Genomic Alteration During Metastasis of Lung Adenocarcinoma

Qiang Tan\textsuperscript{a} Jian Cui\textsuperscript{b} Jia Huang\textsuperscript{a} Zhengping Ding\textsuperscript{a} Hao Lin\textsuperscript{a} Xiaomin Niu\textsuperscript{a} Zhiming Li\textsuperscript{a} Guan Wang\textsuperscript{b} Qingquan Luo\textsuperscript{a} Shun Lu\textsuperscript{a}

\textsuperscript{a}Department of Shanghai Lung Tumor Clinic Center, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, \textsuperscript{b}BioGenius Biotechnology (Shanghai) CO., LTD, Shanghai, People's Republic of China

Key Words
Genome • Lung • Adenocarcinoma

Abstract

\textbf{Background/Aims:} Recurrent gene mutation has been identified by the analysis of exonic DNA from lung adenocarcinoma, but its progression has not been extensively profiled. The investigation of the mutational landscape of tumors provides new insights into cancer genome evolution and further discovers the interplay of somatic mutation, adaptation of clones to their environment and natural selection. Cancer development involves cycles of genomic damage, epigenetic deregulation, and increased cellular proliferation that eventually culminate in the carcinoma phenotype.

\textbf{Methods:} Comparative whole exome sequencing of both primary and metastatic tumor tissues from four patients of stage IV lung adenocarcinoma patients with chest wall metastasis was performed. Both primary and metastatic tumors were diagnosed through biopsy followed by surgical resection. All tumor specimens were cut into several pieces to assess potential heterogenic clones within the tumor tissue. Adjacent normal lung tissue was also obtained to provide germline mutation background.

\textbf{Results:} By modeling and analyzing progression of the cancer metastasis based on non-synonymous variants, we defined the extent of heterogeneity of cancer genomes and identified similar cancer evolution pattern in the four patients: metastasis was an early event occurring right after the primary cancer formation and evolution in the metastatic tumor was continuously and simultaneously in progress with that in the primary tumor. By characterizing the clonal hierarchy of genetic lesions, we further charted a pathway of oncogenic events along which genes may drive lung adenocarcinoma metastasis, such as TAS2R31 and UMODL1, involving in G-protein coupled receptor protein signaling pathway.

\textbf{Conclusion:} The candidate genes identified in this study may become targets for the treatment of lung adenocarcinoma metastasis.

Q. Tan and J. Cui contributed equally to this paper.

Dr Qingquan Luo
Department of Shanghai Lung Tumor Clinic Center, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, 241, Huaihai West Road, Shanghai, 200030, (PR China); Tel. +86-13601860323, E-Mail luoqingquan\_1980@163.com
Introduction

Lung cancer is currently the leading cause of cancer-related death worldwide, accounting for approximately a third of all cancer diagnoses and death. The high mortality rate associated with lung cancer has prompted numerous exhaustive efforts to identify novel therapeutic targets and treatment modalities for this lethal disease [1]. Most lung cancer patients died of metastasis, especially lung adenocarcinoma, which often metastasizes when the primary tumor size is still quite small. The molecular mechanism underlying the early metastasis of lung adenocarcinoma has been instrumental in developing novel therapeutic strategies to target this aggressive metastatic disease. It is necessary to further investigate the molecular determinants of treatment resistance and to develop novel therapeutic strategies†…‡…–‡† •‘‡…‹ϐ‹…ƒŽŽ› ƒ‰ƒ‹–– –Ї ’އ…—Žƒ“ †“‹˜‡“•’ˆ ‡–ƒ•–ƒ–‹… Ž—‰ …ƒ…‡” ȏʹǦͶȐǤ 
‹˜‡ the heterogeneity and convergence of signaling pathways underlying lung adenocarcinoma [5, 6], efforts are needed to improve a better understanding of the molecular origins and evolution of it in the future.

Previous studies in lung adenocarcinoma genome focuse d on mutations for targeted chemotherapy, recurrence of genetic mutations, genome-wide structure variants and as targeted chemotherapy for lung cancer treatment [7, 8]. In lung adenocarcinoma, KRAS, BRAF, ERBB2 and PIK3CA or translocations in RET and ROS1 are found as frequently recurrent and loss-of-function mutations, or deletions in tumor suppressor genes TP53, STK11, RB1, NF1, CDKN2A, SMARCA4 and KEAP1 [9-11]. Recently, many array-based profiling [12], targeted re-sequencing [12, 13], and next generation sequencing-based exome and/or whole genome sequencing data [14, 15] have all characterized the special targeted mutation genes or structure variants in lung adenocarcinoma genomes. A recent study describing whole-exome sequencing of 16 lung adenocarcinomas enumerated several mutated genes but did not identify genes undergoing positive selection for mutation in the studied tumors [16]. Using arrays profiling, somatic focal amplifications of NKF21 (NK2 homeobox 1) was identified in lung adenocarcinomas [12]. In another systemic characterization of somatic mutations, it was found that 26 genes out of 623 genes in 188 human lung adenocarcinomas, including ERBB4, EPHA3, KDR and NTRK, were in mutation with significantly high frequencies and thus were considered to be probably involved in carcinogenesis [9]. Meanwhile, somatic mutations of several tumor suppressor genes, including NF1, APC, RB1, ATM, PTPRD, and LRP1B, are considered as contributing to the development of primary lung adenocarcinoma and other cancers. A transforming KIF5B and RET gene fusion were found in lung adenocarcinoma by whole-genome and transcriptome sequencing [13]. A study on exome and genome sequences of 183 lung adenocarcinoma (tumor/normal DNA pairs) revealed that recurrent somatic mutations occurred in the splicing factor gene U2AF1, truncating mutations affected RBM10 and ARID1A, and the mutation signatures were correlated with smoking history [16]. Unfortunately, such recurrent alterations in genes or transcriptome are difficult to exploit the molecular basis of lung cancer metastasis. Therefore, knowledge of additional genes altered in lung adenocarcinoma metastasis is needed to further guide diagnosis and treatment.

Recently, increasing number of somatic mutations in variety kinds of cancers have been found with the development of sequencing technique. The historically proven approach of understanding evolution is the comparative analysis of extant species, whose power was greatly increased by whole genome sequencing in recent years. The concept of clonal evolution in cancer was originally proposed by Peter Nowell, who theorized that acquired genetic instability increases as a neoplasm progress, resulting in heterogeneity [17]. Consequently, numerous genetically distinct subclones, which developed within the neoplasm, lead to intra-tumor diversity. As sublines evolve from the original primary clone, a Darwinian selection process occurs to allow the cancer to become more malignant [18]. This "survival of the fittest" within cancers, and cancer cell heterogeneity in general, have been studied with increasing sensitivity as technology has developed [19]. Based on the
evolutionary lineage trees, comparisons of somatic genomes from a single individual could shed light on somatic evolution in principle. In this regard, given the large number of genomic changes during tumor evolution, it may be possible to dissect the evolutionary history of a cancer by comparing its genome to clinically recognized precursor lesions.

In the current study, therefore, using next-generation sequencing of the whole exomes of DNA from 4 lung adenocarcinomas with seven multiple regions for each subject, including 1 matched normal adjacent tissue, 3 pieces from primary adenocarcinomas section and 3 pieces from metastasis adenocarcinomas section, we not only verified the genes with frequent somatic alterations that has been reported in the previous studies of lung adenocarcinoma, but also determined novel mutation genes in total 28 exome sequencing data that likely contribute to metastatic evolution; confirmed lineage relationship of primary and metastatic tumor; quantified mutational load and mutation spectra during progression from normal tissue to primary tumor, and to metastatic tumor; and found the earliest detectable mutations. Findings of the current analyses revealed that variation exist among the four cases in the specific evolution of primary and metastatic tumor. The mechanistic commonalities among the cases, however, bear significant implications for our conceptualization of lung adenocarcinoma origins and progression.

Materials and Methods

Patient and Sample Characteristics

Primary lung adenocarcinomas (4 patients), metastatic lung adenocarcinomas and adjacent normal tissue specimens were obtained from surgically resected lung tissue. Both cancerous and normal tissue specimens were grossly dissected and preserved immediately in liquid nitrogen. Each specimen of lung adenocarcinoma was segregated into six parts, that is, three primary adenocarcinomas and three metastatic adenocarcinomas. DNA was extracted from each segregated tumor tissue and adjacent normal lung tissue. Protocol of tissue collection was approved by the Ethic Committee of School of Medicine, Shanghai Jiao Tong University, and a written consent was obtained from each patient. Samples estimated to contain more than 80% tumor cells were used. SNP arrays were conducted in all samples to demonstrate that they were from the same individual and to confirm the presence of copy number changes.

Massively Parallel Sequencing and Analysis

Whole-exome sequencing of paired tumor/normal DNA samples from the 4 patients was performed. Briefly, DNA libraries were prepared according to the Illumina library generation protocol version 2.3. Each sample was tagged with a custom-designed unique 4-base-long index within the Illumina adaptor, pooled, captured by TruSeq Exome Enrichment Kit and sequenced using Illumina Hiseq2000 system. All samples were co-sequenced together within the Genery Genomic Sequencing Center to an average coverage of 48.5 x for each sample, with 74.60 % of targeted bases covered at ≥ 20 × and 86.09 % of targeted bases covered at ≥10 × per base position. Median 94.47 % of the exonic regions were covered at average ≥ 10 × in all twenty-eight samples. Paired-end sequencing was carried out for 100 bases from each end of around 200-bp insert fragment libraries using standard Illumina protocols, and each read was aligned to hg19 (Feb.2009 GRCh37/hg19) from UCSC Genome Browser using BWA (Burrows-Wheeler Aligner) with default parameters [20]. PCR duplicates with the same start site for both ends were removed using picard tools. Variants were called using GATK (The Genome Analysis Toolkit 1.6) tools and VQSR (Variant quality score recalibration) method, which builds an adaptive error model using known variant sites, and it was applied to estimate the probability that each variant is a true genetic variant or a machine artifact [21]. The variants were annotated using the annovar (Functional annotation of genetic variants from high-throughput sequencing data) [22] and all annotation databases were downloaded from the UCSC Genome Browser [23]. Last, the somatic variants from each paired samples with the following criteria [24] were identified: (1) SNP and indels quality score ≥ 255; (2) Genotype score as a Phred-scaled confidence at the true genotype >= 40.0; (3) The number of read depth which read’s mapping quality >= 17 be calculated, >= 20. (4) Variants between tumor and normal were different (tumor contains more variant alleles than normal) based on the genotypes called by GATK and aforementioned filtering criteria (somatic mutations: 0/1 in tumor; 0/0 in normal; 1/1
in tumor, and 0/1 in normal, or 1/1 in tumor, 0/0 in normal). Variants were further filtered by examining whether they were within the protein coding sequence as defined by RefSeq gene models, and not listed in dbSNP132.

**Evolution analysis**

All non-synonymous variants passed the filtering in exonic regions were assembled across all specimens from the same patient and was then conducted the phylogenetic tree using neighbor-joining algorithm in ClustalW to construct trees from the distance matrix. Visualization of the tree was performed under the software Phylovieer. In each phylogenetic tree, the nested structure of variants defined a phylogenetic tree of relationships among the metastases and primary tumor. The length of black lines is proportional to the genetic distance between nodes.

The GO-enriched analysis was done by DAVID (https://david.ncifcrf.gov/gene2gene.jsp) using all confident somatic genes. (Cutoff of enriched GO: $P <= 1.00e-3$). We extended the traditional enrichment analysis logic so that a gene group was more important if the majority of its gene members were associated with highly enriched annotation terms as found in the traditional enrichment analysis of the total gene list. Thus, the enrichment score of each group was measured by the geometric mean of the EASE Scores (modified Fisher Exact) associated with the enriched annotation terms that belonged to this gene group. Importantly, the multiple testing correction issues were considered in the individual EASE scores, and all EASE scores (significant or insignificant) associated with the group participation in the algorithm. In order to emphasize that the geometric mean was a relative score rather than an absolute $p$ value, minus log transformation was applied in the geometric mean (Additional data file 6). Therefore, the group enrichment scores were intended to order the relative importance of the gene groups instead of a absolute decision values. A higher score for a group indicates that the group members were involved in more important (enriched) roles. However, all gene groups were potentially interesting despite lower rankings.

**Statistical Analyses**

The hierarchical clustering was performed on the somatic mutations on gene level for each patient using R scripts. All figures were output as JPEG format using illustrator software.

**Results**

**Exome sequence analysis of lung adenocarcinoma**

We examined four primary lung adenocarcinoma and their chest wall metastasis nodules (S183604, S183744, S183948, S183138) in comparison to their matched normal tissues by paired-end massively parallel sequencing technology [25]. In order to identify heterogeneity of the tumor and to reconstruct the path of the cancer metastasis, both primary tumor and metastatic nodules were cut into 3 pieces. The primary lung adenocarcinoma, metastatic nodule and corresponding normal lung tissue were sequenced to a median 5727.379 Mb data for each sample. Patients’ clinical and histopathological information was presented in the Table 1. Sequence characteristics, and major variants for each patient in the study were presented in the Table 2. Exome was sequenced to median fold coverage of 48.5 on 62 Mb of target sequence with 61.77% capture efficiency (Table 3). Sequencing variances across the
four patients was not statistically significant because all patients contained similar exome capture efficiency and sequencing depth according to the total data yields. Otherwise, exome was sequenced to a median coverage of 48 X in the normal control, 49X in the primary lung adenocarcinoma, and 48 X in the metastatic specimen.

**Somatic mutations in lung adenocarcinoma**

Somatic mutations and small insertions or deletions (indels) through statistical comparison of paired primary and metastatic tumor versus normal sequence data were identified by using algorithms calibrated for heterogeneous cancer tissues. Officially targeted regions of the 28 specimens (including normal tissue) contained 542,064 SNP variants in average, and 86.4% SNP variants were annotated in dbSNP database and 76% in 1000G. Exonic regions in 7 specimens from the 4 patients (S183138, S183604, S183744 and S183948, respectively) contained 33294, 33457, 34345, and 33429 SNP variants (Mean 33631), in which non-synonymous SNP number is 13268, 13110, 13480 and 13291, respectively. As the indels, high quality of 26938, 26191, 29243
and 26165 indels were called in meantime. Exonic indels in 7 of each patient were 12, 209, 25 and 616 for each subject (Table 4).

Using matched normal as control and remaining exonic or splicing non-synonymous variants that changed the protein amino acid sequence, the S183138 patient had 277 somatic variants corresponding to 166 somatic mutation genes and a mean of 4.47 mutations/Mb; the S183604 patient had 319 somatic variants (SNP) in 133 genes with 5.14 mutations/Mb rate; the S183744 patient had 361 somatic variants in 125 genes corresponding to 5.82 mutations/Mb; and the fourth patient had 317 somatic variants in 124 genes corresponding
to 5.11 mutations/Mb. The somatic point mutations were mapped onto their regional distribution across the tumors and shown in Fig. 1-4.

The somatic mutation genes found in the current study were compared with that described in the latest COSMIC database (version v64) ([http://www.sanger.ac.uk/genetics/CGP/cosmic](http://www.sanger.ac.uk/genetics/CGP/cosmic)) [26], which includes study of multiple lung adenocarcinomas. Of the 21,030 total driver and passenger mutated genes, 9,698 genes had more than 1% recurrence and 870 genes had more than 5% recurrence. Of 419 somatically mutated genes found in all four specimens, 63 (15.04%) were overlapped with the top 5% recurrent variants and 297 (70.88%) were overlapped with the top 1% recurrent variants in COSMIC. In addition, compared to another study of 183 Whole-genome/Exome sequencing of lung adenocarcinomas [16], 362 (86.40%) somatically mutated genes found in the current study were among the 14,760 somatically mutated genes published in the literature (14), including TP53, MUC16, and USH2A (> 30% mutated frequency reported in COSMIC, too). These findings suggested that lung adenocarcinoma have most of consistency with knowledge about recurrent genes in lung adenocarcinoma, and the data quality was normal and passed for our further analysis. Although it was highly consistent between our results and that of others, we further explored the driver somatic genes in cancer clonal evolution, which may contribute to the cancer metastasis, and to focus on the recurrent genes.

**Heterogeneity of the mutational landscape and branched evolution within an individual cancer**

Data was analyzed from the whole exome sequencing on multiple spatially separated samples of the primary and metastatic biopsies. The results showed that lung adenocarcinoma tissues were heterogeneous. Heterogeneity is central to cancer genome evolution, and we described this at the level of cancer genes and within individual patients. Total of 166 somatic non-synonymous point mutations were identified in exonic regions in the patient S183138. A metastasis specimen (183138M) contained very similar genetic and somatic variant pattern with a primary tumor specimen (183138T), when compared with the common normal control (183138L) (Fig. 1A). There were a total of 133, 125 and 124 somatic mutations found in patients S183604 (Fig. 2A), S183744 (Fig. 3A) and S183948 (Fig. 4A) across the primary or metastasis cancer, respectively.

To further elucidate the evolution of lung adenocarcinoma, we analyzed ancestral relationships and reconstructed a phylogenetic tree of every lung adenocarcinoma by ordering clonal sequence, following the previously described methods [27, 28]. The evolutionary tree revealed branched evolutionary tumor growth in all intra-tumors with their special patterns. Mutations in a primary and a metastatic tumor tissue have nested convergent event, which suggested a punctuated evolution event just after the primary tumor development. In analysis for four patients, branch lengths are proportional to the number of non-synonymous mutations separating the branching points (cancer local specimen with particular tumor developmental stage). Potential driver mutations were acquired by the indicated genes in the branch.

Next, according to the functional mis-sense mutations, we conducted phylogenetic analysis to reconstruct the evolution path. It was shown that this metastasis onset event occurred right after the primary tumor event. As known from clinical phenotypes, the primary and the metastasis specimen may have experienced punctuated evolutionary events. In contrast to conditional knowledge, the patient S183138 had different evolution pattern and early onset of metastasis (Fig. 1B). Phylogenetic analysis of other three patients revealed the similar branching pattern with punctuated evolution characters between primary and metastasis.

The 183604M and 183604T located in nearest distance from the normal control specimen (183604L) (Fig. 2B). Another two metastasis specimens 183604m-1 and 183604m-2 harbored branches near the two primary specimens: 183604-1 and 183604-2. The 183744M and 183744T had converged and punctuated evolution from 183744L (Fig. 3B). Primary tumor, 183744-2 and 183744-1, and metastasis tumor, 183744m-1 and
Fig. 1. Genetic intra-tumor heterogeneity and phylogeny in the patient S183138. (A) Heat map showing regional distribution of non-synonymous mutations detected in the whole exome sequencing of S183138. Primary lung adenocarcinoma regions of the specimen T, t1 and t2, metastasis regions of the specimen M, m1 and m2 were detected by exome sequencing. The green indicated presence of a mutation and the white indicated absence of a mutation. Gene name was shown in the bottom. (B) Phylogenetic relationships of the tumor regions. Branch lengths are proportional to the number of non-synonymous mutations separating the branching points. Potential driver mutations were acquired by the indicated genes in the branch.

183744m-2, were converged enduring more accumulation of mutations and harboring long distances from 183744M and 183744T loci. Phylogenetic tree also revealed similar punctuated evolution for metastasis and primary tumor. The 183948M and 183948T were converged, and the other two primary (183948t-1) or two metastasis loci (183948t1 and t2) were converged in another branch direction and seemed to extend from the former primary (183948T) and metastasis (183948M) clones (Fig. 4B).

_Heterogeneity drives mutation of genes in cancer metastasis_

To examine which key mutation genes is the driver force to initialize metastasis, we compared the nearest primary tumor and metastasis lesions with the normal control using phylogenetic analysis. In the case of S183138, somatic non-synonymous point mutations in exonic regions of metastasis tumor (183138M) and primary tumor (183138T) were compared. It was found that NT5C3 and FAM194B were the only genes, which were specifically occurred in the metastasis specimen 183138M and at least one of two metastasis loci. The NT5C3 gene recurrently appeared in 183138m-2 specimens. From clustering analysis of non-synonymous point mutations, we also identified 183138M and 183138m-2 were nearest in distance.

In the case of S183604, FAM194B and PRG4 were detected as driver genes for tumor metastasis. PRG4 (proteoglycan 4), or known as migration-stimulating factor (MSF) recurrently shown in two branched primary loci (183604-1 and 183604-2) and metastasis loci (183604m-1 and 183604m-2), not in 183604T.

In the case of S183744, we detected eight genes including PABPC1, CYLC2, KRTAP4-5,
Highly recurrent somatic mutation genes in primary tumor

In order to further determine the driver and passenger mutation genes in different patients, frequency of recurrent somatic mutation genes in primary or metastatic tumors were counted and sorted. The frequency of recurrent somatic mutation genes in primary or metastasis was defined as the ratio of the frequency in primary/metastasis tumor and the total number of the tested specimens in particular group (for example, 12 primary and 12 metastasis in total, respectively). We identified 329 somatically mutated genes correlated with primary tumors with 8.33%~75% frequency in 12 specimens from the 4 subjects, and 298 somatic mutated genes correlated with the metastasis tumor progression with 8.33~83.33% frequency in other 12 specimens from the 4 subjects. Of total 329 somatic primary-associated mutated genes, 208 genes were recurrently shown in metastasis. A total of 14 genes had more than 50% frequency, including MUC4, PRG4, KRTAP5-7, IGFN1, KRTAP9-1, MUC5B, PRR21, CHIT1, ZNF141, MYO5B, IVL, NACAD, NEFH, and HOMEZ.

Recurrent mutation genes for secondary metastasis

In order to understand differences of the genetic architecture between any primary and any metastasis, using somatic mutation pattern of the four patients, we counted frequency
of the recurrently mutated genes, which was occurred only in the metastatic tumors. It was found that total of 113 genes were detected as driver genes from global scale via comparing primary and metastasis tumor, and that the genes TAS2R31 and UMODL1 had 50 % recurrent frequency (Table 5).

In the aforementioned 113 metastasis-potentially somatic mutation genes, we used functional enrichment analysis and found that most of the mutated genes function in the Olfactory transduction pathway (p=0.037) and Gene Ontology. Analysis revealed that these genes involve in functions like sensory perception of chemical stimulus, sensory perception, inositol or phosphatidylinositol phosphatase activity, G-protein coupled receptor protein signaling pathway, and phosphatase activity.

**Discussion**

Whole exome sequencing and integrative analysis of genomic data provided the mutational landscape of tumors and new insights into cancer genome evolution, laying bare the interplay of somatic mutations, adaptation of clones to their environment and natural selection. Multiregional genetic analysis of four consecutive tumors provided evidence of intra-tumor heterogeneity in every lung adenocarcinoma, with spatially separated heterogeneous. Heterogeneous somatic mutations led to phenotypic intra-tumor diversity activating mutation and uniformity. Of all somatic mutations found on multiregional sequencing, most of them were heterogeneous and thus not detectable in every sequenced
region. In the metastasis lesions of every lung adenocarcinoma, it was also showed that high heterogeneity due to the different patterns of somatic mutations to drive cancer metastasis. Evolution is the result of natural selection, so is the progression of cancer. In this study, we updated the understanding of the dynamic evolution of the cancer genome in a lung adenocarcinoma. The implications for lung adenocarcinoma metastasis is counterintuitive that most of metastasis occur early after the primary tumor with punctuated mode in our study. Potential driver mutations were acquired by the indicated.

**Table 5.** Functional enrichment for the highly frequent somatic mutated genes in the primary tumor

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>%</th>
<th>PValue</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0005882—Intermediate filament</td>
<td>3</td>
<td>21.43</td>
<td>5.39E-03</td>
<td>KRTAP5-7, KRTAP9-1, NEFH</td>
</tr>
<tr>
<td>GO:0045111—Intermediate filament cytoskeleton</td>
<td>3</td>
<td>21.43</td>
<td>5.63E-03</td>
<td>KRTAP5-7, KRTAP9-1, NEFH</td>
</tr>
<tr>
<td>GO:0005856—Cytoskeleton</td>
<td>5</td>
<td>35.71</td>
<td>6.64E-03</td>
<td>KRTAP5-7, KRTAP9-1, NEFH, IVL, MYOSB</td>
</tr>
<tr>
<td>Signal transduction mechanisms</td>
<td>2</td>
<td>14.29</td>
<td>1.63E-02</td>
<td>PBR21, MUC4</td>
</tr>
<tr>
<td>GO:0044430—Cytoskeletal part</td>
<td>4</td>
<td>28.57</td>
<td>1.74E-02</td>
<td>KRTAP5-7, KRTAP9-1, NEFH, MYOSB</td>
</tr>
<tr>
<td>GO:0005198—Structural molecule activity</td>
<td>3</td>
<td>21.43</td>
<td>4.52E-02</td>
<td>IVL, MUCSB, MUC4</td>
</tr>
<tr>
<td>GO:0005201—Extracellular matrix structural constituent</td>
<td>2</td>
<td>14.29</td>
<td>4.55E-02</td>
<td>MUCSB, MUC4</td>
</tr>
</tbody>
</table>
had occurred through intermediate stages. The detection of private mutations in specific regions suggested clonal evolution. Cancer metastasis occurs after long time progression of primary tumor and accumulation of serial mutations or epigenetic changes, such as renal-cell carcinoma [27], secondary acute myeloid leukemia [29, 30], relapsed acute myeloid leukemia [31] or acute myeloid leukemia evolved from pre-leukemic hematopoietic stem cells [32] and triple-negative breast cancers [33].

Primary and metastasis cancer can still evolved simultaneously in a prolonged time with clonity properties. Branched tumor evolution underscores the importance of targeting ubiquitous alterations in the trunk of the phylogenetic tree. Larger multiregional series will probably identify genes that can be targeted in the trunks of the phylogenetic tree for lung carcinoma. Intra-tumor heterogeneity within the primary tumor may account for ubiquitous alterations in the trunk of the phylogenetic tree. Larger multiregional series will probably identify genes that can be targeted in the trunks of the phylogenetic tree for lung carcinoma. Intra-tumor heterogeneity within the primary tumor may account for—considered as a biomarker of response to chemotherapy with cytidine analogues and genetic study has characterized genotypes in gemcitabine pathway genes have association with tumor cell survival and adaption in local tissue environment.

In summary, as gemcitabine pathway specifically regulates S-phase of the cell cycle [46], it suggests that, as a member gene in gemcitabine pathway, somatic mutation in NT5C3 may affect the regulation of checkpoint signaling pathways in cancer and thus may influence tumor cell survival and adaption in local tissue environment.

PRG4 has been reported in lung adenocarcinoma [16], as well as multiple cancers, like melanoma [47], prostate cancer [48], endometrial tumors [49], and medulloblastoma [50]. PRG4 is a novel human growth factor acting on the primitive cells of both hematopoietic and endothelial cell lineages, and stimulate endothelial cell proliferation through PI-3K/
Akt pathway [51]. It was also characterized as a factor in the fetal-like fibroblasts of breast cancer patients but not from normal adult cells, and cancer patient fibroblasts undergo certain transitions [52] and was presumably responsible for the characteristically elevated levels of migration [53]. In the mice immunized with human esophageal cancer endothelial cells (HESEC), PRG4-targeted treatment significantly suppressed tumor growth through inhibition of human tumor-related angiogenesis [54]. PRG4 is a new molecule associated with the M2 polarization of M phase and expressed by tumor-associated macrophages (TAMS). Its biological function may contribute to M-mediated promotion of cancer cell invasion and metastasis [55].

We detected eight genes in the patient S183744, including PABPC1, CYLC2, KRTAP4-5, FAM194B, IGFn1, MCF2, Muc4, and PAK2. These genes may contribute to the initial metastasis from the primary tumor 183744T, and represented intratumor heterogeneity in other primary and metastasis loci. Except for CYLC2, other 7 genes were also shown in at least one of the other two primary tumors (183744-1 or 183744-2). The mutation events for the above genes have been reported in the lung adenocarcinoma [16], leaving FAM194B gene alone. CYLC2 mutation was only shown in 183744M, but not in other primary and metastasis loci. It indicated that the tumor heterogeneity between each of the two metastasis loci or primary tumor loci. CYLC2 may be the driver mutation gene only for 183744M, but not for other metastasis tumor loci.

Most of these genes were reported in COSMIC and previous lung adenocarcinoma study [16], except for FOXH1 (forkhead box H1). However, function of these genes in metastasis is rarely known. FOXH1, one of FOX gene family protein, can binds SMAD2 and activates activin responsive element via binding the DNA motif and deregulation of FOX family genes may lead to carcinogenesis [56]. Consistently, it has been reported that Smad/FOXH1-mediated transcriptional activation may play a role in lung tumorigenesis [57].

This comprehensive integrative analysis of 4 lung adenocarcinoma/normal pairs provides a number of insights into the biology and identifies potential therapeutic targets. We demonstrated that 329 somatically mutated genes correlated with primary tumors with 8.33%~75% frequency in 12 specimens from the 4 subjects, and 298 somatic mutated genes correlated with the metastasis tumor progression with 8.33%~83.33% frequency in other 12 specimens from the 4 subjects. Of total 329 somatic primary-associated mutated genes, 208 genes were recurrently shown in metastasis. MUC4, PRG4, KRTAP5-7, IGFn1, KRTAP9-1, MUC5B, PRR21, CHIT1, ZNF141, MYO5B, IVL, NACAD, NEFH, and HOMEZ were shown high frequency.

From the GO and pathway enrichment analysis, we learned that most of the primary somatically mutated genes were correlated with intermediate filament and cytoskeleton signal transduction regulations, and involved in extracellular matrix structural constituent. As we know, genes involving in intermediate filament are usually expressed in non-small cell lung cancer [58], prostate cancer [59], pancreatic cancer [60], breast cancer [61] and malignant melanoma18 from transcriptional and post-transcriptional levels via immunohistochemical analysis, and were specified their functions in cell motility and a cancer progression pathways. It suggested that these keratin genes with genetic changes maybe the drivers for the primary tumor.

TAS2R31 and UMODL1 have also been reported in lung adenocarcinoma [16], but limited descriptions about their functions was found. Of the 113 somatic mutation genes, mutation occurred only once in many genes in any metastasis loci in one patient. It revealed great intra-tumor heterogeneity in potential metastasis genes. For the two highest frequent somatic genes, UMODL1 was up-regulated in about 20 % non-small cell lung cancers [21], and was also up-regulated in metastatic prostate cancer [22] or acute myeloid leukemia [23]. Mutation rate of TAS2R31 was 0.98 % in lung and 1.82% in upper aerodigestive tract in COSMIC (v69). In colon cancer, TAS2R31 was also highly mutated in protein coding region [24].

The heterotrimeric G-protein subunits Gα and Gβγ are involved in cellular transformation and tumor development. Heterotrimeric G-proteins and cAMP signaling pathways play
a pivotal role in post-receptor transduction and immediate cellular responses, such as secretory processes (e.g. hormone release and the production of autocrine/paracrine growth factors, cytokines, and matrix metalloproteinases), actin cytoskeleton reorganization, cellular adhesion and migration, and delayed responses linked to gene transcription, cell growth and differentiation. Several lines of evidence [62] demonstrate that the signaling pathways governed by Rho-like GTPases and heterotrimeric G-proteins are interdependent and contribute to cellular adhesion, transformation, and tumor cell invasiveness. In our study, mutation of genes such as GPR144, OR5M10, OR1A1, OR4M2, OR5M3, GABBR2, OR4C16, EMR1, OR4A16, OR13F1, TAS2R43, TAS2R31 and PRKD3, which are associated with G-protein coupled receptor protein signaling pathway, was detected (data not shown). These findings suggested that G-protein coupled receptor protein signaling pathway act very important function in the lung cancer metastasis.

Taken together, development of tumor involves a various evolution modes, which makes therapy complicated. Genomic analyses from single tumor-biopsy specimens may underestimate the mutational burden of heterogeneous tumors. Intra-tumor heterogeneity may explain the difficulties encountered in the validation of oncology biomarkers owing to sampling bias. Tumor clonal architectures and common mutations located in the trunk of the phylogenetic tree may be novel targets to explore more robust biomarkers and therapeutic approaches.

Acknowledgements

This research was supported by the Chinese Ministry of Science and Technology, Sino-Swiss International Collaboration project (2012DFG31320). The authors thank all the patients who participated in this study.

Disclosure Statement

All the authors declare that they have no conflict of interest.

References

Tan et al.: Genomic Alteration in Lung Adenocarcinoma

Tan et al.: Genomic Alteration in Lung Adenocarcinoma

Cellular Physiology and Biochemistry 2016;38:469-486
DOI: 10.1159/000438644
Published online: February 01, 2016
© 2016 S. Karger AG, Basel
www.karger.com/cpb


