Genomic Alterations in the RB Pathway Indicate Prognostic Outcomes of Early-Stage Lung Adenocarcinoma

Seongmin Choi1,2, Hyeong Ryul Kim3, Chang Ohk Sung4, Jongkyu Kim1,2, Sukjun Kim1,2, Sung-Min Ahn5,6,7, Chang-min Choi8, Sung-Min Chun4, Eun Kyung Choi9, Sang-We Kim6, Yong-Hee Kim3, Ji-Young Lee10, Joon Seon Song4, Deokhoon Kim7, Farhan Haq7,11, Sun Young Lee7, Jong-eun Lee12, Wang-rim Jung12, Hye Yoon Jang12, Eunho Yang12, Charles Lee13,14, Eunsil Yu4, Gu Kong15, Daehyun Baek1,2,16, and Se Jin Jang4,6

Authors’ Affiliations: 1Center for RNA Research, Institute for Basic Science; 2School of Biological Sciences, College of Natural Sciences, Seoul National University; 3Department of Thoracic and Cardiovascular Surgery; 4Department of Pathology; 5Department of Oncology; 6Center for Cancer Genome Discovery, Asan Institute for Life Science; 7Department of Bioinformatics; 8Department of Pulmonology and Medical Oncology; 9Department of Radiation Oncology, University of Ulsan College of Medicine, Asan Medical Center; 10Asan Institute for Life Sciences, Seoul; 11Lee Gil Ya Cancer and Diabetes Institute, Gachon University, Incheon; 12DNA Link, Inc., Asan Institute for Life Sciences, Seoul, South Korea; 13The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut; 14Department of Graduate Studies, Seoul National University College of Medicine; 15Department of Pathology, College of Medicine, Hanyang University; 16Bioinformatics Institute, Seoul National University, Seoul, South Korea.
Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

S. Choi, H.R. Kim, C.O. Sung, G. Kong, D. Baek, and S.J. Jang contributed equally to this work.

Corresponding Authors:
Gu Kong, Department of Pathology, College of Medicine, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, South Korea. Phone: 82-2-2290-8251; Fax: 82-2-2295-1091; E-mail: gkong@hanyang.ac.kr; Daehyun Baek, Rm 423 Bldg 504, School of Biological Sciences, Seoul National University, Seoul, South Korea, 151-747. Phone: 82-2-880-4264; Fax: 82-2-887-0244; E-mail: baek@snu.ac.kr; and Se Jin Jang, Department of Pathology, University of Ulsan College of Medicine, Asan Medical Center, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 138-736, South Korea. Phone: 82-2-3010-5966; Fax: 82-2-472-7898; E-mail: jangsejin@amc.seoul.kr

Running title: RB pathway mutations in early-stage lung adenocarcinoma

Disclosure of Potential Conflicts of Interest
Jong-eun Lee, Wang-rim Jung, Hye Yoon Jang, and Eunho Yang are employees of DNALink, a public company that provides genomic analysis and bioinformatics services. The analysis team of DNALink contributed to the data generation of this study, but no DNALink patents or products (either marketed or in development) were used. No other authors have any competing
interests to declare.
Abstract

**Purpose:** To better understand the complete genomic architecture of lung adenocarcinoma (LA).

**Experimental Design:** We used array experiments to determine copy number variations and sequenced the complete exomes of the 247 LA tumor samples along with matched normal cells obtained from the same patients. Fully annotated clinical data were also available, providing an unprecedented opportunity to assess the impact of genomic alterations on clinical outcomes.

**Results:** We discovered that genomic alternations in the RB pathway are associated with significantly shorter disease-free survival in early-stage LA patients. This association was also observed in our independent validation cohort. The current treatment guidelines for early-stage LA patients recommend follow-up without adjuvant therapy after complete resection, except for high-risk patients. However, our findings raise the interesting possibility that additional clinical interventions might provide medical benefits to early-stage LA patients with genomic alterations in the RB pathway. When examining the association between genomic mutation and histological subtype, we uncovered the characteristic genomic signatures of various histological subtypes. Notably, the solid and the micropapillary subtypes demonstrated great diversity in the mutated genes, while the mucinous subtype exhibited the most unique landscape. This suggests that a more tailored therapeutic approach should be used to treat LA patients.

**Conclusion:** Our analysis of the genomic and clinical data for 247 LAs should help provide a more comprehensive genomic portrait of LA, define molecular signatures of LA...
subtypes, and lead to the discovery of useful prognostic markers that could be used in personalized treatments for early-stage LA patients.

**Translational Relevance**

Lung cancer is the leading cause of cancer-related mortality worldwide and its most predominant subtype is lung adenocarcinoma (LA). For early-stage LAs, surgical resection has been shown to increase curative outcomes and thus remains the recommended treatment approach. However, the survival rate after curative resection in early-stage LA cases is still lower than 50%. In our present study, we performed microarray analyses of 247 LAs to detect copy number variations and also sequenced the complete exomes of these tumor specimens. As a result, we discovered that genomic alternations to the RB pathway are strongly linked to a significantly shorter disease-free survival outcome for early-stage LA patients. The current treatment guideline for early-stage LA patients comprises mainly follow-up without adjuvant therapy after resection. However, our current findings raise an interesting possibility that additional clinical interventions might provide medical benefit for early-stage LA patients who have genomic alterations to the RB pathway.
Introduction

Worldwide, lung cancer is the leading cause of cancer-related mortality (1). The incidence of lung adenocarcinoma (LA) continues to rise and is now the most frequent histological subtype according to nationwide surveys in East Asian countries (2, 3) and the United States (4). Recent large-scale genomic studies have identified novel therapeutic targets and accelerated the development of personalized treatment for LA. A previous study detected high-resolution copy number variations (CNVs) in 371 tumors, revealing novel candidates such as NKX2-1 and VEGFA (5). Sanger and high-throughput sequencing have validated tumor-suppressing candidates (e.g., NF1, ATM, APC, and CTNNB1) and possible proto-oncogenes such as ERBB2, ERBB4, and EPHA3 (6). The whole-genome sequencing of a single patient revealed additional mutations in NEK9 (7). More recently, a hybrid approach that combines whole-exome and whole-genome sequencing was used to identify possible candidates, including U2AF1, ARID1A, and MAST2 (8). Another study, which sequenced whole exomes and whole transcriptomes, identified driver candidates such as LMTK2, ARID1A, and CCDC6-ROS1 fusion (9). However, newly discovered genomic targets have not been applied to patient treatment, in part because of disparities in drug discovery. Despite new insights into the genomic portrait of LA acquired using large-scale studies, the association and causality between such genomic alterations and distinct clinical outcomes largely remain unexplored.

Most genomic studies on early-stage LA analyze resected tumor specimens that did not receive adjuvant chemotherapy mainly because surgical resection is the recommended choice of treatment and has been shown to increase the chance of successful cure (10). However, the
survival rates following curative resection for clinical and pathologic stage I LA are <50% and <75%, respectively (11, 12). The most common cause of treatment failure is disease recurrence. Therefore, the development of genome-based personalized approaches that stratify early-stage LA patients may provide better treatment alternatives. Particularly, the use of prognostic genomic markers, which can determine high-risk patients who require additional treatments such as adjuvant chemotherapy following surgery, may help improve clinical outcomes.

When selecting possible therapeutic options for LA patients and predicting prognosis, it is common to classify tumor samples as various predominant subtypes according to histology and morphology (13). For instance, it is evident that the \textit{EGFR-TK} inhibitor gefitinib is a more effective treatment for patients of the papillary subtype (14, 15), while patients of the solid subtype are less likely to have \textit{EGFR} mutations (16). \textit{KRAS}, one of the most significantly mutated genes in LA, is also associated with the mucinous subtype (17). However, despite anecdotal evidence that a few specific mutations are associated with histological subtype, the comprehensive genomic landscape of these predominant subtypes remains undetermined.

In this study, we performed CNV array experiments and sequenced whole-exome pairs (tumor and matched nontumor DNA) in a discovery cohort \((n = 170)\) and an independent validation cohort \((n = 77)\). This large-scale genomic dataset, in combination with high-quality clinical data, gave us an unprecedented opportunity to gain insight into the complete genomic portrait of LA, the molecular signatures of LA subtypes, and potential prognostic markers that could be used to improve the clinical outcomes of LA patients.
Materials and Methods

Patients and tumor specimens

This study included a whole-exome and CNV analysis of 247 Korean LA patients. LA tumor and matched normal tissue samples were obtained from two independent patient cohorts that include a discovery cohort ($n = 170$) and a validation cohort ($n = 77$). Patients in both cohorts had completely annotated clinical data, including overall survival, recurrence, histological subtype, clinical stage, tumor stage, and smoking status. All specimens used in this study were obtained with the approval of the institutional review board of Asan Medical Center (Seoul, South Korea), and documented informed consent was obtained from all patients. The specimens and data used in this study were provided by the Asan Bio-Resource Center of the Korea Biobank Network (Seoul, South Korea).

Sample preparation and evaluation

Sections of 291 flash-frozen samples were stained with hematoxylin and eosin (H&E) and subjected to histological examination by pathologists in order to determine tumor area in the slides (coverage area) and cellularity. DNA was only extracted from the 256 tumors with coverage areas $\geq 80\%$ and tumor cellularity $\geq 60\%$. Genomic DNA was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer’s instructions. After elution in DNase- and RNase-free water, genomic DNA was quantified using the NanoDrop spectrophotometer and PicoGreen system (Invitrogen). Six samples failed DNA quality control and the final 247 samples were sequenced. Then, from the 247 patients, 170 patients were
randomly chosen to be the discovery cohort, leaving 77 patients in the validation cohort. Using a multieheaded microscope, H&E-stained tumor sections were reviewed by two pathologists (J.S.S. and S.J.J.) blind to the study protocol, and were evaluated simultaneously until a consensus was reached. Tumor subtypes were classified according to the IASLC/ATS/ERS classifications (13). Minimally invasive and invasive adenocarcinomas were divided into five subtypes: papillary-predominant, acinar-predominant, micropapillary-predominant, solid-predominant, and invasive mucinous adenocarcinoma.

**Exome capture and massive parallel sequencing**

SureSelect sequencing libraries were prepared using the SureSelect All Exon 50Mbp kit (Agilent) according to the manufacturer’s instructions. The quality of the amplified fragment libraries was verified by capillary electrophoresis using the Agilent Bioanalyzer. Cluster generation in the flow cells was achieved using the cBot automated cluster generation system (Illumina). Then, the flow cells were loaded onto an Illumina HiSeq2000 sequencing system, and sequencing was performed using a read length of 2×100bp.

**Validation of detected mutations, bioinformatics, and statistical analyses**

Detailed methods for validation of detected mutations, immunohistochemistry, bioinformatics, and statistical analyses can be found in the Supplementary Methods.
Results

Clinicopathological features of studied population

We obtained tumor and matched normal tissue samples from two independent patient cohorts ($n = 247$): a discovery cohort ($n = 170$) and a validation cohort ($n = 77$). Demographics of patients information including smoking status, histological subtype, and recurrence is provided in Table 1. According to the staging criteria of the American Joint Committee on Cancer, 157, 44, 40, and 6 patients were classified as clinical stage I–IV, respectively. About half were never-smokers ($n = 122$), and the rest were either former or current smokers ($n = 125$). Such fraction shows a clear difference from previous studies, which include only 10–15% of never-smokers (6, 8). Almost all smokers were male (94%), and the positive association between smoking status and sex was statistically significant ($P < 2.2\times10^{-16}$, Fisher’s exact test). Most tumor samples included a heterogeneous mixture of multiple histological subtypes. By assessing the most predominant subtype, we classified each tumor sample as one of the following five major subtypes according to new international classifications (13): acinar, papillary, solid, micropapillary, or mucinous. Accordingly, we identified 81 acinar, 66 papillary, 64 solid, 20 micropapillary, and 16 mucinous subtypes.

Results of whole-exome sequencing and validation

The whole exomes of the tumor and normal samples obtained from patients in both cohorts were analyzed using high-throughput sequencing. After applying multiple stringent filters, we obtained a median of 28 Mbp of exome sequences that have sufficient coverage
(≥10×). Overall, the mean depth was 49.8× (range: 13.9–117.3×) (Supplementary Table S1). To reduce potential false positives, which were caused by a lenient parameter cutoff recommended by the published bioinformatics tool that we utilized (18), we used a more stringent cutoff (Supplementary Table S2). We obtained a mean rate of 3.9 mutations (substitutions and small indels) per Mbp, which is lower than that of a previous study (8), perhaps because of the larger number of never-smokers included in our cohort. The overall pattern of detected mutations is fairly similar to those of previous reports (8, 9). Smokers demonstrated a higher overall mutation rate than never-smokers (5.5 vs. 1.9 per Mbp, \( P = 1.3 \times 10^{-9} \), rank sum test), as previously reported (6, 8, 9). Especially, C to G and C to A substitutions demonstrated pronounced differences dependent on smoking status consistently with a previous study (Supplementary Fig. S1 and Supplementary Table S3) (8).

To evaluate the accuracy of the whole-exome sequencing, we performed Sanger sequencing on 161 randomly chosen mutations (31 small indels and 130 substitutions). The validation rate was 100% for small indels (31 of 31), 95% for substitutions (123 of 130), and 96% for both types of mutations combined (154 of 161). Furthermore, we used mass spectrometry to independently validate additional 90 randomly chosen substitutions and obtained an overall validation rate of 71% (64 of 90) (see Supplementary Methods). The high validation rates of these two independent technologies show that rigorous filtering is able to detect somatic mutations with a high level of accuracy.

**Detection of potential cancer drivers and CNVs**

To discover genes that are specifically mutated in cancer cells, we attempted to detect
genes that accumulated significantly more mutations than the estimated number of mutations calculated from their background mutation rates (BMRs). To compute the BMR, we estimated the nonsynonymous BMR from the synonymous BMR of each gene to reflect the fluctuating local mutation rate (Supplementary Methods). We also analyzed each base type separately to reflect the different mutation rate of each base type. This approach appears to improve the overall detection accuracy compared with previous approaches (Supplementary Table S4 and Supplementary Methods) (6, 19). Using this improved method, we detected 48 significantly mutated genes (Supplementary Table S5). To further filter out potential false positives, we used the expression data obtained from 87 LA patients (9) and filtered out genes not likely to be expressed in LA (Supplementary Methods), to determine our final list of 22 significantly mutated genes (Fig. 1A, Table 2). Among these, 6 genes have already been reported (EGFR, TP53, KRAS, PIK3CA, RB1, and SETD2) and the other 16 genes appear to be novel driver candidates. Consistently with previous reports (6, 8, 20), there was a significant anti-association between EGFR and KRAS mutations ($P = 8.2 \times 10^{-3}$, Fisher’s exact test). Possibly due to our stringent cutoff criteria, some previously reported genes were absent from our final list; however, we were able to detect $\geq 1$ mutation(s) in each of previously reported 29 genes (Supplementary Table S6) (6, 8, 9). The known functions of these 16 novel driver candidates suggest that some of them might be related to LA occurrence or progression. For instance, COL11A1 has been proposed as a poor prognostic marker for non-small cell lung cancer (21). CENPF, a centromere protein-coding gene, and COL6A3 have been identified as poor prognostic markers for breast cancer (22, 23). SLIT2, which is involved in the SLIT/ROBO-signaling pathway, is also associated with poor prognosis in lung cancer (24, 25).
To detect the CNVs associated with LA, we used CNV arrays for 247 pairs of tumor and their matched normal samples (Supplementary Methods) and processed them using a publicly available tool (26). After removing CNV regions that did not harbor any genes expressed in LA, 13 significantly altered CNV regions (including 60 genes) were detected. Among these, 6 CNV regions have been previously reported (8), including regions that encompass TERT, CDKN2A, MDM2, MYC, ERBB2, and RB1 (Fig. 1B and Supplementary Table S7). Additionally, dozens of novel driver candidates within the 13 significantly altered CNV regions were identified (Supplementary Table S7). For instance, FOXA1 has been reported as a driver candidate in prostate cancer and identified as a therapeutic target for breast cancer (27, 28), YEATS4 has been identified as a novel oncogene that regulates the p53 pathway (29), and the amplification of a potential oncogene ZNF217 has been shown to regulate expression of ErbB3 receptor tyrosine kinase in breast cancer (30).

We searched for drugs that could potentially target the significantly altered genes in the discovery cohort. When looking at the 22 significantly mutated genes and the 60 genes located on the 13 significantly altered CNV regions, eight of them can be targeted by known drugs (Supplementary Table S8). Among these 8 genes, 4 can be targeted by drugs with known tumor-inhibiting activities. The oncogenes EGFR, ERBB2, PIK3CA, and TERT can be targeted by tailored drugs such as erlotinib (15), trastuzumab (31), NVP-BEZ235 (32), and GRN163L (33), respectively. These target drugs and their respective target genes are mostly present in different pathways (Supplementary Fig. S2). The expanded list of novel driver candidates discovered in this study suggests that additional investigations into these candidates might lead to the discovery of novel drug targets for the treatment of LA.
Association between RB pathway alterations and poor prognosis in early-stage LA patients

We performed pathway enrichment tests on the 22 significantly mutated genes and 60 genes within significantly altered CNVs. We detected 19 significantly enriched pathways (Supplementary Table S9), almost all of which are related to cancer, including non-small cell lung cancer (KEGG pathway, \( P = 1.6 \times 10^{-7} \), modified Fisher’s exact test). This is another indication that these significantly mutated genes and CNVs are involved in LA occurrence or progression.

Next, we examined whether genomic alterations accumulated on the genes of known pathways could result in differential clinical outcomes (Figs 2A to 2D). We found that genomic alternations in the RB pathway (Supplementary Fig. S3) are strongly associated with significantly shorter disease-free survival in stage I and II LA patients (\( P = 1.4 \times 10^{-3} \) and \( q = 0.031 \), log-rank test; Fig. 2A). When we excluded early-stage LA patients who were treated with adjuvant chemotherapy from the analysis, we observed an even stronger association (\( P = 6.7 \times 10^{-4} \) and \( q = 0.026 \), log-rank test; Supplementary Fig. S4A). On the other hand, this association was not observed in stage III or IV LA patients, regardless of including patients treated with adjuvant chemotherapy (\( P = 0.11 \) and \( q = 0.21 \), log-rank test; Fig. 2C) or not (\( P = 0.28 \) and \( q = 0.25 \), log-rank test; Supplementary Fig. S4C).

To rule out the possibility that this discrepancy originated from the small number of late-stage patients included in our discovery cohort, we repeated our analysis using the same number of randomly sampled early-stage patients and found that the observed association between RB pathway mutations and poor disease-free survival is specific to early-stage LA patients,
regardless of including patients treated with adjuvant chemotherapy ($P = 0.025$, log-rank test; Supplementary Fig. S5) or not ($P = 0.020$, log-rank test; Supplementary Fig. S4E). Furthermore, our independent validation cohort of 77 LA patients confirmed this association ($P = 0.017$ and $4.0 \times 10^{-3}$ respectively, log-rank test; Fig. 2B and Supplementary Fig. S4B). When the discovery and validation cohorts were combined, the association was statistically significant ($P = 5.2 \times 10^{-5}$, log-rank test).

To verify that the association between RB pathway alterations and poor prognosis was not due to confounding effects (e.g., sex, age at diagnosis, node metastasis), we performed univariable and multivariable Cox regression analyses. Accordingly, we found that RB pathway alterations were significantly associated with poor prognosis in the discovery cohort, regardless of correcting for the confounding factors in the survival prediction model ($P = 4.6 \times 10^{-3}$, multivariable Cox regression analysis; Supplementary Table S10) or not ($P = 1.9 \times 10^{-3}$, univariable Cox regression analysis; Supplementary Table S11). These results were also reproduced in the validation cohort ($P = 9.4 \times 10^{-3}$ and 0.024 according to multivariable and univariable Cox regression analyses, respectively). In contrast to the current treatment guidelines for early-stage LA patients, which recommends follow-up without adjuvant therapy after complete resection except for high-risk patients (10), these results raise the interesting possibility that early-stage LA patients with genomic alterations in the RB pathway might benefit from additional medical interventions.

**RB1 inactivation alters expression of cell-cycle-related proteins in LA tumors**

We evaluated whether *RB1* inactivation, resulting from mutations or CNV deletion, is
associated with protein level changes of other genes related to cell-cycle regulation. Protein expression levels of pRB1, E2F1, cyclin D1, and cyclin E1 within 247 tumor samples were evaluated by immunohistochemistry and compared for statistical differences (Fig. 3A). Tumors with inactivated \textit{RB1} showed significantly lower pRB1 expression but higher E2F1 expression \((P = 3.2 \times 10^{-5} \text{ and } 1.0 \times 10^{-4}, \text{ respectively, rank sum test; Fig. 3B}),\) consistent with previous reports \((34, 35).\) Higher E2F1 expression correlated with higher expression of cyclin E1 \((P = 5.5 \times 10^{-4}; \text{ Fig. 3C}).\) Additionally, tumors with inactivated \textit{RB1} exhibited significantly lower expression of cyclin D1 \((P = 1.6 \times 10^{-5}; \text{ Fig. 3B}),\) consistent with a previous report that demonstrated that \textit{RB1} mutations promote cyclin degradation, leading to lower expression of cyclin D1 \((36).\) Taken together, these results indicate that tumors with inactivated \textit{RB1} have altered expression of other proteins essential to cell-cycle regulation, providing a possible link between RB pathway alterations and poor prognosis in early-stage LA patients.

\textbf{Molecular definitions of LA subtypes}

To gain more comprehensive insight into the molecular definitions of LA subtypes, we investigated the associations between significantly mutated genes and tumor subtypes, which were classified on the basis of the most up-to-date tumor classification criteria \((13).\)

The overall molecular definition demonstrated the similarities between the acinar and the papillary subtypes, and between the micropapillary and the solid subtypes \((\text{Fig. 4A}).\) Consistent with this observation, when subjected to hierarchical clustering according to the mutation profile, the acinar and the papillary subtypes were included in the same cluster, while the micropapillary and the solid subtypes formed another cluster \((\text{Fig. 4B}).\) These clustering
patterns agree with the known histological and morphological characteristics of these tumor subtypes (13).

The molecular definitions of the acinar and the papillary subtypes appear to be dramatically different from those of the micropapillary and the solid subtypes. One of main reasons for this discrepancy stems from EGFR mutations, which are strongly associated with the papillary subtype ($P = 4.7 \times 10^{-6}$, Fisher’s exact test) and inversely associated with the solid ($P = 1.5 \times 10^{-3}$, Fisher’s exact test), the mucinous ($P = 2.7 \times 10^{-3}$, Fisher’s exact test), and the micropapillary subtypes ($P = 8.0 \times 10^{-3}$, Fisher’s exact test). Some of these associations have been previously reported (16), but our large-scale dataset enabled us to look into these associations more completely. Interestingly, the mucinous subtype demonstrated a unique mutation profile, in that the most prevalent mutations accumulated in KRAS (Fig. 4A). It is also noteworthy that the solid subtype demonstrated a very heterogeneous mutational profile (Fig. 4A), which is consistent with the observation that this subtype often demonstrates unusual histological and morphological characteristics that are fairly different from other subtypes (13). Taken together, our detailed analyses call for a more careful investigation of the molecular definitions of the LA subtypes and indicate that a better tailored therapeutic approach should be employed when treating LA patients in order to improve overall therapeutic outcomes.
Discussion

For this study, we generated the whole exome and CNV dataset together with fully annotated clinical data on 247 LAs, which provide an unprecedented opportunity for investigating the genomic architecture of LA and genomic alternations associated with clinical phenotypes and outcomes.

By processing the exome data according to our stringent statistical approaches, we detected 22 significantly mutated genes with either base substitutions or small indels, which accounts for 78% of our discovery cohort. 69% of the discovery cohort had such mutations in 6 previously reported genes \((\text{EGFR, TP53, KRAS, PIK3CA, RB1, and SETD2})\). The remaining 9% of cases had base substitutions or small indels that accumulated in the 16 novel driver candidates, including \(\text{COL11A1, CENPF, COL6A3, LRBA, and SLIT2}\). Another 6% of the discovery cohort demonstrated \(\geq 1\) significant CNV(s), leaving 15% of LAs unexplained by our exome or CNV data (Supplementary Fig. S6). Potential reasons for the 15% of unexplained LAs include tumor impurity, relatively low overall depth of exome sequencing, translocation or fusion genes such as \(\text{EML4-ALK}\) (37, 38), genomic alterations accumulating in noncoding regions, epigenetic changes (39, 40), and technical underrepresentation by exome capture. This large-scale effort extensively utilized whole-exome sequencing, Sequenom technology, and CNV arrays to detect dozens of novel driver candidates and CNVs, and yet only a small fraction (9%+6%=15%) of LAs can be additionally explained by these genomic alternations compared with 69% of LAs that can be explained by the genomic alternations that accumulated on 6 previously reported genes (Supplementary Fig. S6). In order to comprehensively verify the
molecular causes responsible for the 15% of unexplained LAs, we speculate that future efforts should be directed toward utilizing multiplatform technologies that concurrently monitor the transcriptome, methylome, and noncoding genome, in addition to coding genomes and CNVs.

Although we detected a large number of significantly mutated genes in most of the 170 patients in our discovery cohort, the overall prevalence of mutations in a few genes was lower than previously reported. For instance, \textit{KRAS} has been reported as mutated in about 12% of Asian LA patients (9), whereas in the present study \textit{KRAS} was only affected in 6% of patients in both the discovery and validation cohorts. In order to validate the prevalence of \textit{KRAS} and a few other genes in our dataset, we employed the more sensitive Sequenom approach, which utilizes mass spectroscopy-based genotyping. Accordingly, 81\% of \textit{EGFR} substitutions (29 of 36) and 100\% of \textit{BRAF} substitutions (3 of 3) detected by the Sequenom technology were also detected by exome sequencing (Supplementary Table S12). However, only 21\% of \textit{KRAS} substitutions (3 of 18) detected using this technology were detected in our exome dataset.

To understand the underlying reasons for this discrepancy, we looked at the depths of the sequence reads and found that the median read depth of the first \textit{KRAS} exon (where the recurrent mutations were located) was significantly lower than in other \textit{KRAS} regions ($P < 2.2 \times 10^{-16}$, rank sum test). However, this phenomenon appears to be specific to our exome data because another exome sequencing study on AML-M5 did not show a lower read depth for the first \textit{KRAS} exon (Supplementary Fig. S7) (41). Therefore, we speculate that the low detection sensitivity of our exome dataset may be due to low capturing efficiency in specific regions of the exome, including the first \textit{KRAS} exon. In support of this hypothesis, the read depth of the first \textit{KRAS} exon in patients in whom we were unable to detect \textit{KRAS} mutations was significantly lower than in
patients who have detectable KRAS mutations (P < 2.2×10^{-16}, rank sum test; Supplementary Fig. S7C). These findings illustrate that, depending on the exome-capturing platform, genomic studies based on exome sequencing might include significantly underrepresented regions. Thus, alternative technologies should be employed to compensate for low detection sensitivity in such regions. In this study, we complemented EGFR, KRAS, and PI3KCA mutations using Sequenom-detected mutations. This increased the overall prevalence of EGFR, KRAS, and PIK3CA mutations from 28%, 6%, and 5% to 44%, 14%, and 6%, respectively.

There are several genes that have been previously reported to be driver candidates in LA but that were not detected in our analysis, such as KEAP1 and STK11 (6, 8, 20). Of the 247 patients, only four patients had KEAP1 mutations, and only two patients had STK11 mutations. The lower prevalence of these mutations, compared to a previous study done in the U.S. where 22 and 27 out of 183 patients bore mutations in KEAP1 and STK11, respectively (8), may be the major reason these genes were not detected as driver candidates in our analysis. The low prevalence of mutated KEAP1 and STK11 may be due to the low capture efficiency discussed above or to the different ethnic backgrounds of the study populations. The median sequence depths of KEAP1 and STK11 (22 and 7, respectively; Supplementary Figure S8) are much lower than the median depth of 40 for the overall tumor samples and are lower than the depths of known driver genes such as EGFR, KRAS, and BRAF (45, 48, and 43, respectively), indicating that the lower capture efficiency may be at least partially responsible for the low prevalence. It is also possible that the prevalence of these genes in the Asian population may differ from that of Western populations. Supporting this possibility, another independent exome study on LA patients in South Korea reported low prevalence of mutated STK11 (2 out of 87) and KEAP1 (3
The Cancer Genome Atlas (TCGA) recently reported a comprehensive molecular profile of 230 LA patients with diverse ethnic backgrounds. While the TCGA study agrees well with ours on many key findings such as a higher overall mutation rate in smokers than in never-smokers and the negative association between \( EGFR \) and \( KRAS \) mutations, there are noticeable differences between the two studies. For example, among the 18 and 22 significantly mutated genes detected in the studies, TCGA and ours, respectively, only 6 genes (\( EGFR, TP53, KRAS, PIK3CA, RB1, \) and \( SETD2 \)) were found to be present in both lists. Also, the prevalence of \( EGFR \) mutations was much higher in our study (42.4%) than in the TCGA study (14.3%). These differences may have originated from the differences in ethnic backgrounds and the relative fraction of smokers included in the analysis. Indeed, the TCGA study had a smaller fraction of never-smokers (14.3%, \( n = 33 \)) than in ours (49%, \( n = 122 \)) and the different mutation profiles related to ethnic backgrounds have been previously reported. These two studies should be complementary to each other, providing a more comprehensive molecular profile of LA and a useful resource for future comparative studies that aim to analyze the complexity and the heterogeneity of LA.

To reduce potential false positives in this study, we used a more stringent MuTect parameter, and by doing so, we might have sacrificed the overall detection sensitivity. To estimate the number of mutations we might be missing due to our stringent parameters, we compared the overall detection sensitivity between our current stringent parameter (MuTect \( \theta_T = 8.0 \)) and the more lenient default detection parameter (MuTect \( \theta_T = 6.3 \)). The more lenient cutoff increased the total number of detected mutations by 50%, from 25,065 to 37,548. However, the
prevalence of known driver genes was not significantly affected. For instance, the prevalence of \textit{EGFR} was 42.9\% using the lenient cutoff compared to 42.4\% using our current stringent cutoff. Similarly, the prevalence of \textit{TP53} and \textit{PIK3A} only increased marginally, from 25.3\% to 28.8\% for \textit{TP53} and from 14.1\% to 14.7\% for \textit{PIK3CA}. The prevalence of other driver genes was similarly unchanged (Supplementary Table S13). For 45 known driver genes, using the more stringent cutoff reduced the average prevalence only by 0.5\% in comparison to the default parameter, indicating that the additional mutations detected using the lenient cutoff might include a high fraction of false positives.

It has been reported that LA has a high overall mutation rate (MR), and that the MR is higher in smokers than in never-smokers (8, 43). In comparison to previous reports, our overall MR was a bit lower, probably due to our high detection stringency. Using the current stringent cutoff, we obtained MRs of 5.5/Mbp and 1.9/Mbp for smokers and never-smokers, respectively (Supplementary Fig. S9A). When the lenient cutoff was applied, the MR of smokers increased to 7.7/Mbp. However, the MR of never-smokers also increased, to 4.4/Mbp, resulting in too little difference in MR between smokers and never-smokers (Supplementary Fig. S9B). Repeated analysis with an even more lenient parameter ($\theta_T = 5.5$) resulted in nearly equal MRs (11.1/Mbp for smokers and 8.5/Mbp for never-smokers, Supplementary Fig. S9C). These analyses also suggest that the additional substitutions detected using the lenient cutoff might include many false positives. We concluded that the stringent cutoff results in a more robust set of substitutions at the expense of lowered detection sensitivity. Therefore, we decided to continue to use the more stringent parameter instead of the lenient parameter.
It is noteworthy that the current treatment guidelines for early-stage LA provided by the National Comprehensive Cancer Network recommend follow-up without adjuvant therapy after complete resection, except for patients who have residual tumors or those considered at high risk (10). However, since the long-term treatment failure rate is close to 50% (10), it is important to identify subgroups of patients who may benefit from adjuvant treatments such as chemotherapy, even in early-stage LA. Our discovery that RB pathway mutations are associated with poor prognosis calls for additional clinical interventions that could improve the clinical outcomes of early-stage LA patients. In particular, a set of validated diagnostic markers that could detect genomic alterations in the RB pathway would be useful for patient stratification.

Rb protein acts as a tumor suppressor that inhibits cell cycle progression by binding and inhibiting transcription factors of the E2F family. When Rb is phosphorylated to pRb by cyclin-dependent kinases, it is unable to form a complex with E2F, which allows E2F to promote cell cycle progression (44). In this study, lower levels of pRb and cyclin D1 and higher levels of E2F1, a member of the E2F family, were detected in tumors with RB1 alterations compared to tumors without RB1 alterations. These findings agree with a previous report showing a lack of cyclin D1 overexpression and high E2F1 levels in tumors with Rb loss (36). Lack of cyclin D1 overexpression in Rb loss can be explained by the increased disassembly of cyclin D-cdk4 complexes and increased cyclin D turnover (36, 45). We also observed that tumors with RB1 alterations are characterized by low pRb levels and high E2F1 levels, which may explain the poor prognosis of these patients. In addition, our study reveals that high E2F1 protein expression is associated with high cyclin E1 levels, which is consistent with the fact that E2F binds the promoter of the cyclin E gene and increases its expression. High levels of E2F1 and cyclin E1 in
tumors with RB1 alterations suggests that the prognostic effect of RB1 alterations may be related to the actions of these downstream molecules of Rb1. Our analysis suggests that either pan-CDK inhibitors or selective CDK inhibitors, being used more frequently in clinical trials (46), might be effective to target these downstream molecules of Rb1.

Our analysis of the whole exome, CNV dataset, and fully annotated clinical data provide comprehensive insights into the genomic portrait of LA and help define the molecular signatures of LA subtypes. These findings could lead to the discovery of useful prognostic markers and help devise personalized treatment strategies for patients with early-stage LA.
Authors’ Contributions

Conception and design: Hyeong Ryul Kim, Chang Ohk Sung, Gu Kong, Daehyun Baek, and Se Jin Jang

Development of methodology: Seongmin Choi, Chang Ohk Sung, Jongkyu Kim, Sukjun Kim, and Daehyun Baek

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Hyeong Ryul Kim, Chang-min Choi, Sung-Min Chun, Eun Kyung Choi, Sang-We Kim, Yong-Hee Kim, Ji-Young Lee, Joon Seon Song, Jong-eun Lee, Wang-rim Jung, Hye Yoon Jang, Eunho Yang, Eunsil Yu, and Se Jin Jang

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Seongmin Choi, Hyeong Ryul Kim, Chang Ohk Sung, Jongkyu Kim, Sukjun Kim, Wang-rim Jung, Daehyun Baek, and Se Jin Jang

Writing, review, and/or revision of the manuscript: Seongmin Choi, Hyeong Ryul Kim, Chang Ohk Sung, Jongkyu Kim, Charles Lee, Daehyun Baek, and Se Jin Jang

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Sung-Min Ahn, Deokhoon Kim, Farhan Haq and Sun Young Lee

Study supervision: Gu Kong, Daehyun Baek, and Se Jin Jang

Grant Support
This work was supported by the National Project for Personalized Genomic Medicine, Ministry of Health and Welfare of Korea (A111218-SC01), by the Leading Foreign Research Institute Recruitment Program (2011-0030105) and the Basic Science Research Program (2011-0014523), the Ministry of Education, Science, and Technology of Korea, and by the Institute for Basic Science (EM1302).
FIGURE LEGENDS

Figure 1. Global mutational landscape of the 170 LA patients (discovery cohort). A, clinical data (mutation count, sex, smoking status, and subtype) and types of mutation for the 22 genes that were significantly mutated and expressed in LA. Mutation color code: dark blue, missense mutation; red, nonsense substitution; green, splice-site mutation; light blue, frame-preserving small indel; orange, frame-shifting small indel. Subtype color code: dark blue, acinar; light blue, papillary; red, solid; orange, micropapillary; green, mucinous. B, global pattern of CNV amplification (red) and deletion (blue). The amplitude of CNV amplification or deletion was determined using the G-score obtained by GISTIC (26). The chromosomal loci and representative genes of the 13 significantly altered CNVs are indicated on top (amplification) and bottom (deletion) of the panel (see Supplementary Table S7 and Supplementary Methods). The $q$ value cutoff of 0.01 is indicated by the green lines.

Figure 2. Association between RB pathway alterations and poor prognosis in early-stage LA patients. A, Kaplan-Meier curves of disease-free survival (DFS) for early-stage LA patients in the discovery cohort ($n = 129$). The survival curves of the patients with RB pathway mutations and the remaining patients are shown in red and blue, respectively. The log-rank $P$ value and $q$ value are also shown. B, DFS curve for early-stage LA patients in the validation cohort ($n = 59$). Otherwise, as in A. C, DFS curve for late-stage LA patients in the discovery cohort ($n = 27$). Otherwise, as in A. D, DFS curve for late-stage LA patients in the validation cohort ($n = 13$). Otherwise, as in A.
**Figure 3.** Rb1 gene inactivation associated with expression changes of cell-cycle-related proteins in LA tumors. **A,** protein expression of cell-cycle related proteins in 170 tumor samples as measured by immunohistochemistry (see Materials and Methods). **B,** altered protein expression of cyclinD1, E2F1, and pRB1 in tumors with inactivated RB1 compared to tumors with unaltered RB1. **C,** elevated protein expression of cyclinE1 in tumors with high expression of E2F1.

**Figure 4.** Molecular definitions of LA subtypes. **A,** for each of the five tumor subtypes (acinar, papillary, solid, micropapillary, and mucinous), the frequencies of the most prevalent 10 significantly mutated genes and significantly altered CNV regions are illustrated. Each CNV region was shown as its representative gene (see Supplementary Table S7 and Supplementary Methods). **B,** LA subtypes are clustered according to their overall mutation profile. In the left panel, tumor subtypes are clustered by the frequencies of mutated genes. For each gene, its mutation fraction was calculated by dividing the number of patients with the mutated gene by the total number of patients included in the discovery cohort. In the right panel, tumor subtypes are clustered by the degree of the association between each subtype and the mutated genes in the subtype, as measured by Fisher’s exact test. Positive associations and negative associations are represented as -log_{10}(P value) (red) or log_{10}(P value) (blue), respectively. Asterisks indicate the level of significance for each association (*P < 0.05, **Bonferroni-adjusted P < 0.05).
**Table 1.** Clinical characteristics of 247 LA patients.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Trait</th>
<th>Discovery (n = 170)</th>
<th>Validation (n = 77)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>95 (56%)</td>
<td>46 (60%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>75 (44%)</td>
<td>31 (40%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>≥70</td>
<td>31 (18%)</td>
<td>12 (16%)</td>
</tr>
<tr>
<td></td>
<td>≥60, &lt;70</td>
<td>77 (45%)</td>
<td>36 (47%)</td>
</tr>
<tr>
<td></td>
<td>&lt;60</td>
<td>62 (36%)</td>
<td>29 (38%)</td>
</tr>
<tr>
<td>Smoking status</td>
<td>Never-smoker</td>
<td>82 (48%)</td>
<td>40 (52%)</td>
</tr>
<tr>
<td></td>
<td>Smoker</td>
<td>88 (52%)</td>
<td>37 (48%)</td>
</tr>
<tr>
<td>Overall stage&lt;sup&gt;a&lt;/sup&gt;</td>
<td>I</td>
<td>110 (65%)</td>
<td>47 (61%)</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>28 (16%)</td>
<td>16 (21%)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>28 (16%)</td>
<td>12 (16%)</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>4 (2%)</td>
<td>2 (3%)</td>
</tr>
<tr>
<td>Predominant subtype</td>
<td>Acinar</td>
<td>53 (31%)</td>
<td>28 (36%)</td>
</tr>
<tr>
<td></td>
<td>Papillary</td>
<td>48 (28%)</td>
<td>18 (23%)</td>
</tr>
<tr>
<td></td>
<td>Solid</td>
<td>45 (26%)</td>
<td>19 (25%)</td>
</tr>
<tr>
<td></td>
<td>Micropapillary</td>
<td>13 (8%)</td>
<td>7 (9%)</td>
</tr>
<tr>
<td></td>
<td>Mucinous</td>
<td>11 (6%)</td>
<td>5 (6%)</td>
</tr>
<tr>
<td>Adjuvant chemotherapy treatment</td>
<td>Treated</td>
<td>30 (18%)</td>
<td>21 (27%)</td>
</tr>
<tr>
<td></td>
<td>Untreated</td>
<td>140 (82%)</td>
<td>56 (73%)</td>
</tr>
<tr>
<td>Recurrence status&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Recurrent</td>
<td>80 (47%)</td>
<td>26 (34%)</td>
</tr>
<tr>
<td></td>
<td>Not recurrent</td>
<td>77 (45%)</td>
<td>47 (61%)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Determined according to AJCC criteria (7th edition). See Supplementary Table S14 for additional stage classification data (TNM stage).

<sup>b</sup>Recurrence data on 17 patients are missing; these patients were excluded from subsequent analyses.
Table 2. Twenty-two significantly mutated genes ($q < 0.05$) expressed in LA.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>RefSeq ID</th>
<th>$q$ value $^a$</th>
<th>Prevalence $^b$</th>
<th>Number of affected patients $^c$</th>
<th>Remark $^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR</td>
<td>NM_005228</td>
<td>$&lt; 2.2 \times 10^{-16}$</td>
<td>42.4%</td>
<td>72</td>
<td>D, I, S</td>
</tr>
<tr>
<td>TP53</td>
<td>NM_000546</td>
<td>$&lt; 2.2 \times 10^{-16}$</td>
<td>25.3%</td>
<td>43</td>
<td>D, I, S</td>
</tr>
<tr>
<td>KRAS</td>
<td>NM_033360</td>
<td>$3.9 \times 10^{-9}$</td>
<td>11.8%</td>
<td>20</td>
<td>D, I, S</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>NM_006218</td>
<td>$5.4 \times 10^{-8}$</td>
<td>8.2%</td>
<td>14</td>
<td>I, S</td>
</tr>
<tr>
<td>ACACB</td>
<td>NM_001093</td>
<td>$9.7 \times 10^{-6}$</td>
<td>5.3%</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>COL6A3</td>
<td>NM_004369</td>
<td>$2.3 \times 10^{-5}$</td>
<td>4.1%</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>FRAS1</td>
<td>NM_025074</td>
<td>$8.6 \times 10^{-5}$</td>
<td>5.9%</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>PRRC2C</td>
<td>NM_015172</td>
<td>$1.7 \times 10^{-4}$</td>
<td>2.9%</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>LAMA2</td>
<td>NM_000426</td>
<td>$2.7 \times 10^{-4}$</td>
<td>3.5%</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SETD2</td>
<td>NM_014159</td>
<td>$1.1 \times 10^{-3}$</td>
<td>5.3%</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>PDE4DIP</td>
<td>NM_01198834</td>
<td>$1.3 \times 10^{-3}$</td>
<td>4.7%</td>
<td>8</td>
<td>I, S</td>
</tr>
<tr>
<td>SYNE1</td>
<td>NM_182961</td>
<td>$5.4 \times 10^{-3}$</td>
<td>6.5%</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>CENPF</td>
<td>NM_016343</td>
<td>$7.0 \times 10^{-3}$</td>
<td>4.7%</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>HMGN1</td>
<td>NM_031935</td>
<td>$9.8 \times 10^{-3}$</td>
<td>5.9%</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>COL11A1</td>
<td>NM_080629</td>
<td>$1.5 \times 10^{-2}$</td>
<td>5.3%</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>RB1</td>
<td>NM_000321</td>
<td>$3.6 \times 10^{-2}$</td>
<td>5.9%</td>
<td>10</td>
<td>D, I</td>
</tr>
<tr>
<td>COPA</td>
<td>NM_001098398</td>
<td>$4.4 \times 10^{-2}$</td>
<td>4.1%</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>LRBA</td>
<td>NM_006726</td>
<td>$4.4 \times 10^{-2}$</td>
<td>3.5%</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>F8</td>
<td>NM_000132</td>
<td>$4.4 \times 10^{-2}$</td>
<td>2.9%</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SLIT2</td>
<td>NM_004787</td>
<td>$4.4 \times 10^{-2}$</td>
<td>2.9%</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>GCN1L1</td>
<td>NM_006836</td>
<td>$4.5 \times 10^{-2}$</td>
<td>2.9%</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>MAP1B</td>
<td>NM_005909</td>
<td>$4.5 \times 10^{-2}$</td>
<td>2.9%</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Calculated after taking into account the background mutation rate, which varies with respect to genomic region and base type.

$^b$The number of affected patients divided by the number of patients in the discovery cohort (170 patients).

$^c$Indicates how many patients demonstrated substitutions and/or small indels on the gene.

$^d$D, I, and S indicate previous studies that report candidate driver genes: D, Ding et al. (2008); I, Imielinski et al. (2012); S, Seo et al. (2012).
REFERENCES

Figure 2

A) Discovery cohort

Early stage (I, II)

- Altered
- Unaffected

Survival probability

Disease-free survival (months)

$P = 1.4 \times 10^{-3}$

$q = 0.031$

B) Validation cohort

Survival probability

Disease-free survival (months)

$P = 0.017$

C) Late stage (III, IV)

Survival probability

Disease-free survival (months)

$P = 0.11$

$q = 0.21$

D) Late stage (III, IV)

Survival probability

Disease-free survival (months)

$P = 0.37$
Figure 3

A

Tumors with inactivated RB1

<table>
<thead>
<tr>
<th>Protein</th>
<th>Expression Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRB1</td>
<td>0</td>
</tr>
<tr>
<td>E2F1</td>
<td>0</td>
</tr>
<tr>
<td>CyclinD1</td>
<td>0</td>
</tr>
<tr>
<td>CyclinE1</td>
<td>0</td>
</tr>
</tbody>
</table>

Tumors with unaltered RB1

B

Protein expression level (% of positive cancer cells)

C

High E2F1 expression

Low E2F1 expression

E2F1

Protein expression level (% of positive cancer cells)

P = 5.0 × 10^-5

P = 1.8 × 10^-4

P = 6.3 × 10^-5

Inactivated RB1 by mutation or CNV deletion

mut

del

mut / del

mut

del

mut

del

mut

del

P = 4.6 × 10^-4

E2F1 expression

High

Low

CyclinE1
Figure 4

**A**

Acinar (n = 53)

- EGFR: 51%
- TP53: 8%
- TERT CNV: 9%
- NKX2-1 CNV: 11%
- KRAS: 13%
- MYC CNV: 13%
- CDKN2A CNV: 17%
- PDE4DIP: 17%
- PIK3CA: 17%
- SETD2: 17%

Micropapillary (n = 13)

- TP53: 23%
- TERT CNV: 8%
- CDKN2A CNV: 8%
- NKX2-1 CNV: 8%
- ZNF217 CNV: 8%
- EGFR: 9%
- KRAS: 15%
- PIK3CA: 15%
- ACACB: 15%
- COL6A3: 15%

Mucinous (n = 11)

- KRAS: 45%
- TP53: 9%
- TERT CNV: 9%
- ACACB: 9%
- COL6A3: 9%
- PDE4DIP: 9%
- SYNE1: 9%
- COPA: 9%
- F8: 9%
- GCN1L1: 9%

Papillary (n = 48)

- EGFR: 71%
- TERT CNV: 10%
- TP53: 10%
- MYC CNV: 13%
- PIK3CA: 13%
- NKX2-1 CNV: 17%
- ZNF217 CNV: 25%
- TSPYL4 CNV: 27%
- RB1: 10%

Solid (n = 45)

- TP53: 28%
- EGFR: 11%
- TERT CNV: 13%
- MYC CNV: 17%
- CDKN2A CNV: 13%
- MYC CNV: 13%
- SYNE1: 18%
- CDKN2A CNV: 18%
- TSPYL4 CNV: 9%
- ZNF217 CNV: 9%
- NKX2-1 CNV: 9%

**B**

Mutation fraction

- log10(P) - log10(P)

Substitution and small indel CNV
Genomic Alterations in the RB Pathway Indicate Prognostic Outcomes of Early-Stage Lung Adenocarcinoma

Seongmin Choi, Hyeong Ryul Kim, Chang Ohk Sung, et al.

*Clin Cancer Res* Published OnlineFirst October 7, 2014.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-14-0519

Supplementary Material  Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2014/10/08/1078-0432.CCR-14-0519.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.