Supplemental Materials

Fusion of *KIF5B* and *RET* transforming gene in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing

Young Seok Ju, Won-Chul Lee, Jong-Yeon Shin, Seungbok Lee, Thomas Bleazard, Jae-Kyung Won, Young Tae Kim, Jong-Il Kim, Jin-Hyoung Kang and Jeong-Sun Seo

Contents

A. Supplemental Methods

(1) Detection of fusion gene

1) Evidence search

We searched for discordant read pairs and fusion-spanning reads to detect gene fusion events (Figure 2B). For the paired-end sequencing, most reads in a pair are expected to be mapped to the same chromosome, in the opposite orientation, and within the expected insert size. If the reads in a pair are mapped far away from each other or even mapped to different chromosomes (discordant), there might have been structural variations (such as inversion, deletion, or translocation) which result in a gene fusion event. Furthermore, if a read is broken in two and each half is mapped to different genes (fusion-spanning), it can be additional evidence for gene fusion. From the transcriptome sequencing data, we first extracted such evidence.

2) Fusion candidate detection

When more than two discordant read pairs and fusion-spanning reads are observed together in two distinct genes, we defined those two genes as a fusion candidate.

3) Filtration cascade

a. Strand-orientation filter

DNA strands have directionality, with positive and negative strands. Each human gene is located on either the positive or negative strand of the human genome. Combining the strand-orientation of the candidate fusion genes and that of paired-end reads from transcriptome sequencing enables us to identify the “donor (upstream)” and “receptor (downstream)” genes in the candidates. If the donor (or receptor) gene is not defined uniquely, or strand-orientations are contradictory between the genes and short-reads in a fusion gene candidate, we regarded the

candidate as a sequencing error and discarded it.

b. Homology filter

If two genes are highly homologous, the reads from the two genes can be misaligned. Therefore, we removed homologous fusion candidates using BLAST[1]. We used bl2seq (BLAST 2 Sequences) in the BLAST package and discarded the fusion candidates where there was a significant similarity (E-value < 0.01) between two genes.

c. Fusion-spanning read filter

If a fusion-spanning read is aligned by less than 10bp to either gene in the fusion candidate, we considered it as a “spurious read” representing the fusion point and discarded it (Supplemental Figure 5).

d. Fusion point filter

Genuine fusion points are anticipated to show a stacked/ladder-like pattern with more than two fusion-spanning reads[2]. If multiple fusion-spanning reads are generated by PCR duplication, we cannot see this pattern. Therefore, we only accepted a fusion point when it showed a shifting-pattern of more than 3 bp around the fusion point with more than two fusion-spanning reads (Supplemental Figure 6).

4) Final fusion listing

If the fusion candidates passed through all the filtration steps, they were finally accepted as gene fusion events and reported in a list.

5) Software for fusion gene detection

Software used for fusion gene detection in this study (GFP; Gene Fusion Program) is

available in our website (ftp://ftp.gmi.ac.kr/pub/GFP/). Currently, reads aligned on human reference genome using GSNAP[3] .

(2) Estimation of tumor heterogeneity

To estimate tumor heterogeneity of liver metastatic lung cancer of AK55, we used read counts for 8 somatic SNVs identified (Supplemental Table 1). Because those SNPs qualified following conservative criteria for somatic SNVs (described in METHODS), we regarded those SNVs as existing in the major subclone of the metastatic cancer. In addition, we regarded all of them as heterozygotes in the subclone, since they are not driver mutations and are highly unlikely to be homozygote.

In the 8 somatic SNVs, combined read-counts for SNV and wildtype allele were 67 and 206, respectively (Total = 273).

As shown in Supplemental Figure 1, read-allele frequencies of heterozygote SNVs showed a distribution clustered around 0.5. Their upper and lower boundaries were approximately 0.15 and 0.85. Given the fact that the above somatic SNVs are in heterozygotes, we expect that ~ 67 wildtype-reads originated from the cancer clone. Therefore, out of a total 273 reads, 134 reads (67 x 2) were from a major subclone. As a result, we may estimate that 49.1 % of our sample of liver metastatic lung cancer tissues was major subclone. Considering the upper and lower boundaries of the distribution (0.15 and 0.85), the possible percentage ranges from 28.8 % to 100%.

Regarding the *KIF5B-RET* fusion gene, we found 8 reads and 6 reads supporting the inversion and normal chromosome structure respectively from the whole-genome sequence of liver metastatic lung adenocarcinoma of AK55. By applying the same strategies, we calculate that this fusion gene originates from 67.7 % - 100 % of liver metastatic lung cancer tissue. Although the total number of reads (8 + 6 = 14) may not be sufficient for great precision, we conclude that the fusion gene exists at least in the major subclone of the metastatic cancer tissues.

B. Supplemental Tables

Supplemental Table 1. A full list of 10,390 non-synonymous SNVs identified from the whole-genome sequence of liver metastatic lung cancer tissue of AK55.

SuppTable_1_nsSNVs_liverMets.xls

Depicted below is a preview of the full table.

Chr	Pos	ref	snp	ref_count	wt_count	is_homozygote	annotation	wt_AA	var_AA	ref_count in_Blood	total_RD in_Blood	is_in_KoreanGenomes (Ju_NatGen2011)	is_in_1KGenomes (Durbin_Nature2010)	is_somati
chr1	59374	A	G	25	0	Homozygote	CDS:OR4F5	T	A	13	13	Y	N	N
chr1	867694	T	C	2	0	Homozygote	CDS:SAMD11	W	R	1	1	Y	Y	N
chr1	878522	T	C	14	7	Heterozygote	CDS:NOC2L	I	V	4	9	Y	Y	N
chr1	899101	G	C	6	10	Heterozygote	CDS:PLEKHN1	R	P	5	7	Y	Y	N
chr1	899172	T	C	14	0	Homozygote	CDS:PLEKHN1	S	P	9	9	Y	Y	N
chr1	939471	G	A	8	10	Heterozygote	CDS:ISG15	S	N	2	4	Y	Y	N
chr1	1110294	G	A	21	14	Heterozygote	CDS:TTLL10	S	N	5	9	Y	Y	N
chr1	1212130	G	C	7	0	Homozygote	CDS:SCNN1D	R	P	4	4	Y	Y	N
chr1	1259417	T	C	14	0	Homozygote	CDS:TAS1R3	C	R	7	7	Y	N	N
chr1	1351504	C	T	12	1	Homozygote	CDS:TMEM88B	P	L	1	1	Y	Y	N
chr1	1377627	G	A	12	18	Heterozygote	CDS:ATAD3C	G	R	4	6	Y	N	N
chr1	1411854	G	A	5	10	Heterozygote	CDS:ATAD3B	R	Q	1	3	Y	N	N
chr1	1420831	A	T	8	7	Heterozygote	CDS:ATAD3B	K	M	0	2	N	N	N
chr1	1548655	T	C	6	7	Heterozygote	CDS:MIB2;UTR:MI	M	T	2	4	Y	Y	N
chr1	1589675	C	T	14	28	Heterozygote	CDS:LOC728661;I	V	I	3	20	Y	N	N
chr1	1624871	A	G	8	14	Heterozygote	CDS:CDK11A;Intro	V	A	0	5	N	N	N
chr1	1628785	C	T	13	12	Heterozygote	CDS:CDK11A;Intro	D	N	2	4	Y	N	N
chr1	1640647	T	C	47	55	Heterozygote	CDS:CDK11B,CDK	H	R	17	55	Y	Y	N
chr1	1640657	A	G	45	57	Heterozygote	CDS:CDK11B,CDK	C	R	18	52	Y	Y	N
chr1	1640692	A	G	37	63	Heterozygote	CDS:CDK11B,CDK	V	A	18	61	Y	Y	N
chr1	1640705	G	A	27	72	Heterozygote	CDS:CDK11B,CDK	R	W	13	54	Y	Y	N
chr1	1656111	G	A	16	29	Heterozygote	CDS:SLC35E2	R	W	7	19	Y	Y	N

Supplemental Table 2. A full list of 334 CDS indels identified from the whole-genome sequence of liver metastatic lung cancer tissue of AK55.

SuppTable_2_CDSindels_liverMets.xls

Depicted below is a preview of the full table.

Chr	start_pos(left)	size (bp)	type	var_allele	var_count	wt_count	gene (exon involved)	total_RD in_blood	indel_found in blood	is_in_KoreanGenomes (Ju_NatGen2011)	is_in_1KGenomes (Durbin_Nature2010)	is_somatic_
chr1	1889968	3	ins	CCT	23	3	KIAA1751	11	Y	N	N	N
chr1	2928267	3	del	-	21	12	ACTRT2	8	Y	Y	Y	N
chr1	13978986	3	ins	CCT	30	8	PRDM2	20	Y	N	N	N
chr1	17591258	1	del	-	21	1	PADI6	7	Y	Y	Y	N
chr1	27748104	3	del	-	4	13	AHDC1	3	Y	N	Y	N
chr1	52078651	3	del	-	37	18	NRD1	25	Y	Y	Y	N
chr1	54377907	1	ins	C	6	4	CDCP2	8	Y	Y	Y	N
chr1	62683097	3	del	-	18	44	USP1	39	Y	N	N	N
chr1	67628309	3	del	-	11	32	IL12RB2	34	Y	N	N	N
chr1	74810445	3	del	-	13	19	C1orf173	26	Y	Y	Y	N
chr1	77796937	3	ins	ATT	11	25	AK5	33	Y	N	N	N
chr1	86818484	6	del	-	11	21	CLCA4	18	Y	Y	Y	N
chr1	143326608	2	ins	AA	26	9	NBPF9	10	Y	Y	N	N
chr1	143626980	1	del	-	34	62	PDE4DIP	66	Y	Y	N	N
chr1	143635085	1	del	-	35	59	PDE4DIP	55	Y	Y	N	N
chr1	144747332	4	ins	TATC	6	22	NBPF11	17	N	N	N	Y
chr1	150462352	1	del	-	42	22	HRNR	36	Y	Y	N	N
chr1	154831674	2	ins	CA	20	6	GPATCH4	19	Y	N	N	N
chr1	169823508	1	ins	C	33	15	BAT2L2	21	Y	N	N	N
chr1	201404409	2	ins	CT	8	19	MYBPH	6	N	N	N	N
chr1	243351843	5	ins	TTTTA	15	39	EFCAB2	19	Y	Y	N	N
chr1	245681884	1	del	-	19	6	OR2B11	17	Y	Y	Y	N
chr1	246045164	3	del	-	24	4	OR14A16	22	Y	Y	Y	N
chr1	246525498	1	del	-	25	6	OR2T12	28	Y	Y	Y	N
chr1	246579562	1	del	-	20	26	OR14C36	25	Y	N	N	N
chr1	246868567	7	del	-	13	4	OR2T35	16	Y	Y	Y	N

Supplemental Table 3. A full list of 70 CDS large deletion candidates identified from the whole-genome sequence of liver metastatic lung cancer tissue of AK55.

SuppTable_3_CDSCNloss_LivMets.xls

Depicted below is a preview of the full table.

Chr	Start	Stop	size	# Stretched Reads	RD ratio relative to flanking regions	Genes (exon involved)
chr1	45799117	45800888	1772	8	0.427046263	AKR1A1
chr1	108534856	108538779	3924	14	0.385947037	SLC25A24
chr1	143612529	143618160	5632	13	0.573020264	PDE4DIP
chr1	143804306	143808442	4137	35	0.277033008	SEC22B
chr1	147628609	147921627	293019	7	0.290305917	PPIAL4A
chr1	150594146	150594394	249	23	0.04519774	FLG2
chr1	150593980	150594290	311	15	0	FLG2
chr1	153897963	153999856	101894	4	0.880045093	YY1AP1,DAP3,GON4L
chr1	159786759	159829771	43013	7	0.663020913	FCGR2C
chr1	199446558	199446907	350	7	0.415584416	IGFN1
chr1	199446241	199446843	603	3	0.281690141	IGFN1
chr1	205760579	205932940	172362	5	0.714263265	CR1,CR1L
chr2	110210032	111169571	959540	17	0.277648507	MALL,NPHP1,LIMS3,RGPD8,RGPD6,RGPD5,BUB1
chr2	240630178	240630918	741	3	0	PRR21
chr3	75869890	75870425	536	10	0.730067243	ZNF717
chr3	99364886	99404198	39313	23	0.881484439	OR5H15
chr3	131246078	131289429	43352	21	0.0408921	ALG1L2
chr3	194714883	194715272	390	7	0.373626374	ATP13A4
chr3	196937734	196938314	581	11	0	MUC20
chr3	196990557	196991069	513	7	0.463768116	MUC4

Supplemental Table 4. A full list of 10 RNA editing candidates identified from the whole-genome and transcriptome sequence of liver metastatic lung cancer tissue of AK55.

chr	pos	ref	var	Genome (LivMets) var_count	Genome (LivMets) total_RD	RNA wt_count	RNA var_count	annotation	wt_AA	var_AA
chr1	62725668	T	C	0	25	0	28	CDS:DOCK7	T	A
chr3	44587690	A	G	0	29	8	13	CDS:ZNF167:Intron:ZNF	D	G
chr3	185044746	T	C	0	32	52	17	CDS:PARL	M	V
chr5	171270237	T	C	0	23	42	12	CDS:FBXW11	H	R
chr10	59656822	T	C	0	30	11	17	CDS:IPMK	T	A
chr11	1007466	T	C	0	60	22	10	CDS:MUC6	R	G
chr11	113182423	T	C	0	37	0	14	CDS:USP28	K	E
chr15	73433139	A	G	0	20	2	17	CDS:NEIL1	K	R
chr19	8867764	T	C	0	31	2	13	CDS:MUC16	T	A
chr19	11838035	A	G	0	24	10	15	CDS:ZNF439	K	E

Supplemental Table 5. Full list of 52 fusion genes from transcriptome sequencing.

Category	Index	Donor gene	Acceptor gene	Chr	Distance (Mb)	Number of discordant reads	Number of spanning reads	Evidence in whole-genome sequence
Intra-chromosomal								
	1	*KIF5B*	*RET*	chr10	10.580	34	60	YES (inversion)
	2	*KIF5B*	*KIAA1462*	chr10	1.970	4	4	-
	3	*EEF1DP3*	*FRY*	chr13	0.133	3	5	-
	4	*RPS6KB1*	*TMEM49*	chr17	0.097	4	31	-
	5	*HACL1*	*COLQ*	chr3	0.075	3	4	-
	6	*TMEM56*	*RWDD3*	chr1	0.073	4	11	-
	7	*FAM18B2*	*CDRT4*	chr17	0.065	4	29	-
	8	*CTBS*	*GNG5*	chr1	0.065	6	27	-
	9	*METTL10*	*FAM53B*	chr10	0.054	2	4	-
	10	*AZGP1*	*GJC3*	chr7	0.048	5	15	-
	11	*NKX2-1*	*SFTA3*	chr14	0.046	3	7	-
	12	*ADSL*	*SGSM3*	chr22	0.036	5	6	-
	13	*ART4*	*C12orf69*	chr12	0.034	3	4	-
	14	*LOC100131434*	*IDS*	chrX	0.031	2	11	-
	15	*LOC100130093*	*SNAP47*	chr1	0.030	2	2	-
	16	*C15orf57*	*MRPL42P5*	chr15	0.025	2	7	-
	17	*MIA2*	*CTAGE5*	chr14	0.024	30	102	-
	18	*SH3D20*	*ARHGAP27*	chr17	0.024	2	10	-
	19	*RBM14*	*RBM4*	chr11	0.023	16	24	-
	20	*GOLT1A*	*KISS1*	chr1	0.021	3	2	-
	21	*CLCF1*	*POLD4*	chr11	0.021	2	4	-
	22	*SLC43A3*	*PRG2*	chr11	0.020	13	35	-
	23	*PRKAA1*	*TTC33*	chr5	0.018	2	4	-
	24	*BMS1P4*	*AGAP5*	chr10	0.016	2	3	-
	25	*NOS1AP*	*C1orf226*	chr1	0.015	3	16	-
	26	*MFSD7*	*ATP5I*	chr4	0.014	3	9	-
	27	*SCNN1A*	*TNFRSF1A*	chr12	0.014	15	42	-
	28	*OSGIN1*	*NECAB2*	chr16	0.013	3	10	-
	29	*VAMP8*	*VAMP5*	chr2	0.011	4	8	-
	30	*NDUFB8*	*SEC31B*	chr10	0.010	4	4	-
	31	*C6orf47*	*BAT3*	chr6	0.009	2	8	-
	32	*SEC14L4*	*SDC4P*	chr22	0.008	3	2	-

	33	*WDR81*	*SERPINF2*	chr17	0.007	5	30	-
	34	*ARPC4*	*TTLL3*	chr3	0.007	11	12	-
	35	*PMF1*	*BGLAP*	chr1	0.007	3	6	-
	36	*ANKRD39*	*ANKRD23*	chr2	0.006	7	14	-
	37	*CTSD*	*LOC402778*	chr11	0.006	6	31	-
	38	*ARL6IP1*	*RPS15A*	chr16	0.006	4	7	-
	39	*ADCK4*	*NUMBL*	chr19	0.005	2	4	-
	40	*ZNF606*	*C19orf18*	chr19	0.005	2	6	-
	41	*ORMDL3*	*GSDMB*	chr17	0.004	3	2	-
	42	*FCN3*	*MAP3K6*	chr1	0.004	4	2	-
	43	*PLEKHM1P*	*LOC146880*	chr17	0.004	4	11	-
	44	*APOC4*	*APOC2*	chr19	0.003	4	6	-
	45	*HARS2*	*ZMAT2*	chr5	0.003	5	6	-
	46	*COL7A1*	*UCN2*	chr3	0.002	3	14	-
	47	*NLRP1*	*LOC728392*	chr17	0.002	2	2	-
	48	*C17orf106*	*CDK3*	chr17	0.001	3	9	-
	49	*TAP1*	*PSMB8*	chr6	0.001	2	14	-
	50	*CCDC142*	*MRPL53*	chr2	0.000	5	7	-
	51	*APBB3*	*SRA1*	chr5	0.000	6	5	-
Inter-chromosomal								
	52	*RSPO1*	*HP*	chr16;chr1	-	2	3	-

Supplemental Table 6. A full list of expression levels of human genes identified from the transcriptome sequencing of cancer tissues of AK55, LC_S1 - S5.

SuppTable_6_expression_levels.xls

Depicted below is a preview of the full table.

gene	accession_ID	chr	start	stop	length	strand	AK55	LC_S1	LC_S2	LC_S3	LC_S4	LC_S5
WASH7P	NR_024540	1	14361	29370	1769	-	7.24	3.12	3.43	4.36	0.98	1.97
FAM138A	NR_026818	1	34610	36081	1130	-	0.00	0.00	0.00	0.00	0.00	0.00
FAM138F	NR_026820	1	34610	36081	1130	-	0.00	0.00	0.00	0.00	0.00	0.00
OR4F5	NM_001005484	1	69090	70008	918	+	0.00	0.00	0.00	0.00	0.00	0.00
LOC100132062	NR_028325	1	323891	328581	4370	+	6.76	0.00	0.00	0.00	0.00	0.01
LOC100132287	NR_028322	1	323891	328581	4370	+	6.76	0.00	0.00	0.00	0.00	0.01
LOC100133331	NR_028327	1	323891	328581	4273	+	7.58	0.00	0.00	0.00	0.00	0.00
OR4F16	NM_001005277	1	367658	368597	939	+	0.01	0.00	0.00	0.00	0.00	0.00
OR4F29	NM_001005221	1	367658	368597	939	+	0.01	0.00	0.00	0.00	0.00	0.00
OR4F3	NM_001005224	1	367658	368597	939	+	0.03	0.00	0.00	0.00	0.00	0.00
NCRNA00115	NR_024321	1	761585	762902	1317	-	1.28	2.05	1.98	2.72	1.79	4.12
LOC643837	NR_015368	1	763063	789740	1543	+	5.80	3.42	3.59	3.14	3.48	3.79
FAM41C	NR_027055	1	803450	812182	1706	-	2.15	0.43	0.57	7.45	0.37	1.16
FLJ39609	NR_026874	1	852952	854817	496	-	0.08	0.12	0.03	0.05	0.07	0.00
SAMD11	NM_152486	1	861120	879961	2554	+	2.27	6.68	5.68	2.07	3.74	3.78
NOC2L	NM_015658	1	879582	894679	2800	-	14.54	17.47	20.62	24.50	42.73	20.15
KLHL17	NM_198317	1	895966	901099	2564	+	1.44	2.33	3.11	3.66	3.29	2.85
PLEKHN1	NM_001160184	1	901876	910484	2295	+	1.38	1.08	2.06	1.55	2.56	1.36
PLEKHN1	NM_032129	1	901876	910484	2400	+	1.36	1.06	2.18	1.51	2.52	1.36
C1orf170	NR_027693	1	910578	917473	3040	-	0.36	0.36	0.51	0.24	0.68	0.41
HES4	NM_021170	1	934341	935552	962	-	1.48	1.76	2.85	2.35	4.13	0.86

[Exon-by-exon RET expression among 6 samples]

gene	accession	chrom	exon	start	end	length	strand	AK55	LC_S1	LC_S2	LC_S3	LC_S4	LC_S5
RET	NM_020630	10	exon1	43572516	43572779	263	+	0.03	0.00	0.10	0.00	0.00	0.00
RET	NM_020630	10	exon2	43595906	43596170	264	+	0.00	0.29	0.38	0.11	0.00	0.18
RET	NM_020630	10	exon3	43597789	43598077	288	+	0.18	0.16	0.68	0.18	0.21	0.24
RET	NM_020630	10	exon4	43600399	43600641	242	+	0.06	0.25	0.41	0.01	0.24	0.00
RET	NM_020630	10	exon5	43601823	43602019	196	+	0.07	0.25	0.32	0.17	0.15	0.00
RET	NM_020630	10	exon6	43604478	43604678	200	+	0.24	0.88	0.33	0.01	0.10	0.53
RET	NM_020630	10	exon7	43606654	43606913	259	+	0.14	0.32	0.43	0.12	0.15	0.16
RET	NM_020630	10	exon8	43607546	43607672	126	+	0.11	0.31	0.00	0.04	0.09	0.10
RET	NM_020630	10	exon9	43608300	43608411	111	+	0.26	0.10	0.27	0.14	0.15	0.40
RET	NM_020630	10	exon10	43609003	43609123	120	+	0.40	0.91	0.58	0.22	0.04	0.00
RET	NM_020630	10	exon11	43609927	43610184	257	+	0.24	1.48	0.66	0.24	0.14	0.19
RET	NM_020630	10	exon12	43612031	43612179	148	+	4.25	0.29	12.50	0.00	0.14	0.09
RET	NM_020630	10	exon13	43613820	43613928	108	+	5.82	0.02	7.74	0.00	0.20	0.00
RET	NM_020630	10	exon14	43614978	43615193	215	+	4.49	1.10	8.41	0.16	0.10	0.20
RET	NM_020630	10	exon15	43615528	43615651	123	+	7.13	1.85	14.60	0.25	0.00	0.62
RET	NM_020630	10	exon16	43617393	43617464	71	+	7.45	1.41	17.86	0.04	0.00	0.00
RET	NM_020630	10	exon17	43619118	43619256	138	+	8.94	0.42	18.15	0.15	0.04	0.21
RET	NM_020630	10	exon18	43620330	43620430	100	+	8.88	0.00	15.81	0.21	0.31	0.44
RET	NM_020630	10	exon19	43622022	43622952	930	+	8.21	0.89	8.37	0.05	0.22	0.12

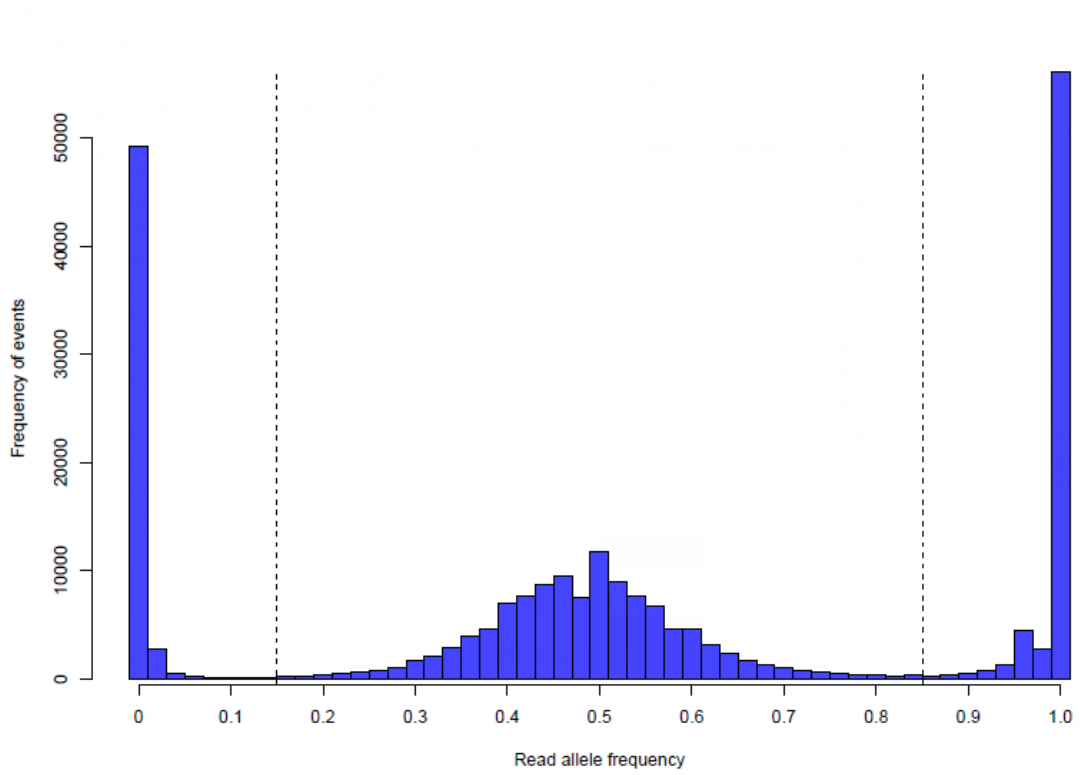
Supplemental Table 7. Summary statistics of transcriptome sequencing from primary lung adenocarcinoma tissues of LC_S1 - S5.

Sample	# of total reads	# of aligned reads	Read length	Total throughput (aligned)	aligned %
LC_S1	71,707,630	52,378,115	2 x 101 bp	5,290,189,615	73.04
LC_S2	89,309,594	66,196,831	2 x 101 bp	6,685,879,931	74.12
LC_S3	104,898,166	78,419,618	2 x 101 bp	7,920,381,418	74.76
LC_S4	96,843,482	70,747,235	2 x 101 bp	7,145,470,735	73.05
LC_S5	92,899,066	69,604,137	2 x 101 bp	7,030,017,837	74.92

C. Supplemental Figures

Supplemental Figure 1. Distribution of read-allele frequency (count of SNV reads / (count of SNV + wildtype reads)) in the whole-genome sequencing of liver metastatic lung cancer tissue of AK55.

We selected 235,706 autosomal SNVs, which were known in dbSNP 130 and which we had observed at minor allele frequency = 50 % from whole-genome sequencing of 10 Koreans[4]. Assuming Hardy-Weinberg equilibrium in AK55, we would then expect to find 25% of the loci to be homozygotic SNV, 25% to be homozygotic wildtype, and 50% heterozygotic.

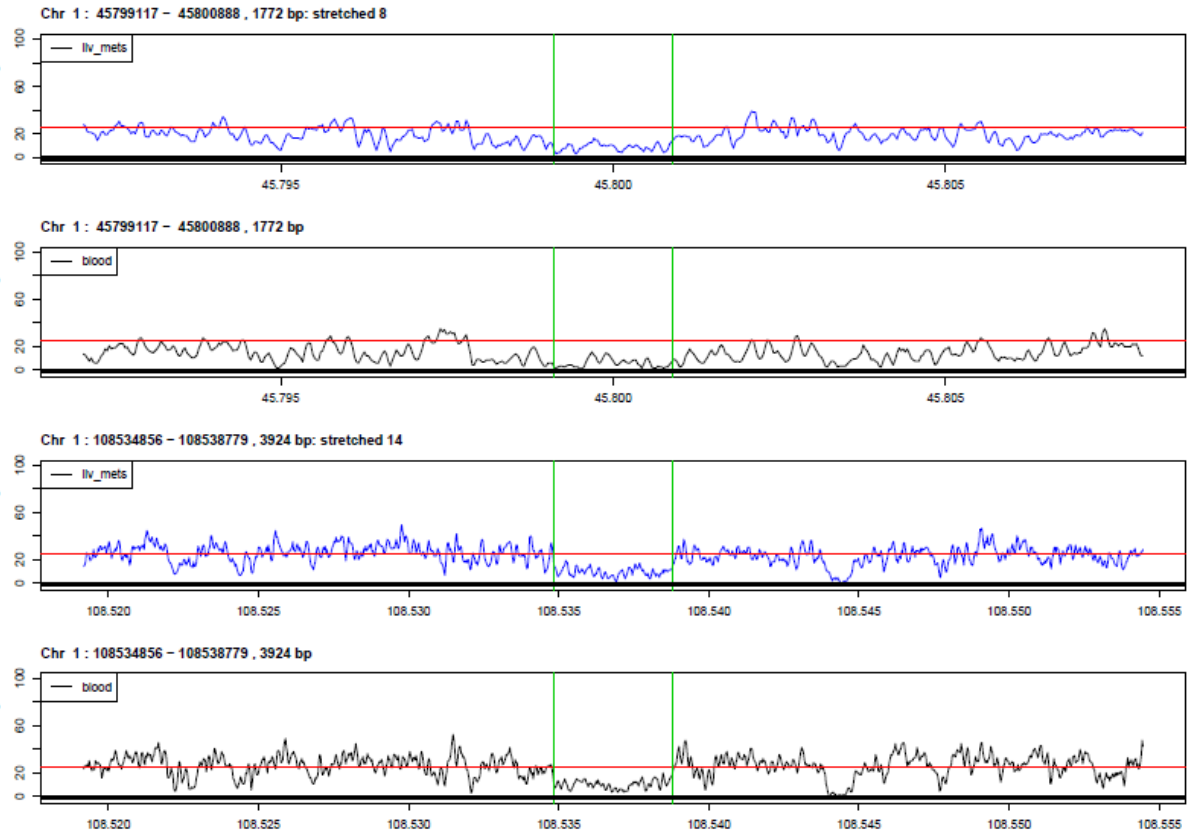


Supplemental Figure 2. Comparison of read-depth of whole-genome sequence for 70 CDS large deletion candidates from liver metastatic lung cancer tissue and blood.

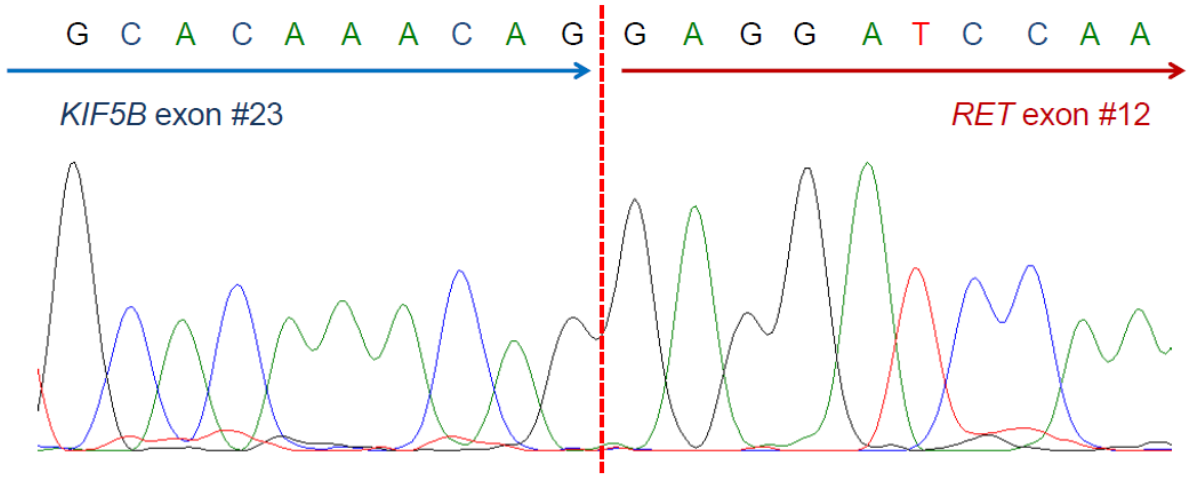
SuppFigure_2_CDSlargedeletion_candidates.pdf

Depicted below is a preview of the full figures.

5' and 3' flanking regions were shown with large deletion candidate regions. The boundaries of large deletions are shown by green vertical lines. Blue line; read-depth of liver metastatic lung cancer tissue of AK55 (normalized to 25 x). Black line; read-depth of blood of AK55 (normalized to 25 x). No somatic large deletion was identified in this analysis.

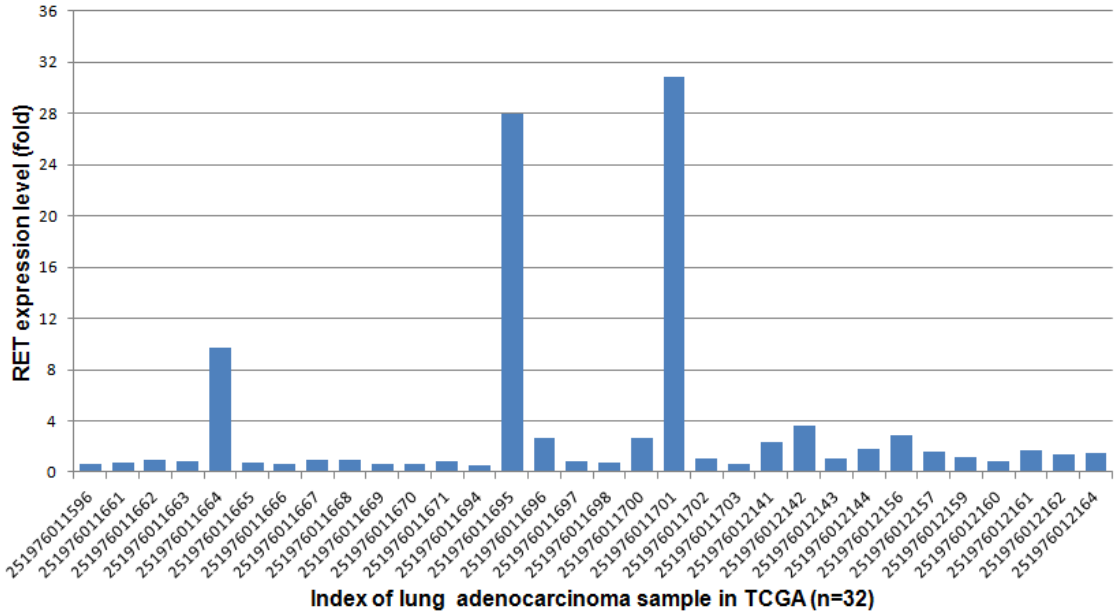


Supplemental Figure 3. Identification of breakpoint of the *KIF5B-RET* fusion gene in LC_S6 using Sanger sequencing.

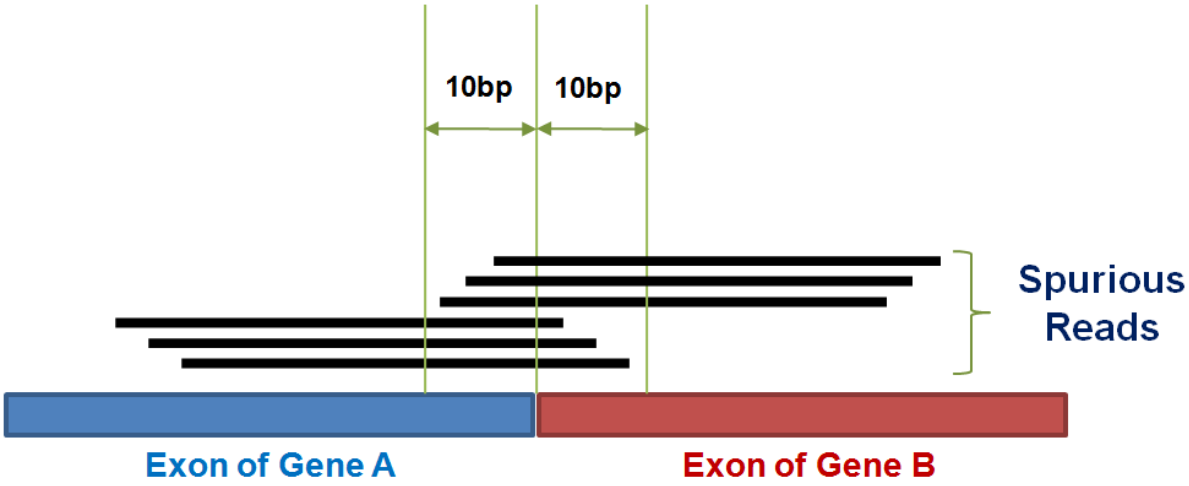


Supplemental Figure 4. RET expression levels in lung adenocarcinomas deposited in TCGA.

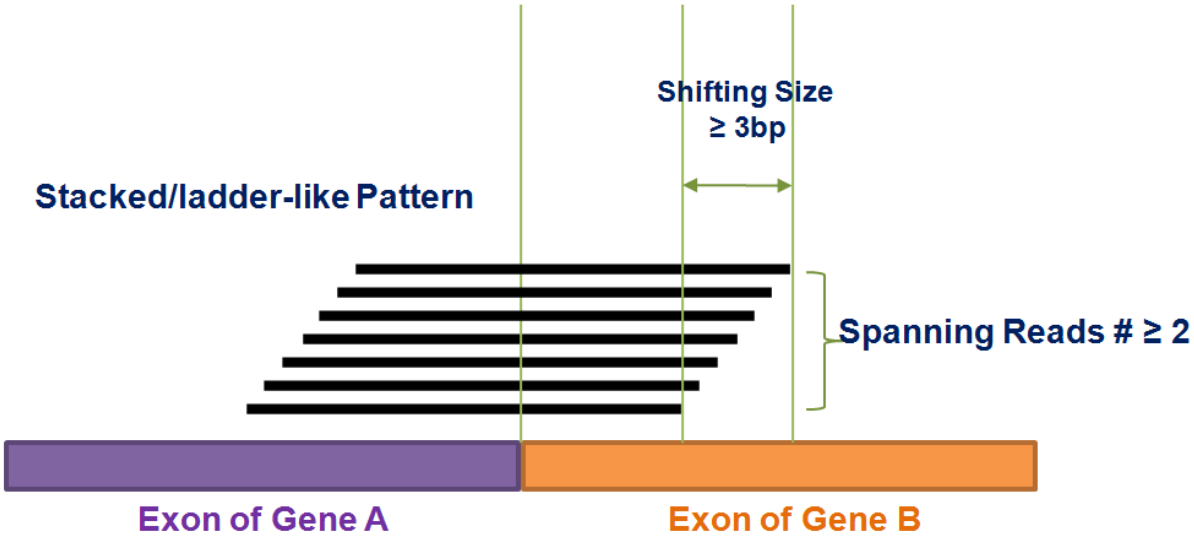
We analyzed the expression microarray data of 32 lung adenocarcinomas in the TCGA project. Of these, 3 samples showed clear over-expression of *RET*. These data cannot confirm that the *KIF5B-RET* fusion gene exists in the cancer genome of those 3 samples. However, the data clearly show that there is a subset of lung cancers, where *RET* proto-oncogene is highly expressed.



Supplemental Figure 5. Schematic diagram representing the “Spurious Reads”.



Supplemental Figure 6. Schematic diagram representing the “Stacked/ladder-like pattern”.



D. References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.

2. Edgren H, Murumagi A, Kangaspeska S, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. Genome Biol 2011;12:R6.

3. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 2010;26:873-81.

4. Ju YS, Kim JI, Kim S, et al. Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals. Nat Genet 2011.