[Supplementary Materials]

The transcriptional landscape and mutational profile of lung adenocarcinoma

Authors

Jeong-Sun Seo, Young Seok Ju, Won-Chul Lee, Jong-Yeon Shin, June Koo Lee, Thomas Bleazard, Junho Lee, Yoo Jin Jung, Jung-Oh Kim, Jung-Young Shin, Saet-Byeol Yu, Jihye Kim, Eung-Ryoung Lee, Chang-Hyun Kang, In-Kyu Park, Hwanseok Rhee, Se-Hoon Lee, Jong-Il Kim, Jin-Hyoung Kang and Young Tae Kim

Table of Contents

Supplementary Information

Supplementary Figures

Supplementary Tables

Supplementary Information

I. Cancer samples analyzed in this study

See also Supplementary Figure 1.

We collected 200 fresh surgical specimens of primary lung adenocarcinoma from patients who underwent lobectomy at Seoul National University Hospital (n=164; from 2010 to September 2011) and Seoul St. Mary's Hospital (n=36; samples deposited in tissue bank from 2009). Twenty patients from our previous report were included in this cohort[1]. For each patient, we recorded diagnosis, gender, cancer stage and smoking status (Supplementary Table 1).

We performed screening genetic tests for three well-known driver mutations (exon 18-21 of *EGFR* (n=164), exon 2 of *KRAS* (n=37); and *EML4-ALK* fusion genes (n=163) (see Methods). Of the 200 cancer tissues, 110 tissues were positive for somatic mutations in one of *EGFR* (n=99), *KRAS* (n=6) and *EML4-ALK* (n=7), with two double-positive samples (1 $EGFR^{+}KRAS^{+}$ and 1 $EGFR^{+}EML4\text{-}ALK^{+}$). Driver mutations in the remaining 90 samples were unknown.

We targeted these 90 samples for RNA sequencing. Excluding 3 samples which did not pass the RNA quality control, we obtained mRNA sequences from 87 lung adenocarcinomas. When available, we performed transcriptome (n=77) and whole-exome (n=76) sequencing of adjacent normal lung tissues for comparison between cancer and normal tissues (Supplementary Table 1 and 2).

II. Gene expression analyses

Using the RNA short-read data, we calculated expression levels for all currently known RefSeq genes (n=36,742; n=22,427 with redundant genes collapsed) in RPKM units[2] (Supplementary Table 10). The gene expression profiles are also available at our cancer-expression browser (http://gene.gmi.ac.kr/index.html). Of these 22,427 genes, we found evidence for active transcription (average expression level > 1 RPKM) of 14,740 and 14,076 genes in the cancer and paired-normal tissues, respectively.

Hierarchical clustering analysis of 3,051 genes for which expression levels showed significant variance among the 164 samples categorized the genes into three subgroups: a group with generally increased abundance in cancer (Subgroup 1; n=1,031), generally decreased abundance in cancer (Subgroup 2; n=1,232) and mixed expression patterns (Subgroup 3; n=614) (Supplementary Figure 10; Supplementary Table 14). These genes clearly differentiate cancer from normal tissues. As expected, group 2 included many genes related to normal lung function, such as surfactant genes (e.g. *SFTPA1)*.

To identify a subset of genes, which are extremely highly expressed not generally but exclusively in a small number of cancer tissues only (which could be undetected by the hierarchical clustering), we performed outlier analyses. We detected a total of 6,719 cancer outlier genes (COGs) from 87 cancer tissues (Supplementary Table 11). We narrowed this list down to more functionally relevant genes (Supplementary Figure 11), such as *GUCY2C*, *CDX2*, *HMGA2*, *ERN2* and *PAX7*, by comparing them with 934 putative cancer related genes (union of 462 genes deposited in COSMIC[3] v57 and 515 protein kinase genes[4]). Of these, *RET* protein tyrosine kinase (3rd strongest COG among protein kinases) was especially interesting, since fusion of this gene with *KIF5B* was recently identified as a transforming driver mutation in lung adenocarcinoma[1]. Of 87 cancer tissues, nine (10.3%) showed clear *RET* expression as expected[1]. Of these, five cancers expressed *RET* tyrosine kinase without evident fusion events.

III. Statistical power of this study for fusion gene detection

To calculate statistical power, we applied a simple binomial model. For example, when the frequency of any specific transforming fusion gene in lung adenocarcinoma is 0.015 (p=0.015), the probability that the fusion gene is not included in our cohort (n=200) can be calculated as follows:

Probability (# of specimens with the fusion gene (r) = 0)

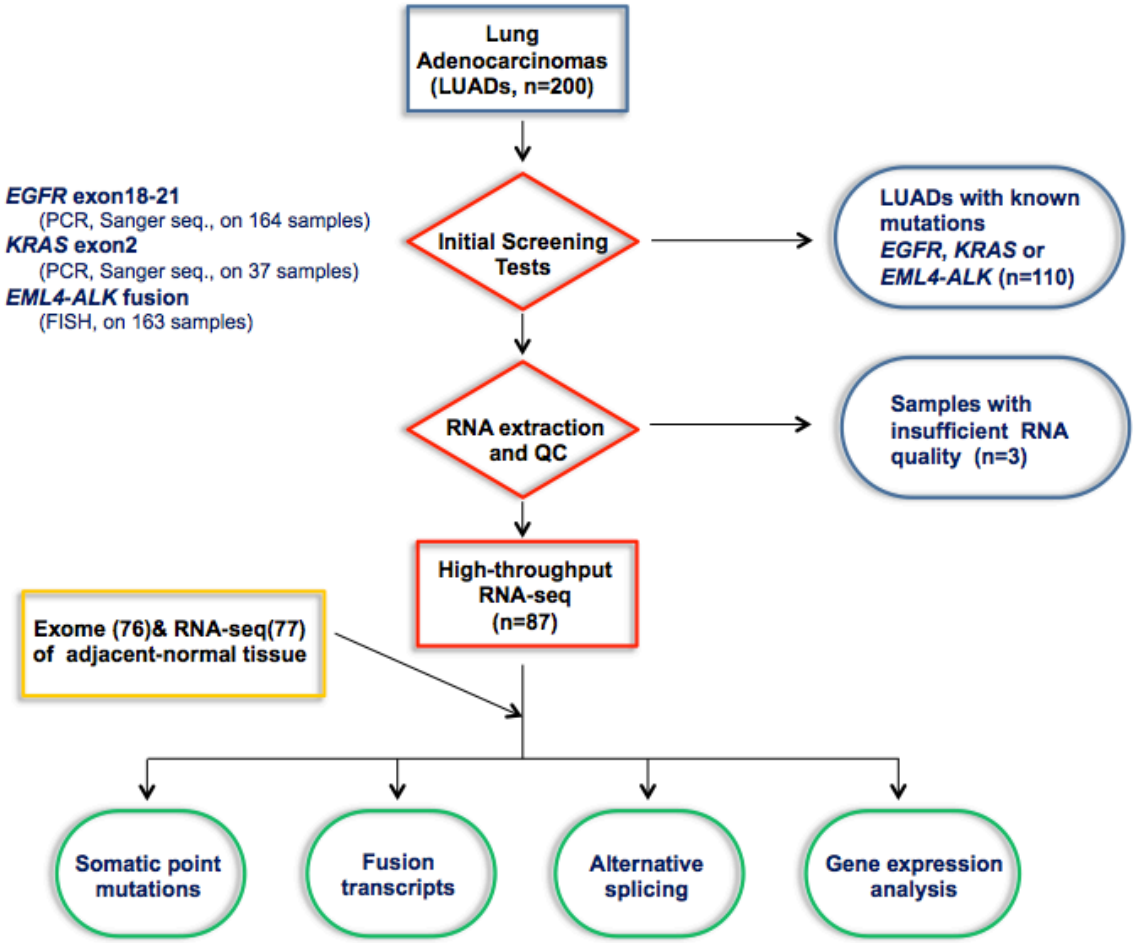
$= nCr \times p^r \times (1-p)^{(n-r)}$
$= {}_{200}C_0 \times (0.015)^0 \times (0.985)^{200}$
$= 0.0487$

Therefore, we expect the probability that the fusion gene is included in our cohort is approximately 95.1%.

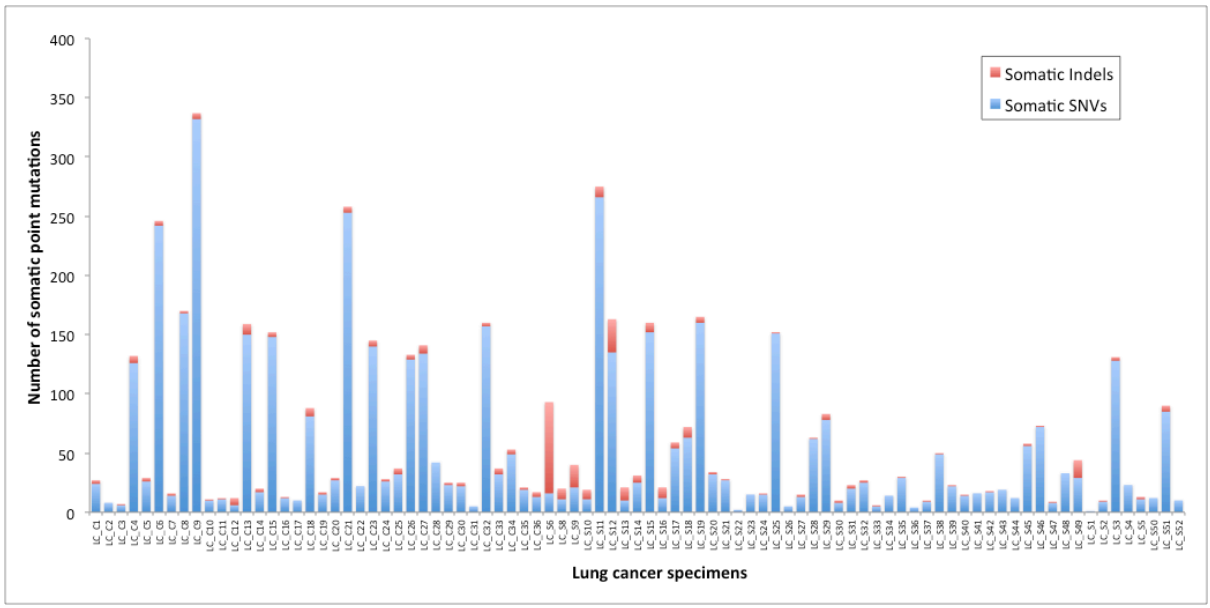
Power $= 1 - P$
$= 1 - 0.0487$
$= 0.9513$

Supplementary Figures

Supplementary Figure 1. Schematic flow chart summarizing this study.

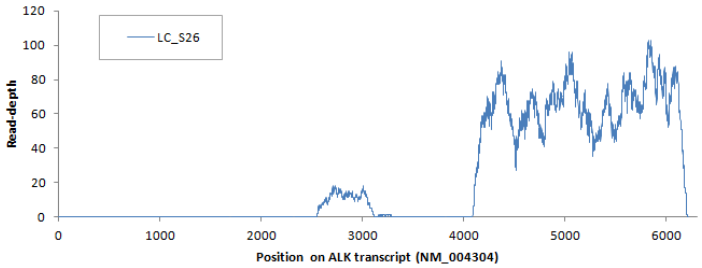


Supplementary Figure 2. Number of somatic point mutations for each lung cancer specimen.

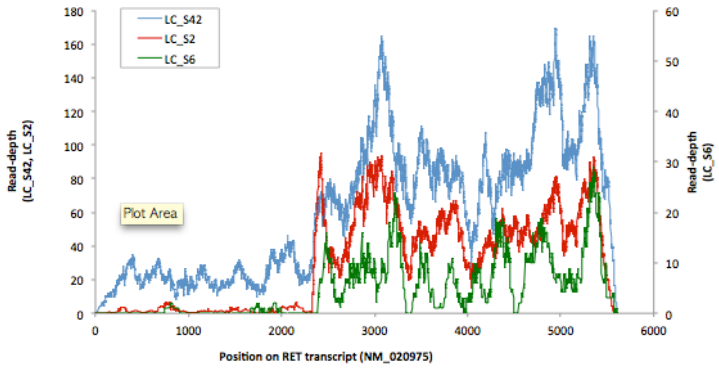


Supplementary Figure 3. Specific gene expression patterns which support the existence of *ALK*, *RET*, *ROS1* and *PDGFRA* fusion genes in a cancer tissue.

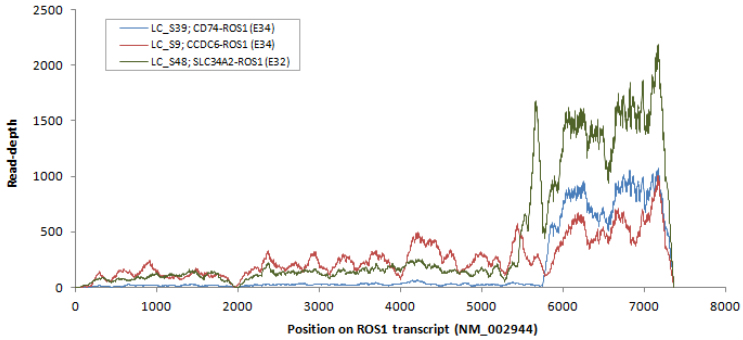
(a) *EML4-ALK*



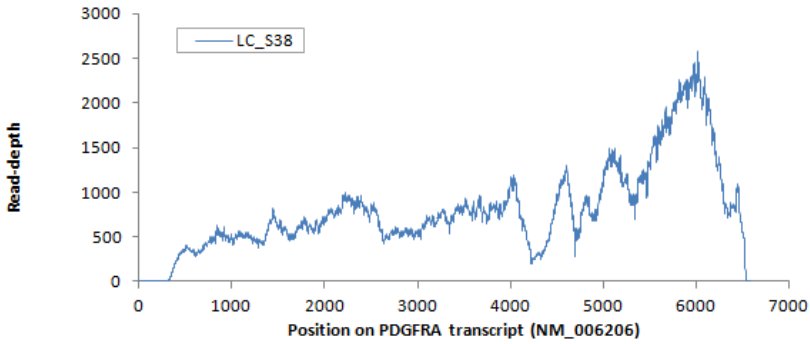
(b) *KIF5B-RET*



(c) *ROS1* fusions

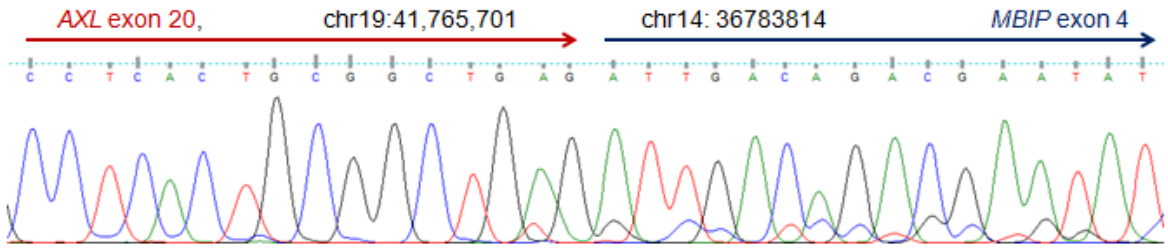


(d) *SCAF11-PDGFRA*

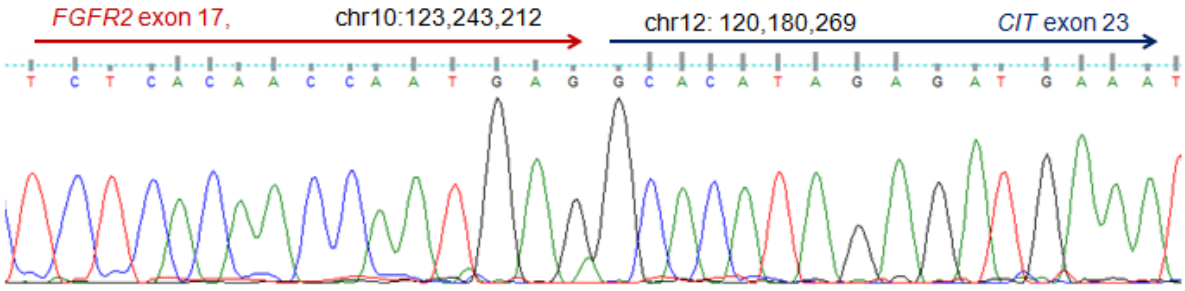


Supplementary Figure 4. Examples of fusion gene validation by cDNA PCR and Sanger sequencing.

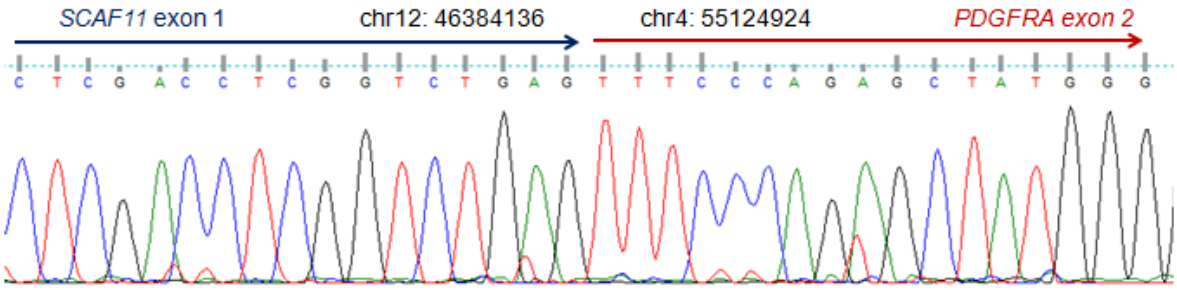
(a) *AXL-MBIP*



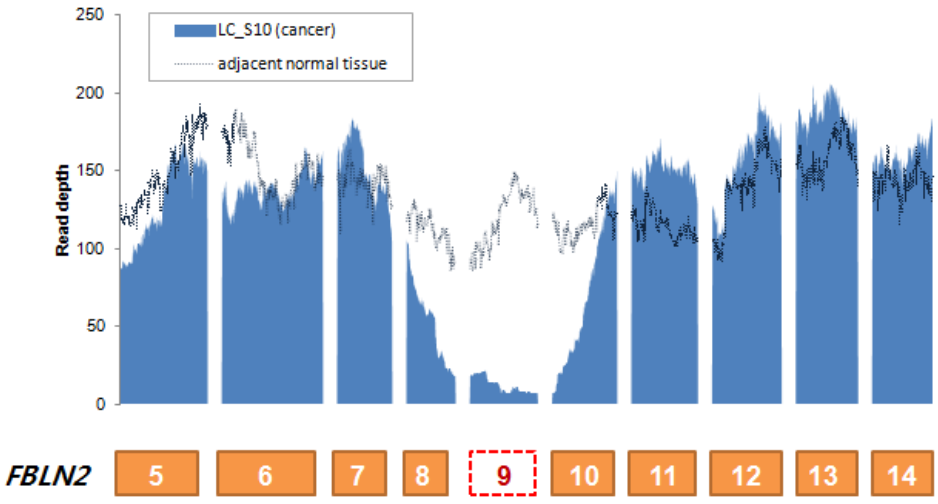
(b) *FGFR2-CIT*



(c) *SCAF11-PDGFRA*

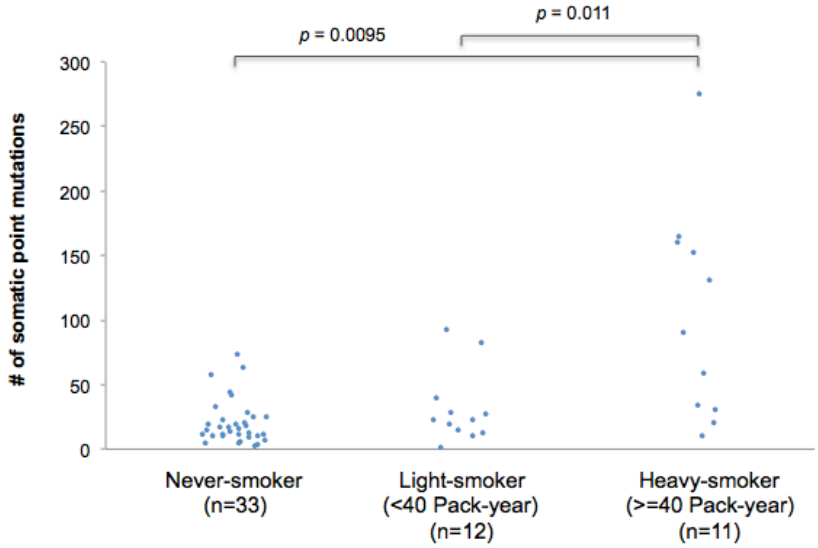


Supplementary Figure 5. Recurrent skipping of exon 9 in the *FBLN2* tumor suppressor gene in lung adenocarcinoma.



Supplementary Figure 6. The amount of smoking and number of somatic point mutations in lung cancer.

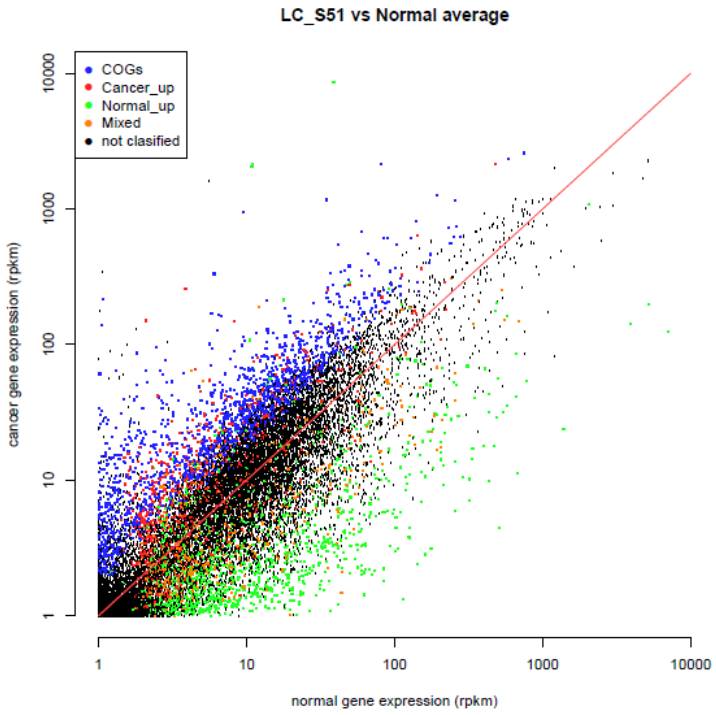
We split smokers into two-groups (heavy- and light-smokers) using a cutoff of 40 pack-years. Heavy-smokers harbored significantly more somatic point mutations than light-smokers and never-smokers (on average, 102.5, 31.3 and 20.6 somatic point mutations were detected from the cancer genomes of heavy-smokers, light-smokers and never-smokers, respectively). Of the 40 smokers studied, the information on the amount of smoking (pack year) was available for 23 cancer patients.



Supplementary Figure 7. Expression profiles of 87 lung adenocarcinomas compared to averaged gene expression levels of 77 adjacent paired normal tissues.

Seo_SuppFig7A.pdf
Seo_SuppFig7B.pdf
Seo_SuppFig7C.pdf
Seo_SuppFig7D.pdf
Seo_SuppFig7E.pdf

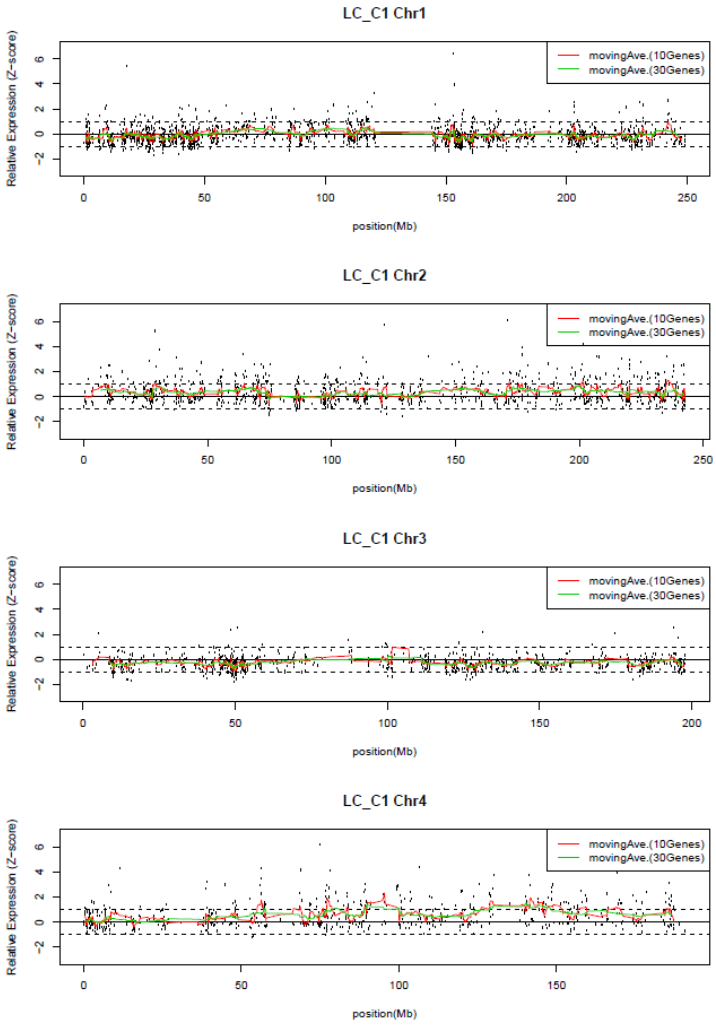
Depicted below is a preview of the full version.



Supplementary Figure 8. Relative expression of genes (Z-score) for 87 lung adenocarcinomas.

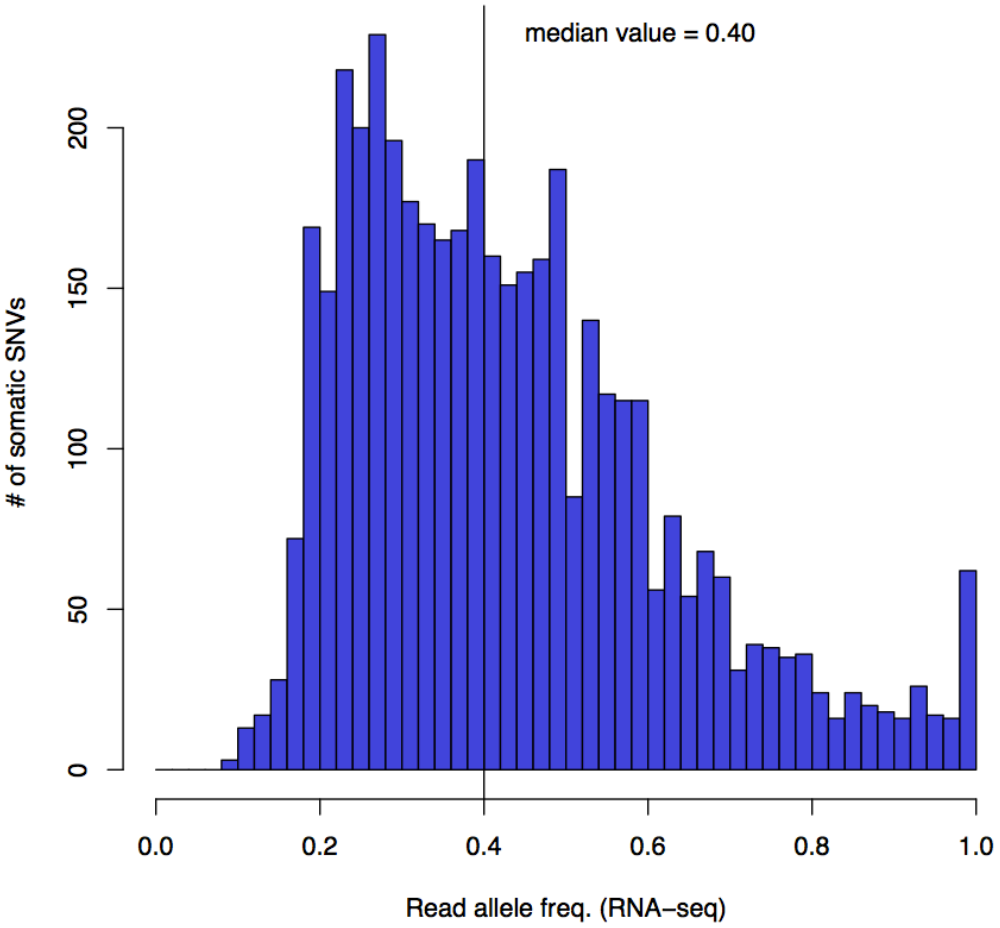
Seo_SuppFig8A.pdf
Seo_SuppFig8B.pdf
Seo_SuppFig8C.pdf
Seo_SuppFig8D.pdf
Seo_SuppFig8E.pdf
Seo_SuppFig8F.pdf
Seo_SuppFig8G.pdf
Seo_SuppFig8H.pdf
Seo_SuppFig8I.pdf

Depicted below is a preview of the full version.

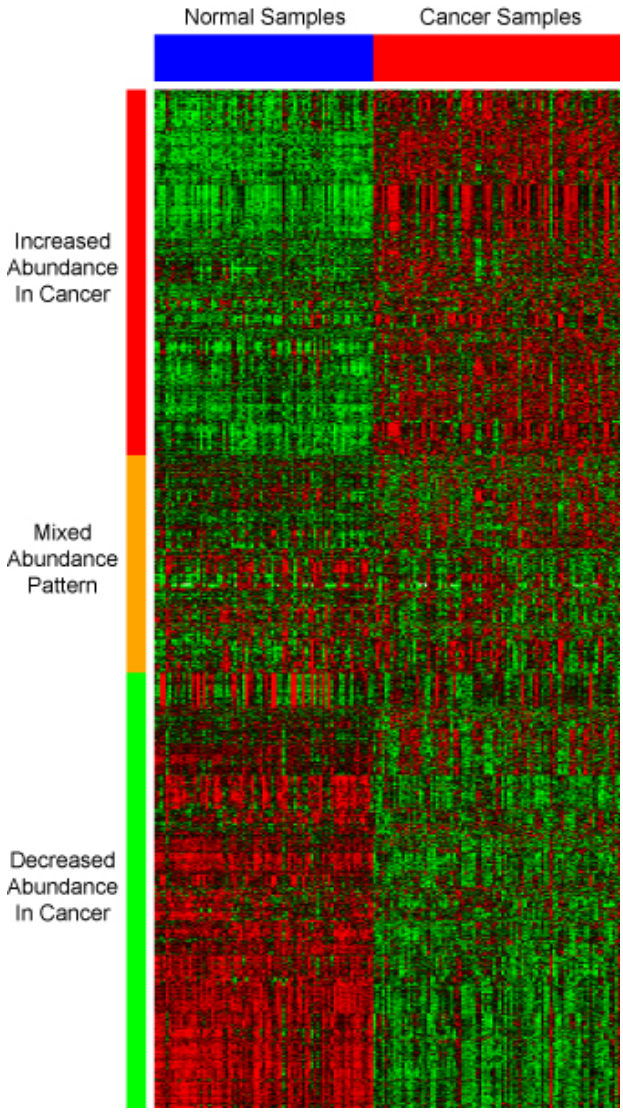


Supplementary Figure 9. Distribution of read-allele frequency of somatic SNVs identified from 87 lung adenocarcinomas for estimating tumor purity.

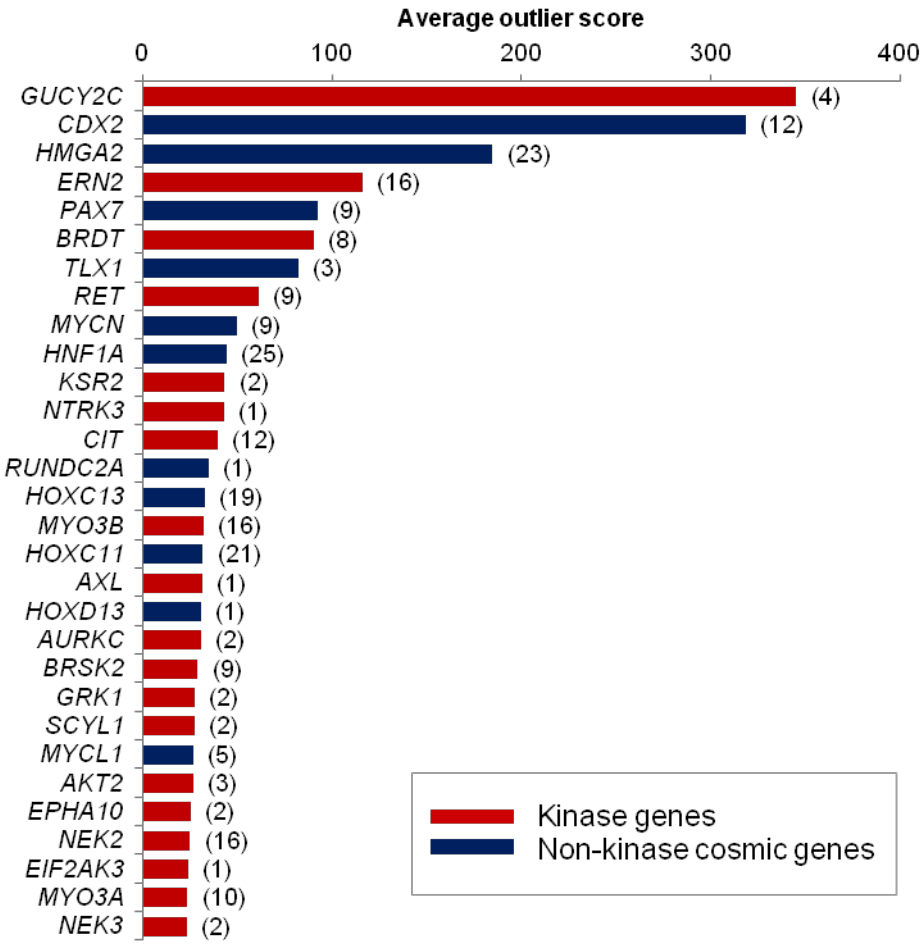
We calculated read-allele frequency (RAF) of cancer RNA sequencing for somatic SNVs. Of the 4,607 somatic SNVs identified, we assessed 4,283 SNVs with sufficient read-depth (>=10) to calculate RAFs with accuracy. The median RAF is 0.40, which suggests that the approximate purity of the major cancer clone is ~ 80%. For detailed calculations, please see Supplemental Methods #2, Ju YS *et al., Genome Research* (2012) 22(3): 436-45 (Ref. 1).



Supplementary Figure 10. Difference of global gene expression levels between primary lung adenocarcinomas (n=87) and adjacent paired-normal tissues (n=77). Vertical bar: Increased abundance in cancer (red; Cancer-up); decreased abundance in cancer (green; Cancer-down); and mixed patterns (orange). Horizontal bar: lung adenocarcinomas (red); paired-normal tissues (blue).



Supplementary Figure 11. Selected cancer outlier genes (COGs) by outlier score among protein kinases and genes deposited in COSMIC database. The number in the brackets shows the number of cancer specimens (out of 87 specimens), which carry the gene as an expression outlier.



Supplementary Tables

Supplementary Table 1. The clinical and mutational information of 200 lung adenocarcinoma patients enrolled in this study.

Seo_SuppTable1_r.xls

Depicted below is a preview of the full version.

									Genetic test at hospital						
Publication_	Age_at	Gender (1=male,2=female	Smoking_status (0=neverSmoker;1=sSmoker;2=	Stage	Cancer_RNA_Quality (0=poor;1=good;NE=NotExtracted)	Cancer_RNAseq (0=no;1=yes)	Normal_RNAseq (0=no;1=ye	Normal_ExomeSeq (0=no;1=yes)	EGFR e18-e21 (NE=notExamined,0=noMutation,des	EGFR Summary (NE=notExamined,0=noMutation,1=Mut	KRAS (NE=notExamined,wt=wildtype,description)	EML4-ALK (NE=notExamined,0=no,1=yes)	FUSION (1=No, 2=YES)	LN Mets (0=No, 1=Yes)	Potential Driver Mutations
LC_C1	54	1	1	1A	1	1	1	1	NE	NE	NE	NE	1	0	EGFR exon 19 microdeletion
LC_C2	51	1	2	2B	1	1	1	1	NE	NE	NE	NE	1	0	
LC_C3	65	2	0	1A	1	1	1	1	NE	NE	NE	NE	1	0	EGFR L858R
LC_C4	52	1	0	3B	1	1	0	0	NE	NE	NE	NE	1	1	KRAS G13D
LC_C5	68	1	3	3A	1	1	1	1	NE	NE	NE	NE	1	1	KRAS G12C
LC_C6	38	1	2	2B	1	1	0	0	NE	NE	NE	NE	1	0	
LC_C7	81	1	1	1A	1	1	1	1	NE	NE	NE	NE	1	0	
LC_C8	85	1	1	1B	1	1	0	0	NE	NE	NE	NE	1	0	
LC_C9	71	1	1	1B	1	1	1	1	NE	NE	NE	NE	1	0	
LC_C10	58	2	0	2A	1	1	1	1	NE	NE	NE	NE	1	1	EGFR exon 19 microdeletion
LC_C11	63	2	0	1A	1	1	1	1	NE	NE	NE	NE	2	0	EGFR exon 19 microdeletion
LC_C12	66	1	3	1B	1	1	1	1	NE	NE	NE	NE	2	0	EGFR L858R
LC_C13	72	1	2	2B	1	1	0	0	NE	NE	NE	NE	1	0	KRAS G12V
LC_C14	50	2	2	3A	1	1	1	1	NE	NE	NE	NE	1	1	EGFR exon 19 microdeletion
LC_C15	60	1	2	1B	1	1	0	0	NE	NE	NE	NE	2	0	MET exon 14 skipping
LC_C16	54	2	0	1A	1	1	1	1	NE	NE	NE	NE	2	0	EGFR L858R
LC_C17	64	2	0	1A	1	1	1	1	NE	NE	NE	NE	2	0	MET exon 14 skipping
LC_C18	68	2	1	1B	1	1	1	1	NE	NE	NE	NE	1	0	
LC_C19	40	2	0	4	1	1	1	1	NE	NE	NE	NE	1	0	NRAS Q61K
LC_C20	65	1	1	1A	1	1	1	1	NE	NE	NE	NE	1	0	EGFR exon 19 microdeletion

Supplementary Table 2. Summary statistics of massively parallel sequencing experiments performed in this study.

Seo_SuppTable2.xls

Depicted below is a preview of the full version.

Sample	Cancer RNA-Seq			Normal RNA-Seq			Normal whole-exome sequencing		
	# of all reads	# of aligned reads	# of aligned reads (bp)	# of all reads	# of aligned reads	# of aligned reads (bp)	# of exon-aligned reads (bp) (on-target)	Total coverage (X) (on-target)	% of exons captured (on-target)
LC_C1	77,494,486	47,470,889	4,794,559,789	45,688,760	30,723,033	3,103,026,333	2,178,031,360	47.33	98.48
LC_C2	107,215,548	63,982,363	6,462,218,663	84,044,300	56,980,300	5,755,010,300	1,401,862,209	30.46	97.76
LC_C3	83,152,108	52,466,930	5,299,159,930	73,116,392	49,425,295	4,991,954,795	1,685,976,503	36.64	97.77
LC_C4	92,656,104	53,906,608	5,444,567,408	-	-	-	-	-	-
LC_C5	78,018,264	50,172,918	5,067,464,718	64,131,160	42,834,089	4,326,242,989	2,385,520,838	51.84	98.7
LC_C6	92,725,904	56,470,698	5,703,540,498	-	-	-	-	-	-
LC_C7	91,073,182	55,521,230	5,607,644,230	71,647,042	50,101,529	5,060,254,429	2,537,090,617	55.13	98.78
LC_C8	94,479,946	59,514,604	6,010,975,004	-	-	-	-	-	-
LC_C9	74,519,406	46,621,418	4,708,763,218	85,984,544	60,050,324	6,065,082,724	1,747,785,171	37.98	97.93
LC_C10	90,885,746	53,008,298	5,353,838,098	57,904,006	38,302,163	3,868,518,463	1,379,192,060	29.97	97.45
LC_C11	72,691,004	38,326,688	3,870,995,488	60,069,120	39,730,037	4,012,733,737	1,390,026,595	30.2	97.22
LC_C12	67,503,666	35,156,178	3,550,773,978	54,398,896	37,732,658	3,810,998,458	1,218,193,317	26.47	97.1
LC_C13	63,150,052	34,157,339	3,449,891,239	-	-	-	-	-	-
LC_C14	79,154,436	43,445,511	4,387,996,611	66,595,976	45,186,779	4,563,864,679	2,019,292,321	43.88	98.31
LC_C15	120,345,132	65,067,364	6,571,803,764	-	-	-	-	-	-
LC_C16	77,198,774	39,908,939	4,030,802,839	60,758,862	40,073,167	4,047,389,867	1,405,362,092	30.54	97.68
LC_C17	75,140,824	40,726,871	4,113,413,971	66,540,006	44,966,927	4,541,659,627	2,405,400,933	52.27	98.49
LC_C18	87,207,996	46,837,953	4,730,633,253	72,409,094	51,842,913	5,236,134,213	1,572,378,178	34.17	97.98
LC_C19	79,215,512	43,241,275	4,367,368,775	75,620,208	51,101,497	5,161,251,197	2,602,075,094	56.54	98.75
LC_C20	73,241,042	41,503,737	4,191,877,437	69,621,988	47,893,928	4,837,286,728	1,334,532,896	29	96.54
LC_C21	66,062,856	35,515,426	3,587,058,026	77,754,752	51,871,683	5,239,039,983	1,602,111,576	34.81	97.13

Supplementary Table 3. List of somatic non-synonymous and coding short-indel mutations identified from transcriptome sequencing of 87 lung adenocarcinomas.

Seo_SuppTable3.xls

There are three spreadsheets inside, for somatic non-synonymous mutations, CDS short-indel mutations and gene-based dataset.

Depicted below is a preview of the full version.

												Somatic SNV candidates #var_reads_in_CaRNAseq | #wt_reads_in_CaRNAseq || #var_reads_in_NormalExomeSeq | #wt_reads_in_NormalE (-), no_variant_detected; (N), notSequenced																							
chr	pos	wt allele	var allele	annotation	wt AminoAcid	variant AminoAcid	Blosum	is_in_COSMIC(v56) (LungCA) (ND=notDetected)	is_in_COSMIC(v56) (OtherCA) (ND=notDetected)	is_in dbSNP132common	is_on segmental duplication	# samples involved	LC_C1	LC_C2	LC_C3	LC_C4	LC_C5	LC_C6	LC_C7	LC_C8	LC_C9	LC_C10	LC_C11	LC_C12	LC_C13	LC_C14	LC_C15	LC_C16	LC_C17	LC_C18	LC_C19	LC_C20	LC_C21	LC_C22
1	899,541	T	C	CDS:KLHL17	V	A	0	ND	ND	no	no	1	-	-	-	-	-	3|4||	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	906,512	G	A	CDS:PLEKH	R	Q	1	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	978,660	G	T	CDS:AGRN	G	W	-2	ND	ND	no	no	1	-	-	-	51|2|	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	1,192,679	G	A	CDS:UBE2J2	H	Y	2	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	1,271,848	G	A	CDS:DVL1	R	C	-3	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	27|3|	-	-	-	-	-	-	-	-	-	-
1	1,355,905	G	T	CDS:LOC44	L	M	2	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	1,563,912	C	T	CDS:MIB2;U	P	L	-3	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	1,564,571	G	A	CDS:MIB2;U	C	Y	-2	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	1,986,912	C	T	CDS:PRKCZ	T	M	-1	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	2,116,739	A	G	CDS:C1orf86	M	T	-1	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	2,125,258	G	A	CDS:C1orf86	P	L	-3	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	2,426,310	G	A	CDS:PLCH2	D	N	1	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	2,444,426	C	A	CDS:PANK4	R	L	-2	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	2,452,304	C	A	CDS:PANK4	C	F	-2	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	2,518,387	G	T	CDS:C1orf93	E	D	2	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	3,328,220	G	A	CDS:PRDM1	E	K	1	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	5|13|	-	-	-	-	-	-	-	-	-
1	3,380,086	G	T	CDS:ARHGE	R	S	-1	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	3,413,341	G	A	CDS:MEGF6	P	L	-3	ND	ND	no	no	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	3,548,900	C	T	CDS:WDR8	E	K	1	ND	ND	no	no	1	-	-	-	-	-	5|13|	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Supplementary Table 4. The accuracy of somatic point mutation detection.

The cancer whole-exomes of patient LC_C5 and LC_C21 were sequenced to estimate the accuracy of RNA-seq in the detection of somatic point mutations. We concluded that somatic mutations were validated when >=1 read supporting the corresponding mutant allele was detected in the exome sequences. Somatic mutations include somatic SNVs and indels. The numbers in parenthesis are for somatic indels.

Specimens	# Somatic mutations	# somatic mutations covered >= 1 x in cancer exome seq	Validated	Rate
LC_C5	29 (3)	28 (3)	24 (2)	85.7% (66.6%)
LC_C21	258 (5)	252 (5)	226 (4)	89.6% (80%)
Total	287 (8)	280 (8)	250 (6)	89.2% (75.0%)

Supplementary Table 5. Mutual exclusivity and concurrence of cancer specific alterations.

Every cancer alteration (somatic point mutations, fusion genes (fusions) and MET ES (exon skipping) in Figure 1 was considered in this testing

Seo_SuppTable5_new.xls

Depicted below is a preview of the full version.

Gene1	Gene2	wt_wt	mut_wt	wt_mut	mut_mut	Fisher's p-value	Pearson's correlation
SETD2	BRAF	Pai	2	0	2	0.00160385	0.69873836
AKAP9	JAK2	83	1	1	2	0.00238691	0.6547619
SMARCA4	MYST4	83	1	1	2	0.00238691	0.6547619
EGFR	KRAS	47	22	18	0	0.00457092	-0.2971436
ALPK3	RNF213	82	1	2	2	0.00473607	0.56007437
SETD2	BAP1	82	2	1	2	0.00473607	0.56007437
CIC	MYH9	82	2	1	2	0.00473607	0.56007437
NOTCH2	CDC42BPB	81	3	1	2	0.00783056	0.49465787
NOTCH2	BAP1	81	3	1	2	0.00783056	0.49465787
TP53	NRAS	65	19	0	3	0.01452899	0.32483762
TP53	JAK2	65	19	0	3	0.01452899	0.32483762
NOTCH2	RNF213	80	3	2	2	0.01529048	0.41740702
KRAS	SETD2	68	15	1	3	0.02666972	0.29432826
CDKN2A	STK11	84	1	1	1	0.0457097	0.48823529
TP53	CIC	64	19	1	3	0.04825699	0.25107822
EGFR	fusions	55	22	10	0	0.05940816	-0.209657
MLL2	BRAF	83	2	1	1	0.06816359	0.39134545
CDKN2A	ARID1A	83	1	2	1	0.06816359	0.39134545
MET	SMARCA4	83	1	2	1	0.06816359	0.39134545
MET	LMTK2	83	1	2	1	0.06816359	0.39134545
MET	MYST4	83	1	2	1	0.06816359	0.39134545
MET	CDC42BPB	83	1	2	1	0.06816359	0.39134545
MET	MAP3K4	83	1	2	1	0.06816359	0.39134545
NACA	LMTK2	82	2	2	1	0.10105194	0.30952381

Supplementary Table 6. List of 45 fusion genes identified from transcriptome sequencing of 87 lung adenocarcinomas.

Seo_SuppTable6_r.xls

Depicted below is a preview of the full version.

Index	Donor Gene	Acceptor Gene	Chromosome (Donor;Acceptor)	# samples observed	Distance (Mb)	Sample observed	Donor Breakpoint (RNA)	Donor protein sequence near breakpoint	Acceptor Breakpoint (RNA)	Acceptor protein sequence near breakpoint	Co-occurrence of other driver mutation
1	*EML4*	*ALK*	chr2;chr2	1	12.252	LC_S26	chr2:42522656]	TPGKGPK+1nt	chr2:29446394]	2nt+YRRKHQE	-
2	*KIF5B*	*RET*	chr10;chr10	4	11.227	LC_S2	chr10:[32317356	NNDVK	chr10:[43612032	EDPKWEF	-
2	*KIF5B*	*RET*	chr10;chr10	4	11.227	LC_S6	chr10:[32306980	KVHKQ	chr10:[43612032	EDPKWEF	-
2	*KIF5B*	*RET*	chr10;chr10	4	11.227	LC_S42	chr10:[32317356	NNDVK	chr10:[43612032	EDPKWEF	-
3	*CD74*	*ROS1*	chr5;chr6	1	Interchromosomal	LC_S39	chr5:[149784243	DAPPK+1nt	chr6:117645578]	2nt+DFWIP	-
4	*SLC34A2*	*ROS1*	chr4;chr6	1	Interchromosomal	LC_S48	chr4:25678324]	SREAQ+1nt	chr6:117650609]	2nt+GVPNK	-
5	*CCDC6*	*ROS1*	chr10;chr6	1	Interchromosomal	LC_S9	chr10:[61572393	AAQLQ+1nt	chr6:117642557]	2nt+WHRRL	-
6	*SCAF11*	*PDGFRA*	chr12;chr4	1	Interchromosomal	LC_S38	chr12:[46384136	5UTR	chr4:[55124924	5UTR, in-frame	-
7	*FGFR2*	*CIT*	chr10;chr12	1	Interchromosomal	LC_S13	chr10:[123243212	LTLTTNE	chr12:120180269]	AHRDEIQ	-
8	*AXL*	*MBIP*	chr19;chr14	1	Interchromosomal	LC_S23	chr19:41765701]	LTAAE	chr14:36783814]	IDRRI	-
9	*APLP2*	*TNFSF11*	chr11;chr13	1	Interchromosomal	LC_C15	chr11:130000061]	AAQMKSQ	chr13:[43174888	ELQHIVG	-
10	*MAP4K3*	*PRKCE*	chr2;chr2	1	6.215	LC_S26	chr2:[39664033	TYGDVYK	chr2:[46070139	IDLEPEGR	EML4-ALK
11	*BCAS3*	*MAP3K3*	chr17;chr17	1	2.23	LC_S20	chr17:59161925]	TVIDAAS+1nt	chr17:61710041]	2nt+EQEALNS	-
12	*KRAS*	*CDH13*	chr12;chr16	1	Interchromosomal	LC_C12	chr12:[25378548	TSAKTRQ	chr16:[83158990	DIFKFAR	EGFR L858R
13	*ZFYVE9*	*CGA*	chr1;chr6	1	Interchromosomal	LC_C25	chr1:52803606]	DKNVSK+2nt	chr6:87797925]	5UTR, in-frame	-
14	*ERBB2IP*	*MAST4*	chr5;chr5	1	0.515	LC_S19	chr5:65372777]	QPGDKIIQ	chr5:[66400194	ATAQMEER	-
15	*TPD52L1*	*TRMT11*	chr6;chr6	1	0.723	LC_S42	chr6:125569529]	SKKFGDM+2nt	chr6:[126342306	1nt+YTEEMVP	KIF5B-RET
16	*TXNRD1*	*GPR133*	chr12;chr12	1	26.694	LC_C17	chr12:104733051]	IHPVCAE	chr12:[131561346	TRKQHS	MET exon 14 skipping
17	*SRSF4*	*SNRNP40*	chr1;chr1	1	2.224	LC_S29	chr1:[29485886	SRCSWQDLK	chr1:31744346]	VWDLRQN	-
18	*EDA*	*MID1*	chrX;chrX	1	57.984	LC_S51	chrX:68836548]	DSQDGHQ	chrX:10463731]	VNASRQE	-
19	*HYOU1*	*C11orf93*	chr11;chr11	1	7.736	LC_S11	chr11:[118921747	SGVLSLDR	chr11:[111175653	5 UTR	-
20	*SLC16A7*	*MUCL1*	chr12;chr12	1	4.831	LC_C36	chr12:60098799]	LAVMYAG+1nt	chr12:[55248900	2nt+NPTTAAPAD	EGFR microdeletion

Supplementary Table 7. List of 43 pairs of primers used for PCR and Sanger sequencing validation of fusion genes.

Seo_SuppTable7_r.xls

Depicted below is a preview of the full version.

index	Donor Gene	Acceptor Gene	Forward Primer Name	Forward Primer Sequence	Reverse Primer Name	Reverse Primer Sequence	Remark
1	KIF5B	RET	GF1_KIF5B:RET_F	TAAGGAAATGACCAACCACCAG	GF1_KIF5B:RET_R	CCTTGACCACTTTTCCAAATTC	Validated
2	KRAS	CDH13	GF2_KRAS:CDH13_F	GGAAATAAATGTGATTTGCCTTC	GF2_KRAS:CDH13_R	AAGGCTGTCTCTGATTCTCTGG	Validated
3	APLP2	TNFSF11	GF3_APLP2:TNFSF11_F	TGCTGAGAACAAAGATCGCTTA	GF3_APLP2:TNFSF11_R	TGTCGGTGGCATTAATAGTGAG	Validated
4	ZFYVE9	CGA	GF4_ZFYVE9:CGA_F	ACTGCAGAGAACATGGATTCCT	GF4_ZFYVE9:CGA_R	GAATGGAGAACATGCAGAAACA	Validated
5	CCDC6	ROS1	GF5_CCDC6:ROS1_F	CCTGCAGGAAAAATTAGACCAG	GF5_CCDC6:ROS1_R	AGCTCAGCCAACTCTTTGTCTT	Validated
6	FGFR2	CIT	GF6_FGFR2:CIT_F	ACATGATGATGAGGGACTGTTG	GF6_FGFR2:CIT_R	ACAGCTGTTACGAAGAGCATCA	Validated
7	AXL	MBIP	GF7_AXL:MBIP_F	GCCTGACGAAATCCTCTATGTC	GF7_AXL:MBIP_R	CAAAATTCCCTGACGTTGTTTT	Validated
8	SCAF11	PDGFRA	GF8_SCAF11:PDGFRA_F	CAGCGGAGTCAGTGTCCTAGAG	GF8_SCAF11:PDGFRA_R	TGAGAAGACAGCCTAAGACCAG	Validated
9	CD74	ROS1	GF9_CD74:ROS1_F	GTCTTTGAGAGCTGGATGCAC	GF9_CD74:ROS1_R	AGCTCAGCCAACTCTTTGTCTT	Validated
10	SLC34A2	ROS1	GF10_SLC34A2:ROS1_F	ATGCCGTCGTCTCCAAGTTC	GF10_SLC34A2:ROS1_R	ATCTTCAGCTTTCTCCCACTGT	Validated
11	TXNRD1	GPR133	GF11_TXNRD1:GPR133_F	TCCAAATGCTGGAGAAGTTACA	GF11_TXNRD1:GPR133_R	AGTACACGAAGACTCGGTTGCT	Validated
12	EML4	ALK	GF12_EML4:ALK_F	GCCAAAATTTGTGCAGTGTTTA	GF12_EML4:ALK_R	GGAGCTTGCTCAGCTTGTACTC	Validated
13	HYOU1	C11orf93	GF13_HYOU1:C11orf93_F	CCAGAATCTGACCACAGTGAAG	GF13_HYOU1:C11orf93_R	AGAAGATGGTGCAACTGGGTCT	Validated
14	MAP4K3	PRKCE	GF14_MAP4K3:PRKCE_F	AGGAGGACTTCGAGCTGATTC	GF14_MAP4K3:PRKCE_R	ACGACCCTGAGAGATCGATGA	Validated
15	RBM14	FGF3	GF15_RBM14:FGF3_F	CCAAGGCCTCTTAATACTTGGA	GF15_RBM14:FGF3_R	CATAGAGTCGTCCCCTCTTGTT	Validated
16	BCAS3	MAP3K3	GF16_BCAS3:MAP3K3_F	CATCCCGTCCAGTCTCTGAT	GF16_BCAS3:MAP3K3_R	CTGCCTATTTGAGTGACCTGTG	Validated
17	SRSF4	SNRNP40	GF17_SRSF4:SNRNP40_F	GAAGTGGCCGAGATAAATATGG	GF17_SRSF4:SNRNP40_R	TAAACTCAGGCCAGTCACTGAA	Validated
18	UBR4	ATP13A2	GF18_UBR4:ATP13A2_F	ACCCTTTCTCTACCTGTGTTGG	GF18_UBR4:ATP13A2_R	AGCTGAGGGGATCTATTGATGT	Validated
19	TTC19	ATPAF2	GF19_TTC19:ATPAF2_F	CGCTTTGATGAGGCCTATATTT	GF19_TTC19:ATPAF2_R	CTGTGTGATGCTGACATTCTGA	Validated
20	TPD52L1	TRMT11	GF20_TPD52L1:TRMT11_F	GAAAACACATGAAACCCTGAGTC	GF20_TPD52L1:TRMT11_R	ATGTGTGACTGGAAAGCTTCTG	Validated
21	IGSF3	MAN1A2	GF21_IGSF3:MAN1A2_F	CTGACCAGGGCGAATTCTACT	GF21_IGSF3:MAN1A2_R	TCTTGCCTCATGGTCTGTTTTA	Validated
22	ERBB2IP	MAST4	GF22_ERBB2IP:MAST4_F	AACAAGGGTACAACCTGAAGGA	GF22_ERBB2IP:MAST4_R	TCAAGGAAGTATCGTGAGGTGA	Validated
23	XAF1	FAM64A	GF23_XAF1:FAM64A_F	GGAGCTCCACGAGTCCTACTGT	GF23_XAF1:FAM64A_R	AGAGGTCTCCTGATGGCTGAC	Validated
24	MIER2	ITGB1BP3	GF24_MIER2:ITGB1BP3_F	AGATCATGGTGGGACCTCAGT	GF24_MIER2:ITGB1BP3_R	AGCAGCGAGTTCTGAATGTCTT	Validated
25	SLC16A7	MUCL1	GF25_SLC16A7:MUCL1_F	GTGGTTGGAGCAGCTTTTATCT	GF25_SLC16A7:MUCL1_R	TCATCATCAGCAGGACCAGTAG	Validated
26	ITGB1BP3	DNM2	GF26_ITGB1BP3:DNM2_F	CCTGGAAGACATTCAGAACTCG	GF26_ITGB1BP3:DNM2_R	TTTGAGAAGATGAGCTGCAGAA	Validated
27	ARHGEF16	TCTEX1D4	GF27_ARHGEF16:TCTEX1D4	GCATGGAGCAGATGTACACG	GF27_ARHGEF16:TCTEX1D4_R	TGTGTTTTAGAACAAGTGGATCAGA	Validated
28	CMBL	C8orf38	GF29_CMBL:C8orf38_F	CTCTCCCAGGAGGCTACGACT	GF29_CMBL:C8orf38_R	TGAGCCAGTTCCACATTAAAGG	Validated
29	EDA	MID1	GF30_EDA:MID1_F	TGACGTTGTGCTGCTACCTAGA	GF30_EDA:MID1_R	ATCTGTCGTCTTTGCTGAATGA	Validated
30	H19	CALR	GF28_H19:CALR_F	CACCGCAATTCATTTAGTAGCA	GF28_H19:CALR_R	GCCTCTCTACAGCTCGTCCTT	Failed

Supplementary Table 8. Mutually exclusivity of protein tyrosine kinase fusion genes and MET exon 14 skipping with known driver mutations of lung adenocarcinomas.

Gene1	Gene2	wt_wt	mut_wt	wt_mut	mut_mut	Fisher's p-value	Pearson's correlation
canonical point driver mutations *	all PTK fusions (EML4-ALK, KIF5B-RET, ROS1 fusions, FGFR2-CIT, AXL-MBIP, and SCAF11-PDGFRA)	30	47	10	0	2.12E-04	-0.391
canonical driver mutations **	novel 4 fusions (CCDC6-ROS1, FGFR2-CIT, AXL-MBIP and SCAF11-PDGFRA)	30	53	4	0	2.08E-02	-0.274
canonical driver mutations **	MET exon 14 skipping	31	53	3	0	0.0565	-0.236

--- Note ---

PTK; protein tyrosine kinase

* Canonical point driver mutations

Total 47 specimens.
EGFR (22), KRAS (18), NRAS (3), BRAF (1), MET (1), CTNNB1 (1), PIK3CA alone (1)

** Canonical driver mutations

Total 53 specimens
Canonical point driver mutations (47), EML4-ALK (1), KIF5B-RET (3), CD74-ROS1 (1) and SLC34A2-ROS1 (1)

Supplementary Table 9. List of 17 recurrent exon-skipping events identified from transcriptome sequencing of 87 lung adenocarcinomas.

Seo_SuppTable9.xls

Depicted below is a preview of the full version.

index	gene	index	Upstream Exon	Downstream Exon	# samples observed	Length of Skipped Exon (bp)	Samples observed
1	*LMO7*	NM_005358	9	11	23	30	LC_C4,LC_C6,LC_C11,LC_C18,LC_C19, LC_C21,LC_C22,LC_C25,LC_C28, LC_C33,LC_S12,LC_S17,LC_S23,LC_S31, LC_S32,LC_S35,LC_S36,LC_S38,LC_S43, LC_S45,LC_S2,LC_S3,LC_S51,
2	*H2AFY*	NM_138609	5	7	5	91	LC_C4,LC_C7,LC_S20,LC_S21,LC_S25,
3	*MET*	NM_000245	13	15	3	141	LC_C15,LC_C17,LC_S4,
4	*FBLN2*	NM_001165035	8	10	3	141	LC_C32,LC_S10,LC_S11,
5	*RIPK2*	NM_003821	1	3	3	154	LC_S8,LC_S14,LC_S3,
6	*CASK*	NM_001126055	17	19	3	36	LC_S38,LC_S45,LC_S51,
7	*CELF2*	NM_001025076	5	7	3	80	LC_C18,LC_S6,LC_S10,
8	*WDFY3*	NM_014991	44	46	3	51	LC_C21,LC_S27,LC_S43,
9	*SLIT2*	NM_004787	14	16	3	24	LC_C22,LC_S42,LC_S3,
10	*OPN3*	NM_014322	1	3	3	320	LC_S6,LC_S8,LC_S45,
11	*SLC33A1*	NM_001190992	1	3	3	188	LC_S6,LC_S9,LC_S9,
12	*EPB41L2*	NM_001431	12	14	2	63	LC_C4,LC_C21,
13	*ORC4*	NM_181741	4	10	2	537	LC_S8,LC_S10,
14	*PKD2*	NM_000297	5	7	2	229	LC_S9,LC_S27,
15	*SETDB2*	NM_031915	1	3	2	16	LC_S9,LC_S12,
16	*YME1L1*	NM_014263	3	5	2	99	LC_S10,LC_S14,
17	*SLC23A2*	NM_005116	6	8	2	89	LC_S45,LC_S49,

Supplementary Table 10. Expression map of 87 cancer and 77 adjacent paired-normal tissues represented in RPKM values on all reference genes.

Seo_SuppTable10.xls

There are two spreadsheets inside, for gene expression levels of cancers and normal tissues..

Depicted below is a preview of the full version.

Gene	index	chr	start	stop	strand	conding length (bp)	LC_C1	LC_C2	LC_C3	LC_C4	LC_C5	LC_C6	LC_C7	LC_C8	LC_C9
WASH7P	NR_024540	1	14,361	29,370	-	1,769	1.83	2.13	4.26	3.26	1.43	2.67	2.82	3.13	3.39
FAM138A	NR_026818	1	34,610	36,081	-	1,130	0	0	0	0	0	0	0	0	0
FAM138F	NR_026820	1	34,610	36,081	-	1,130	0	0	0	0	0	0	0	0	0
OR4F5	NM_001005484	1	69,090	70,008	+	918	0	0	0	0	0	0	0	0	0
LOC100132287	NR_028322	1	323,891	328,581	+	4,370	0.05	0.02	0.04	0	0.02	0.12	0.07	0.01	0.14
LOC100132062	NR_028325	1	323,891	328,581	+	4,370	0.05	0.02	0.04	0	0.02	0.12	0.07	0.01	0.14
LOC100133331	NR_028327	1	323,891	328,581	+	4,273	0.05	0.02	0.04	0	0.02	0.12	0.05	0	0.13
OR4F29	NM_001005221	1	367,658	368,597	+	939	0	0	0	0	0	0	0	0	0
OR4F3	NM_001005224	1	367,658	368,597	+	939	0	0	0	0	0	0	0	0	0
OR4F16	NM_001005277	1	367,658	368,597	+	939	0	0	0	0	0	0	0	0	0
OR4F29	NM_001005221	1	621,095	622,034	-	939	0	0	0	0	0	0	0	0	0
OR4F3	NM_001005224	1	621,095	622,034	-	939	0	0	0	0	0	0	0	0	0
OR4F16	NM_001005277	1	621,095	622,034	-	939	0	0	0	0	0	0	0	0	0
LOC100133331	NR_028327	1	661,138	665,731	-	4,273	2.35	2.65	1.12	2.22	1.39	1.93	3.48	1.97	1.23
LOC100288069	NR_033908	1	700,244	714,068	-	1,371	2.91	2.33	0.18	2.26	1.28	1.64	3.25	1.74	1.39
NCRNA00115	NR_024321	1	761,585	762,902	-	1,317	1.93	2.75	1.47	3.41	2.18	2.59	2.39	3.24	3.11
LOC643837	NR_015368	1	763,063	789,740	+	1,543	4.57	2.71	2.82	2.71	7.7	3.08	3.46	6.62	9.08
FAM41C	NR_027055	1	803,450	812,182	-	1,706	0.54	0.87	0.41	0.28	0.37	0.75	0.84	0.76	0.37
FLJ39609	NR_026874	1	852,952	854,817	-	496	0.21	0	0.05	0	0.07	0	0	0.05	0.1
SAMD11	NM_152486	1	861,120	879,961	+	2,554	4.03	1.85	3.82	3.26	3.63	2.76	3.14	4.04	2.95
NOC2L	NM_015658	1	879,582	894,679	-	2,800	23.14	15.26	18.26	43.48	19.96	19.59	36.53	28.35	26.97
KLHL17	NM_198317	1	895,966	901,099	+	2,564	2.81	1.81	2.69	2.68	2.38	2.32	4.08	1.56	1.22
PLEKHN1	NM_001160184	1	901,876	910,484	+	2,295	2.42	0.98	0.61	1.17	1.04	0.9	0.12	0.57	1.65
PLEKHN1	NM_032129	1	901,876	910,484	+	2,400	2.48	0.96	0.64	1.2	1.09	0.91	0.13	0.61	1.71
C1orf170	NR_027693	1	910,578	917,473	-	3,040	0.54	0.2	0.35	0.23	0.23	0.14	0.02	0.12	0.26

Supplementary Table 11. List of 6,719 cancer outlier genes (COGs) identified from transcriptome sequencing of 87 lung adenocarcinomas and 77 adjacent paired-normal tissues.

Seo_SuppTable11.xls

Depicted below is a preview of the full version.

Gene	# Samples involved	Outlier score	Average Score (per sample)	is_in_COSMIC (v57)	is_kinase	is_CancerUp (Heatmap)	LC_C1	LC_C2	LC_C3	LC_C4	LC_C5	LC_C6	LC_C7	LC_C8	LC_C9	LC_C10	LC_C11	LC_C12
APOA2	5	114801.2	22960.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S100A7	12	25327.5	2110.6	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
F2	4	4033.0	1008.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CRABP1	20	16110.7	805.5	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
CALML3	8	5521.3	690.2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
SERPINC1	1	531.1	531.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASCL1	8	4014.2	501.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KRT6B	15	6409.8	427.3	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
ASGR1	1	412.4	412.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DHRS2	10	4044.5	404.5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
DSG3	7	2735.0	390.7	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
GSTM1	1	378.7	378.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PRAC	2	736.7	368.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SERPINA10	1	353.9	353.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GUCY2C	4	1377.7	344.4	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
VIL1	27	9238.4	342.2	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0
CDX2	12	3811.3	317.6	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
SPRR1B	26	7838.0	301.5	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
SERPINA4	11	3150.9	286.4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
GPX2	30	8307.1	276.9	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0
ANGPTL3	3	785.0	261.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PRAP1	16	4183.1	261.4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
SPRR2D	20	4715.3	235.8	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0
COL11A1	18	4183.4	232.4	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0

Supplementary Table 12. Correlation between the lymph node metastasis and somatic mutations in primary lung cancer tissues.

* Driver mutations: canonical point mutations (n=47; *EGFR*, *KRAS*, *NRAS*, *PIK3CA*, *BRAF*, *MET, CTNNB1*), protein tyrosine kinase fusion genes (n=10; *EML4-ALK*, *KIF5B-RET*, *ROS1* fusions, *FGFR2-CIT*, *AXL-MBIP*, *SCAF11-PDGFRA*) and *MET* exon 14 skipping (n=3)

(a) Between cancer tissues with and without canonical driver mutations

	with LN mets	No LN mets	Subtotals
Driver known cancers (proportion)	15 (0.250)	45 (0.750)	60
Driver unknown cancers (proportion)	3 (0.111)	24 (0.889)	27
Subtotals	18	69	87

Chi-square *p*-value = 0.2327

<Logistic regression>

LN Mets and TP53 mutation, Driver mutation, Age, Gender, Smoking, Cancer Stage

Factors	Estimate	Std. Error	z-value	P-value
TP53 mutation	1.85618	1.05337	1.762	0.07847
Driver mutation	1.49805	1.05577	1.419	0.155925
Age	0.05830	0.05720	1.019	0.308099
Gender	0.05053	1.17829	0.043	0.965797
Smoking	1.26906	1.22444	1.036	0.300000
Cancer stage	2.69636	0.73290	3.679	0.000234
(intercept)	-13.13204	5.47495	-2.399	0.016459

(b) Between cancer tissues with a combination of canonical driver and *TP53* mutations and other groups

	with LN mets	No LN mets	Subtotals
Cancers with driver and TP53 mutations (proportion)	7 (0.438)	9 (0.563)	16
Others (proportion)	11 (0.155)	60 (0.845)	71
Subtotals	18	69	87

Chi-square *p*-value = 0.012

<Logistic regression>

LN Mets and (TP53 mutation x Driver mutation), Age, Gender, Smoking, Cancer Stage

Factors	Estimate	Std. Error	z-value	P-value
Driver x *TP53* mut.	2.82963	1.18588	2.386	0.017028
Age	0.05838	0.05910	0.988	0.323172
Gender	-0.17579	1.16010	-0.152	0.879560
Smoking	1.52051	1.26227	1.205	0.228363
Cancer stage	2.61576	0.69802	3.747	0.000179
(intercept)	-12.04942	5.12642	-2.350	0.0118751

Supplementary Table 13. List of specific aberrations of note for 25 cancer tissues which do not harbor canonical driver mutations.

Seo_SuppTable13_r.xls

Depicted below is a preview of the full version

Index	Age	Gender	Stage	Smoking_status (0=neverSmoker; 1=smoker;2=current_smoker;3=unknown)	#COGs	Length of total JRBs (Mb)	Remarkable gene	type	Mutation position	Protein ID	AminoAcid Change	Prediction (SIFT)	KEGG pathway
LC_C2	51	M	2B	2	67	11.823	*ALK*	overexpression (fusion)					
LC_C6	38	M	2B	2	180	25.547	*EPHA2*	point mutat	chr1:16456865,C>T	ENSP00000:	C842Y	NA	Axon guidance
							JAK2	point mutat	chr9:5080672,T>G	ENSP00000:	L808W	N/A	Measles;Adipocytokine signaling pathway
							CDK9	point mutat	chr9:130548447,C>T	ENSP00000:	S7L	N/A	Transcriptional misregulation in cancer
							MEN1	point mutat	chr11:64572131,C>T	ENSP00000:	G503D	N/A	Transcriptional misregulation in cancer
							TP53	point mutat	chr17:7577141,C>A	ENSP00000:	G266V	N/A	Measles;Hepatitis C;MAPK signaling pat
							GNAS	point mutat	chr20:57484420,C>T	ENSP00000:	R186C	DAMAGING	Vascular smooth muscle contraction;GnR
							NEK2	overexpression					
LC_C7	81	M	1A	1	740	681.076	*MYCN*	overexpression					Transcriptional misregulation in cancer
							FGFR1	overexpression					Adherens junction;MAPK signaling pathw
LC_C8	85	M	1B	1	463	162.305	*EPHA2*	point mutat	chr1:16460020,T>G	ENSP00000:	E607A	N/A	Axon guidance
							JAK1	point mutat	chr1:65307187,T>C	ENSP00000:	Q834R	N/A	Measles;Influenza A;Hepatitis C;Leishma
							NOTCH2	point mutat	chr1:120502080,C>T	ENSP00000:	C654Y	DAMAGING	Dorso-ventral axis formation;Notch signa
							TP53	point mutat	chr17:7577536,T>C	ENSP00000:	R249G	N/A	Measles;Hepatitis C;MAPK signaling pat
LC_C9	71	M	1B	1	389	384.84	*ELK4*	point mutat	chr1:205585730,C>A	ENSP00000:	G414W	DAMAGING	MAPK signaling pathway;Transcriptional
							APC	point mutat	chr5:112179479,G>T	uc011cvt.2	A2712S	Not scored	Wnt signaling pathway;HTLV-I infection;F
							JAK2	point mutat	chr9:5054638,G>T	ENSP00000:	R230S	DAMAGING	Measles;Adipocytokine signaling pathway
							TP53	point mutat	chr17:7577538,C>A	ENSP00000:	R248L	DAMAGING	Measles;Hepatitis C;MAPK signaling pat
							NF1	point mutat	chr17:29665096,G>A	ENSP00000:	G2232E	DAMAGING	MAPK signaling pathway
							SMARCA4	point mutat	chr19:11134270,G>T	ENSP00000:	R979L	DAMAGING	
							CHEK1	point mutat	chr11:125499190,T>G	ENSP00000:	V118G	DAMAGING	Cell cycle;p53 signaling pathway;HTLV-I

Supplementary Table 14. Three subgroups of genes in differentially expressed gene analysis.

Seo_SuppTable14.xls

There are three spreadsheets inside, for increased abundance in cancers (Cancer-UP), decreased abundance in cancers (Cancer-DOWN) and mixed patterns.

Depicted below is a preview of the full version.

Gene symbol	Protein description
ABCA7	ATP-binding cassette sub-family A member 7
ABCB6	ATP-binding cassette sub-family B member 6, mitochondrial
ABCB9	ATP-binding cassette sub-family B member 9 isoform 5
ABCC3	ATP-binding cassette, sub-family C (CFTR/MRP)
ABCC4	multidrug resistance-associated protein 4
ABHD11	abhydrolase domain-containing protein 11
ABL2	Abelson tyrosine-protein kinase 2 isoform i
ABTB2	ankyrin repeat and BTB (POZ) domain containing
ACAD8	acyl-CoA dehydrogenase family, member 8
ACHE	uncharacterized protein LOC606473 precursor
ACOT11	acyl-coenzyme A thioesterase 11
ADAM28	ADAM metallopeptidase domain 28 precursor
ADAM8	a disintegrin and metalloproteinase domain 8 precursor
ADAMDEC1	ADAM DEC1 precursor
ADCK4	aarF domain containing kinase 4

Cancer-UP | Cancer-DOWN | MIXED

References

1. Ju, Y.S. et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res* **22**, 436-45 (2012).
2. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-8 (2008).
3. Forbes, S.A. et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-50 (2011).
4. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-34 (2002).
5. Altshuler, D.M. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
6. Ju, Y.S. et al. Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals. *Nat Genet* **43**, 745-52 (2011).