085H, A1124T, D1164E and L1292 splicing site) and one truncated mutation that resulted in the loss of the RasGAP C terminus domain (Fig. 4b). Furthermore, patients harbouring IQGAP3 SNVs showed better survival status than the other cases though did not reach a statistical significance (Supplementary Fig. 7). By contrast, this gene was overexpressed in tumours affected by copy number amplifications (Fig. 4c). An inter-chromosomal rearrangement occurred between chromosome 17 and chromosome 1 was identified in patient WG04, in both primary and metastasis tumours, which results in a novel gene-gene fusion involving TMC6 and IQGAP3, with reversed 5’ and 3’ orientation (Supplementary Fig. 17a). The fused product consists of exons coding most of the RasGAP domain and the intact RasGAP C terminus domain of IQGAP3, while retained exon 1 and exon 2 of TMC6. RNA-seq data revealed markedly higher, outlier expression of IQGAP3 in primary and metastasis tumour of patient WG04, and investigation of reads coverage indicated that highly covered regions were enriched in exons containing RasGAP and RasGAP C terminus domains (Supplementary Fig.17b,c).

Altered key pathways and therapeutic targeting. Integrative analysis of altered key pathways affected by somatic mutations, CNAs as well as frequent rearrangements was performed to construct a comprehensive view of aberrant characteristics for lung adenocarcinoma (Fig. 5a). The most frequently altered pathways in primary tumours were p53 signalling and cell cycle process, affecting 87% of primary tumours and 89% of metastatic tumours. A majority of the observed mutations were somatic mutations in TP53, amplifications of MDM2, MYC, CCND1, CCNE2, CDK4 and CDK6, as well as loss-of-function variants including deletions of CDKN2A, CDKN2B, mutations and deletions affecting SMAD2 and SMAD4, of which TP53 mutations were mutually exclusive with MDM2 amplification (Supplementary Fig. 12). CHEK2, a serine/threonine-protein kinase involved in regulating cell cycle arrest, was altered either by somatic mutation, deletion or rearrangement. The RTK/Ras/PI3K pathway was also frequently altered in lung adenocarcinoma, with 88% of primary tumours and 80% of metastasis tumours in this study harbouring aberrations in genes involved in this pathway. Included were events affecting several therapeutic targets that were responsive to antibodies or kinase inhibitors, such as alterations of EGFR, PIK3CA, ALK, ERBB2, MET, AKT1 and mTOR (Fig. 6).

Specifically, an additional pathway that involves actin cytoskeleton remodelling and focal adhesion was frequently altered in lung adenocarcinoma, affecting 50% of primary tumours. In addition to IQGAP3, five other genes with important roles in cytoskeleton remodelling exhibited somatic mutations or copy number amplifications, namely, RHPN2, RAC1, RAC3, TRIO and Cdc42. FSCN1 and FSCN2 were found to be frequently amplified in our studied cohort. These two genes belong to the Fascin protein family, which organizes F-actin into parallel bundles and plays an important role in cell migration, motility, adhesion and cellular interactions. Several studies have demonstrated a modulating function of Fascin proteins, especially FSCN1 in tumour metastasis and considered the overexpression of this gene as a prognostic marker for aggressiveness in carcinomas.

Network analysis further extended the interactive association between genes and pathways (Fig. 5b) that promote regulation of cytoskeleton remodelling. It is notable that this process was more frequently in metastasis samples (77%) than primary tumours (50%), implying the potential crucial contributions of these genes in driving metastasis.

Discussion

Although lung adenocarcinoma is one of the most well-studied cancers worldwide, the accuracy of diagnostic and therapeutic effectiveness are far from satisfactory, especially for patients harbouring metastases. In the present study, we carried out a comprehensive genetic characterization of 335 lung adenocarcinomas by integrating WGS, WES and RNA-seq. Sixty-eight per cent of studied cases were metastasis positive, providing an opportunity for the investigation of potential mechanisms and identifying biomarkers for future therapies.

Our study first delineated the similarity and distinctive mutational features of lung adenocarcinomas between Chinese and Caucasians. We discovered signatures associated with smoking and activation of APOBEC molecules that were shared in tumours from both ethnic populations and also detected signature specific to our cohort, which consists mainly late stage lung adenocarcinomas and is associated with the proportion of indel mutations. EGFR was mutated at a much higher frequency in our cohort than in Caucasians, with a dominance of the Leu858Arg change. By contrast, KRAS, the second most commonly mutated gene in Caucasians, was only found in 11% of Chinese patients. Notably, we observed significantly lower frequency or even loss of KRAS mutation in the metastatic tumours as compared with their paired primary tumours. The discordance of KRAS mutations in primary tumours and corresponding metastases has been noted previously in a few cases of lung adenocarcinomas. Independent studies demonstrated that oncogenic KRAS itself initiates the growth of non-invasive tumours in mouse models of lung cancer, implying that KRAS mutations may induce tumorigenesis, but might not be responsible for tumour metastasis, at least in some cases. Alternative interpretation of such discordant observation could attribute to tumoral heterogeneity, thus analysis of more biopsies from both primary tumour and metastatic lymph nodes from the same individuals may give a more complete picture. Our results suggest that alterations in major cancer drivers may change during tumour evolution and the development of metastasis, which should be important in clinical application, for instance, it may not be sufficient to determine therapeutic strategy for metastatic tumours based solely on the mutation spectrum in primary tumours.

We identified three genes, RHPN2, GLI3 and MRC2 to be significantly mutated in our studied cohort. These genes are altered in lower frequency, therefore are not determined as driver genes in lung adenocarcinoma previously. According to the saturation analysis by Lawrence et al., the discovery of candidate cancer genes is significantly associated with the sample set used and taking lung adenocarcinoma as an example ~2,000 cases are required to achieve a comprehensive catalogue of cancer genes with sufficient statistic power when the genes are mutated in >3% of patients. Our discovery of novel significantly mutated genes in lung adenocarcinoma confirms the necessity of continuous sequencing and analysis effort on more tumour-normal pairs of cancer patients. Although our study presented the potential association of these mutations to lung adenocarcinoma...
through bioinformatic exploratory analysis, functional study is essential to demonstrate the exact mechanism of these mutated genes in lung adenocarcinoma carcinogenesis in the future. We also showed a significant enrichment of TP53 mutations in patients with metastases on diagnosis or surgery. TP53 was in fact the most frequently mutated gene in many types of cancers.

Figure 5 | Altered pathways and network in lung adenocarcinomas. (a) Somatic mutations, copy number alterations and genomic rearrangements affecting p53 signalling/cell cycle process, RTK/Ras/PI3K pathway, cytoskeleton remodelling regulation and histone/chromatin modification. Percentage presented alteration frequencies in 101 primary tumours and 35 metastatic tumours, respectively. (b) Network connection of genes involved in cytoskeleton remodelling regulation. Gene–gene interactions are inferred by CytoScape program.
Methods

Emerging evidence indicates that mutant TP53 not only loses its tumour-suppressing activity, but also can gain functions that contribute to tumour progression50–53. Many mutations of TP53 have been implicated in invasion, migration and metastasis in different tumour types (summarized in ref. 54).

Integrative pathway investigation revealed that the most commonly altered processes are related to p53 signalling/cell cycle progression and the RTK/Ras/PI3K (mTOR) pathways in lung adenocarcinoma. Within this group of molecules are several targets for kinase inhibitors that are successfully used for treatment of patients with sensitive mutations. Furthermore, the actin cytoskeleton remodelling process, along with genes involved in focal adhesion was observed to be frequently altered in lung adenocarcinoma. Among the mutated genes involved in cytoskeleton remodelling, IQGAP3 was highlighted as numerous different alterations were identified including a novel fusion event. Clinical association analysis indicated that a high level of IQGAP3 expression is a putative biomarker in disease progression and poor survival status, implicating IQGAP3 as a potentially targetable gene for novel therapeutic strategies. Overall our study provided the first comprehensive description of mutational landscape of lung adenocarcinoma in Chinese patients, defined potential mechanisms for cancer metastasis and provided a basis for effective personalized therapies for this lethal cancer. Whether the discovered novel genomic changes are specific for Asian population and/or represent the disease heterogeneity, remains to be further explored in larger sets of tumour samples.

Patients and clinical samples description

The primary tumour specimens, adjacent normal tissues and lymph node metastases (if available) were obtained from 335 patients with lung adenocarcinoma who underwent surgical resection in Department of thoracic surgery, the First Affiliated Hospital of Guangzhou Medical University. None of the patients were subjected to chemotherapy or radiotherapy before surgery. Tumours and adjacent normal tissues were frozen in liquid nitrogen before sequencing. The primary tumour specimens, 61% of tumours were obtained from 61% of the total patients, 63% of the total patients had lymph node metastases. All the constructed libraries were sequenced on HiSeq 2500 platform using 2 × 100-bp paired-end reads.

Genomic DNA and total RNA preparation

The genomic DNA from tissues was prepared using the QIAamp DNA Mini Kit (Qiagen) following the manufacturer’s instructions. The total RNA was prepared using TRIzol Reagents (Life Technologies). The quantity and quality of DNA/RNA were determined by agarose gel electrophoresis or Agilent 2100 analyzer. Mass spectrometric fingerprint genotyping of 21 common SNPs was performed for all the samples before sequencing, to verify that paired tumour (primary and metastatic) and normal tissues were derived from the same patient.

Whole-genome and WES

Prior to the library construction, 2–3 μg of genomic DNA from each sample was fragmented using a Covaris sonication system to mean sizes of ~500 bp. After fragmentation, libraries were constructed according to the IlluNova Paired-End protocol. Briefly, the purified, randomly fragmented DNA was treated with a mix of T4 DNA polymerase, Klenow fragments, T4 polynucleotide kinase and a nucleotide triphosphate mix to repair the ends by blunting and phosphorylation. The Blunted DNA fragments were subsequently 3’-adenylated using the Klenow fragment (3’-5’exo) and ligated by T4 DNA ligase to BGI-designed PE Index Adapters that had been synthesized with 5’-methylcytosine in place of cytosine. After each step, the DNA was purified using the QIAquick PCR Purification Kit (Qiagen). For exome sequencing, 3 μg genomic DNA from each sample was fragmented and ligated PE Index Adapters then captured with SureSelect Human All Exon 50 Mb Kit as described in the protocol for the SureSelect Target Enrichment System for an Illumina Paired-End Sequencing Library. All the constructed libraries were sequenced on HiSeq 2000 platform using 2 × 100-bp paired-end reads.

Validation of somatic mutations

The raw lists of somatic indels were detected by Platypus20 based on default parameter. After this process, germline variants could be effectively removed. Identified mutations were validated either by MassARRAY or by target region captured resequencing. For mass spectrum resequencing, specific primers were designed for PCR amplification and base extension that covered the mutation sites. Genotyping assay and base calling of successfully amplified products were performed on the MassArray platform of Sequenom. Ultra-deep target region sequencing on a set of 51 selected genes was used to evaluate the validation status of identified SNPs and indels. One microgram of genomic DNA from each sample for validation was used for hybrid capture and library construction. Libraries were then sequenced on HiSeq 2500 platform using 2 × 100-bp paired-end reads sequencing.
Mutational signature analysis. Mutational signature analysis was performed as follows: briefly, the computational framework for mutational signatures deciphering was downloaded from http://www.mathworks.com/matlabcentral/fileexchange/38724. Mutational signatures were extracted from the WGS data of 24 primary tumours and corresponding metastasis tumours as follow: (1) Somatic base substitutions of each data set were classified into 96 possible mutated trinucleotides from 4 types of substitutions (C:G→A:T, C:G→T:A, T:A→G:C, T:A→C:G) × 4 types of 3′ base (A, C, G, T) × 4 types of 5′ base (A, C, G, T), to generate mutational catalogue. Then the prevalence of somatic mutations was calculated for each type of substitution. (2) Signatures of 15 mutational processes from the mutational catalogue were deciphered by the computational framework for mutational signature deciphering. The number of signatures extracted (N) is determined by estimating the signature reproducibility and reconstruction error rate and as described by Alexandrov et al.60. N here is the number where the lowest reconstruction error is achieved without decreasing the reproducibility. (3) The minimal set of mutational signatures was then determined to optimally explain the proportion of each mutation type found in the catalogue, based on reproducibility of their signatures and low error for reconstructing the original catalogue. Somatic mutations of WGS from 22 European lung adenocarcinomas were downloaded from ftp://ftp.sanger.ac.uk/pub/cancer/AlexandrovEtAl.Pearson correlation analysis was performed to compare signature patterns between primary and metastasis tumours in this study and those of European lung adenocarcinomas. Contributions of distinct signatures in individual tumour were also profiled and Pearson correlation analysis was performed to compare the signature contributions between paired primary and metastasis tumours13,60.

Identification of potentially significantly mutated genes. Considering the high background mutation rate in lung adenocarcinoma, analysis of significance of mutated genes was performed through two statistical procedures. First, a cancer driver gene prediction model23 was used to calculate the significance of detected mutations for each mutated gene. The mutation prevalence and functional impact was evaluated by this approach, and a background distribution of mutation score was assigned to each gene then. Taking together the background distribution of mutational score and test statistics of observed mutation scores across samples, a P value was computed for each gene and adjusted using the Benjamini–Hochberg method to further determine the false discovery rate (FDR). Only genes with FDR (q-value) <0.1 were selected for following analysis. A subsequent prediction of significantly mutated genes was performed using the MutSig2 algorithm and genes with FDR q-value <0.1 were regarded as significant.

Transcriptome sequencing. About 20 μg total RNA from each sample was treated with RNase-free DNase I for 30 min at 37°C (New England BioLabs) to remove residual DNA. Beads with oligo (dT) were used to isolate poly (A) mRNA. First-strand complementary DNA (cDNA) was synthesized using random hexamer-primer and reverse transcriptase (Invitrogen). The second-strand cDNA was synthesized using RNase H (Invitrogen) and DNA polymerase I (New England BioLabs). Then the cDNA libraries were prepared following the manufacturer’s instructions and were sequenced on the HiSeq 2000 platform.

The raw reads filtering process include three criteria: (1) Remove reads with sequence adaptors; (2) Remove reads in which unknown bases are >10%; (3) Remove low-quality reads, which have >50% in one read. All subsequent analyses were based on clean reads.

Mapping clean reads to the human genome and transcriptome, the reference sequences used were genome and transcriptome sequences downloaded from the UCSC website (version hg19). The Clean data were respectively aligned to the reference genome and transcriptome by SOAP2. Less than five mismatches were allowed for each read in the alignment.

The gene expression level is calculated by using RPKM method (reads per kb transcriptome per million mapped reads), and the formula is shown as follows:

\[
\text{RPKM} = \frac{10^{6} \times C}{N \times L}
\]

In which C is the number of reads that are uniquely mapped to the given gene, N stands for the total number of reads that are uniquely aligned to all genes and L is the total length of exons from the given gene.

Identification and validation of fusion genes. Soapfuse61 was used to identify fusion genes between paired tumour and normal samples from transcriptome sequencing using the default parameters. Reads were mapped against the human reference genome sequence (hg19) and annotated transcript. SOAPfuse then detected two kinds of reads to identify potential fusion events: spliced reads that were defined as paired-end reads mapping to two distinct genes and junction reads that covered the exact junction sites, after filtering duplicated reads a candidate fusion event was determined if more than two reads support the paired-end connection and junction site. WGS data of samples harbouring fusion genes were also analysed to confirm fusions that were generated by somatic rearrangements with breakpoints in intron or exon regions. Confirmed fusion events were validated at both RNA and DNA level by PCR and Sanger sequencing.

Concurrence and mutual exclusion analysis. Concurrence and mutual exclusion analysis was performed on significantly mutated genes, significant copy number variations and translocations or fusion events involving ALK, as well as clinical features such as smoking, age and metastasis status. Significance was calculated by Fisher’s exact test.

Pathway and network analysis. Pathway enrichment analysis was carried out by using WebGestalt62 to investigate the distribution of genes affected by somatic mutations, CNAs and genomic rearrangements within the KEGG database. To inquire the potential interaction network of altered genes, Cytoscape (V3.1.1) software63 (http://www.cytoscape.org/) was used to identify pathways and network patterns related to cancer through the reactome functional interaction plugin64.

Real-time RT-PCR assay. Total RNAs of lung cancer cells and lung adenocarcinoma specimens were extracted using Trizol reagent (Invitrogen). The quantity and quality of RNA were determined by agarose gel electrophoresis and photometry. One microgram of RNA was used for cDNA synthesis using the RT System (Promega) according to the manufacturer’s instructions. Real-time PCR was performed using GoTaq qPCR Master Mix (Promega). Thermal cycling conditions were as follow: 1 step for 10 min at 95°C followed by 40 cycles at 95°C for 15 s and at 60°C for 1 min. The quality of the PCR products was monitored using post-PCR melting curve analysis. Relative mRNA levels were calculated with ΔΔCT methods using β-actin for normalization.

Immunohistochemistry. Immunohistochemistry analysis of IQGAP3 in the primary tumours of lung adenocarcinoma was carried out using a standard avidin–biotin–peroxidase technique. In brief, the formalin-fixed, paraffin-embedded tissue sections were deparaffinized and rehydrated through xylene, graded ethanol and water.

The endogenous peroxidase activity was blocked by incubating the sections in 3% hydrogen peroxide for 15 min, and heat-induced epitope retrieval was performed by incubating the sections in boiling sodium citrate buffer (10 mM sodium citrate, 0.05% Tween-20, pH 6.0) for 20 min in microwave. Then the sections were blocked with 10% normal goat serum in PBS for 10 min at room temperature, and incubated with primary antibody against IQGAP3 (1:80, ab151437, Abcam, USA) diluted in 1% BSA (ZhongshanJinqiao, Beijing, China) overnight at 4°C. The Maxin Detection System Peroxidase/DAF, Rabbit/Mouse kit was applied to determine immunoreactivity according to the manufacturer’s instructions, with 1:50 dilutions of DAB Chromogen to HRP Substrate Buffer. Sections were then washed, and counterstained for 30 s in filtered Harris Haematoxylin (ZhongshanJinqiao). After washing with PBS and dehydration with graded ethanol followed by xylene, the sections were mounted in neutral balsam (ZhongshanJinqiao). The slides were viewed and images were captured using Nikon Eclipse 80i microscope with NIS-Elements F software. Analysis of cytoplasmic staining was performed by two pathologists independently through measuring the percentage of positively stained tumour cells. A consensus was reached for intensity by taking the mean if scores were within 20% of each other; otherwise, images were reviewed and a consensus reached.

Statistical analysis for clinicopathological data. The correlation of IQGAP3 expression with clinicopathological characteristics was determined by Pearson’s χ2-test. The differences of IQGAP3 mRNA and protein between tumours and normal tissues were tested by student’s t-test and Wilcoxin signed rank test, respectively. Overall survival and disease-free survival were estimated using Kaplan–Meier method with log rank test. The prognostic significance of IQGAP3 was evaluated by Cox regression model with forward likelihood ratio (LR) method. All statistical analyses were performed using GraphPad Prism 5 and SPSS16.0.

References
1. Jeral, A. et al. Global cancer statistics. CA Cancer J. Clin. 61, 69–90 (2011).
2. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. CA Cancer J. Clin. 63, 11–30 (2013).
3. Herbst, R. S., Heymach, J. V. & Lippman, S. M. Lung cancer. N. Engl. J. Med. 359, 1367–1380 (2008).
4. Iemielski, M. et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150, 1107–1120 (2012).
5. Pao, W. & Girard, N. New driver mutations in non-small-cell lung cancer. Lancet Oncol. 12, 175–180 (2011).
6. Pao, W. & Hutchinson, K. E. Chipping away at the lung cancer genome. Cell 1107–1120 (2012).
7. Weir, B. A. Characterizing the cancer genome in lung adenocarcinoma. Science 351, 251–256 (2014).
8. Weir, B. A. et al. Characterizing the cancer genome in lung adenocarcinoma. Nature 450, 893–898 (2007).
37. Beroukhim, R. et al. Prevalence and heterogeneity of KRAS, BRAF, and PIK3CA mutations in primary colorectal adenocarcinomas and their corresponding metastases. Cancer Res. 66, 790–799 (2010).

38. Jackson, E. L. et al. Association of fascin-1 with mortality, disease progression and metastasis in carcinomas: a systematic review and meta-analysis. J. Pathol. 232, 144–152 (2014).

39. Briggs, M. W. & Sacks, D. B. IQGAP proteins are integral components of cytoskeletal regulation. EMBO Rep. 4, 571–574 (2003).

40. White, C. D., Brown, M. D. & Sacks, D. B. IQGAPs in cancer: a family of scaffold proteins underlying tumorigenesis. FEBS Lett. 583, 1817–1824 (2009).

41. Tan, V. Y., Lewis, S. J., Adams, J. C. & Martin, R. M. Association of fascin-1 with mortality, disease progression and metastasis in carcinomas: a systematic review and meta-analysis. J. Pathol. 232, 144–152 (2014).

42. Park, S. H. et al. Fascin1 expression in high-grade serous ovarian carcinoma is a prognostic marker and knockdown of fascin1 suppresses the proliferation of ovarian cancer cells. Int. J. Oncol. 44, 637–646 (2014).

43. Al-Alwan, M. et al. Fascin is a key regulator of breast cancer invasion that acts via the modification of metastasis-associated molecules. PLoS ONE 6, e27339 (2011).

44. Schmid, K. et al. EGFR/KRAS/RAF mutations in primary lung adenocarcinomas and corresponding loco-regional lymph node metastases. Clin. Cancer Res. 15, 4554–4560 (2009).

45. Monaco, S. et al. A comparison of EGFR and KRAS status in primary lung carcinoma and matched metastases. Hum. Pathol. 41, 94–102 (2010).

46. Jackson, E. L. et al. Analysis of lung tumour initiation and progression using conditional expression of oncogenic K-ras. Genes Dev. 15, 3243–3248 (2001).

47. Meuwissen, R., Linn, S. C., van der Valk, M., Mooi, W. J. & Berns, A. Mouse model for lung tumorigenesis through Cre/lox controlled sporadic activation of the K-Ras oncogene. Oncogene. 20, 6551–6558 (2001).

48. Baldus, S. E. et al. Prevalence and heterogeneity of KRAS, BRAF, and PIK3CA mutations in primary colorectal adenocarcinomas and their corresponding metastases. Clin. Cancer Res. 16, 790–799 (2010).

49. Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505, 495–501 (2014).

50. Adorno, M. et al. A Mutant-p53/Smad complex opposes p63 to empower TGFbeta-induced metastasis. Cell 137, 87–96 (2009).

51. Muller, P. A. et al. Mutant p53 drives invasion by promoting integrin recycling. Cell 139, 1327–1341 (2009).

52. Morton, J. P. et al. Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. Nat. Proc. Acad. Sci. USA 107, 246–251 (2010).

53. Weissmueller, S. et al. Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor beta signalling. Cell 157, 382–394 (2014).

54. Muller, P. A. & Vouken, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. Cancer Cell 25, 304–317 (2014).

55. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 27, 15–25 (2011).

56. Wang, J. et al. SOAP2: an alternative assembler for short pair-ended DNA sequencing reads. Genome Res. 20, 1297–1303 (2010).

57. Li, J. et al. CONTRA: copy number analysis for targeted resequencing. Bioinformatics 28, 1307–1313 (2012).

58. Wang, J. et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat. Methods 8, 652–654 (2011).

59. Wang, K. L., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).

60. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 3, 248–259 (2013).

61. Jia, W. et al. SOAPFusion: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol. 14, R12 (2013).

62. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 33, W741–W748 (2005).

63. Shannon, P. et al. Cytoscape: a software environment for integrating models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).

64. Croft, D. et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 39, D691–D697 (2011).
sequencing data. D.L., L.S., X.Y., N.A. and H.W. analysed the whole-exome sequencing data. D.L., F.L., J.Y. and N.L. analysed the mRNA data. F.L. performed mutational signature analysis. F.L., D.L., N.L., Y.H. and S.Z. performed significantly mutated gene, copy number alteration, translocation and fusion analyses. K.W. and N.L. performed pathway and network analyses. D.X. and L.P. performed clinical association analysis. M.Y., L.I. and J.L. provided assistance in genomic data analysis. D.X. designed the functional experiments. L.H., W.L., H.H., M.T. and J.V. carried out the functional experiments. Q.D. performed immunohistochemistry analysis. F.L., D.L., K.W., N.L., L.P., H.L., Y.Z. and G.X. constructed the figures and tables. Y.H., S.Z. and X.Y. contributed bioinformatics support and scientific advice. M.Y. and Z.P. provided administration assistance. F.M., Xiuqing Z., X.X., H.Y., K.K., Jian W. and N.Z. provided organizational support. K.W. wrote the manuscript. Q.P.-H., K.W., D.X., Y.H. and S.Z. revised the manuscript.

Additional information
Accession code: All the sequencing data of 335 lung adenocarcinomas have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted by the EBI, under the accession code EGAS00001000982.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

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

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wu, K. et al. Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas. Nat. Commun. 6:10131 doi: 10.1038/ncomms10131 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/