<붙임 4>

기관고유연구사업 최종보고서

연구분야(코드)		과제번호	1110100	지원 프로그램	*(일반연구)과제*
과제성격(기초,응용,개발)	응용	실용화 대상여부	실용화	공개가능여부 (공개,비공개)	*공개*

연구과제명	(국문) 차세대 염기서열분석법을 이용한 폐암의 유전체 연구
	(영문) Comprehensive genomic characterization of lung cancer using next generation sequencing (NGS)

과제책임자	소 속	폐암연구과	직 위	부장
	성 명	한지연	전 공	내과종양학

세부과제	구분	세부과제명	세부과제책임자		
			성명	소속(직위)	전 공
	1				
	2				
	3				

총 연구기간	년 월~ 년 월(총 년)	참여연구원수 (단위: 명, MY)	

연구기간 및 연구비 (단위:천원)	구분	연 구 기 간	계	국 립 암센터	기 업 부 담 금		
					소계	현금	현물
	계	2011.1~2013.12	360,000	360,000			
	제1차	2011.1~2011.12	120,000	120,000			
	제2차	2012.1~2012.12	120,000	120,000			
	제3차	2013.1~2013.12	120,000	120,000			

참 여 기 업	명 칭		전화		FAX	

기관고유연구사업관리규칙에 따라 본 연구개발사업을 성실히 수행하였으며 아래와 같이 최종보고서를 제출합니다.

2013년 10 월 30 일

과제책임자 한 지 연 (서명)

국 립 암 센 터 원 장 귀 하

(첨부서류)

작 성 요 령

○ 반드시 편집순서에 따라 작성하여야 함

○ 전년도 연차실적을 포함하여 전체 사업기간에 대한 연구결과와 성과를 중심으로 기술함

○ 필요한 경우 소제목을 설정하여 체계적인 형식을 갖추도록 함

○ 요약문은 연구목표, 연구내용 및 방법, 연구성과 등을 중심으로 작성함

○ 요약문중 중심단어(key words)는 5개 이내로 반드시 기재해야 함

○ 번호나 기호를 사용한 보고서 형태로 작성하고 표나 그림을 이용할 수 있음. 단, 동 보고서와 함께 제출하는 전산파일에도 같은 표와 그림이 첨부되어 있어야 함

목 차

※ 여러개의 세부과제로 과제가 구성된 경우 위 목차와 동일하게 세부과제별로 작성함

(I. 총괄과제, II. 제1세부과제, III. 제2세부과제.....)

< 요 약 문 >

<table>
<tr><td>연구목표
(200자 이내)</td><td colspan="4"><최종목표>
◆ 최신의 암 유전체 분석 기법인 NGS를 이용하여 폐암의 암 유전체 연구를 수행, 새로운 유전자 변형 등을 발굴하여 폐암의 진단 및 치료에 기여를 할 수 있는 원천 지식을 제공하고 암 유전체 연구를 위한 기반 확보

<당해연도목표>
◆ 비흡연 폐선암 환자에서 EGFR-TKI 내성에 관여하는 새로운 유전체 변형을 발굴하여 EGFR-TKI 일차적 내성기전 제시
◆ 비흡연 폐선암의 새로운 driver oncogene 발굴
◆ NGS 분석을 통한 소세소폐암의 발생기전 분석</td></tr>
<tr><td>연구내용 및 방법
(500자 이내)</td><td colspan="4">◆ EGFR-TKI 일차내성 관련 기전 분석 및 폐선암 새로운 driver oncogene 발굴
1) 수술 후 재발하여 EGFR-TKI 치료를 받은 비흡연 폐선암 환자 37명의 암 조직으로부터 DNA 분리함.
2) IonTorrnet PGM을 이용하여 targeted sequencing 실시함.
3) IonTorrent 분석법으로 내성 기전 발굴에 실패한 7명 환자에서 Whole exome sequencing과 RNA sequencing 실시함.
◆ Whole genome sequencing을 통한 소세포폐암 발생 기전 분석
1) 수술적 적제술을 시행한 stage I SCLC (pT2N0) 1례 종양 및 정상 조직에서 DNA 분리
2) whole genome sequencing using 90 bp paired-end reads & validation variations by Sanger sequencing & SNP genotyping</td></tr>
<tr><td>연구개발에 따른 기대성과</td><td colspan="4"><정량적 성과[1)]>
<table><tr><th>구분</th><th>달성치/목표치[1)]</th><th>달성도(%)</th></tr><tr><td>SCI 논문 편수</td><td>18/6</td><td>300</td></tr><tr><td>IF 합</td><td>75.516/32</td><td>236</td></tr><tr><td>기타 성과</td><td></td><td></td></tr></table>1) 총연구기간 내 목표연구성과로 기 제출한 값

<정성적 성과>
-주요연구성과를 개조식으로 간단히 작성(5줄 이내)</td></tr>
<tr><td rowspan="4">색인어</td><td rowspan="2">국문</td><td>폐선암</td><td>소세포폐암</td><td>차세대연기서열분석</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td rowspan="2">영문</td><td>lung adenocarcinoma</td><td>small cell lung cancer</td><td>NGS</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

Project Summary

Title of Project	Comprehensive genomic characterization of lung cancer using next generation sequencing (NGS)
Key Words	lung adenocaarcinoma, SCLC, NGS
Project Leader	Ji-Youn Han
Associated Company	

Clinical application of targeted next generation sequencing for predicting responsiveness to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) in never-smoking lung adenocarcinoma

Background: To investigate the clinical utility of targeted next-generation sequencing (NGS) for predicting the responsiveness to epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitor (TKI) therapy, we compared the efficacy with conventional sequencing in never-smokers with lung adenocarcinoma (NSLAs).

Methods: We obtained DNA from 48 NSLAs who received gefitinib or erlotinib for their recurrent disease after surgery. Sanger sequencing and peptide nucleic acid clamp polymerase chain reaction (PCR) were used to analyze *EGFR, KRAS, BRAF,* and *PIK3CA* mutations. We analyzed *ALK, RET,* and *ROS1* rearrangements by fluorescent in situ hybridization or reverse transcriptase-PCR and quantitative real-time PCR. After molecular screening, Ion Torrent NGS was performed in 31 cases harboring only *EGFR* exon 19 deletions (19DEL), an L858R mutation, or none of the above mutations.

Results: The 31 samples were divided into four groups: 1) responders to EGFR-TKIs with only 19DEL or L858R (n=15); 2) primary resistance to EGFR-TKI with only 19DEL or L858R (n=4); 3) primary resistance to EGFR-TKI without any mutations (n=8); 4) responders to EGFR-TKI without any mutations (n=4). With NGS, all conventionally detected mutations were confirmed except for one L858R in group 2. Additional uncovered predictive mutations with NGS included one *PIK3CA* E542K in group 2, two *KRAS* (G12V and G12D), one *PIK3CA* E542K, one concomitant *PIK3CA* and *EGFR* L858R in group 3, and one *EGFR* 19DEL in group 4.

Conclusions: Targeted NGS provided a more accurate and clinically useful molecular classification of NSLAs. It may improve the efficacy of EGFR-TKI therapy in lung cancer.

Identification of novel oncogenes in a stage I small cell lung cancer through whole-genome sequencing

Background: Small cell lung cancer (SCLC) is rapidly progressive and most cases are presented with advanced stage. It hampers access to early stage cancer samples, which are essential to understand the early development of this disease. We performed whole-genome sequencing (WGS) of a stage I SCLC to analyze the genomic feature of early tumorigenesis of SCLC.

Methods: We obtained normal and tumor genomic DNAs from a patient with stage I SCLC underwent curative resection and analyzed the normal and cancer genomes by WGS using 90 bp paired-end reads and validated the cancer variations by Sanger sequencing and single-nucleotide polymorphism (SNP) genotyping.

Results: WGS revealed a lot of SNVs, small insertion/deletions, and chromosomal abnormality. Chromosomes 4p, 5q, 13q, 15q, 17p, and 22q contained many block deletions. Especially, copy loss was observed in tumor suppressor genes RB1 and TP53, and copy gain in oncogene hTERT. Recurrent somatic mutations were found in two genes, TP53 and CREBBP. Novel non-synonymous (ns) single nucleotide variations (SNV) were found in C6ORF103 and SLC5A4 genes. Sequencing of SLC5A4 gene in 23 independent limited disease SCLC samples revealed another ns SNV in SLC5A4 gene, indicating ns SNVs in SLC5A4 gene are recurrent in early – stage SCLC. This cancer genome had mutated genes associated with Notch and WNT signaling pathways.

Conclusion: WGS analysis in this study provides the genomic landscape of early-stage SCLC, which may provide new insight into the early development of SCLC.

※ 연구목표, 연구방법, 연구성과를 영문으로 요약하여 2쪽이내의 분량으로 작성

1. 연구의 최종목표

최신의 암 유전체 분석 기법인 NGS를 이용하여 폐암의 암 유전체 연구를 수행, 새로운 유전자 변형 등을 발굴하여 폐암의 진단 및 치료에 기여를 할 수 있는 원천 지식을 제공하고 암 유전체 연구를 위한 기반 확보

1) 비흡연폐선암 환자에서 EGFR-TKI 내성 원인 기전 발굴

2) 소세포폐암의 발생기전 분석

2. 연구의 내용 및 결과

○ 연구의 이론적, 실험적 연구 방법, 연구 내용 및 결과를 객관적으로 기술

◆ Clinical utility of targeted next-generation sequencing for predicting the responsiveness to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in never-smokers with lung adenocarcinoma

Introduction

The treatment strategy for advanced lung adenocarcinoma has been changed by the steady progress in identifying targetable oncogenic mutations. Although first generation epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitors (TKIs), such as gefitinib and erlotinib, had been approved for unselected non-small cell lung cancer (NSCLC) patients, the efficacy was modest [1]. In contrast, the targeted clinical trials of EGFR-TKIs in the molecularly selected patients whose tumors bear sensitive mutations in the tyrosine kinase domain of EGFR demonstrated unprecedented improvements in the response rates and progression-free survival (PFS) [2-5]. Subsequently, evidence-based treatment guidelines recommend basing treatment decisions for lung adenocarcinoma on *EGFR* testing results [6-8].

EGFR mutations are the most common actionable target oncogene in never-smokers with lung adenocarcinoma (NSLAs) and are found in 40-50% of NSLAs [9]. Ninety percent of sensitive *EGFR* mutations are in-frame deletions in exon 19 (19DEL) or a missense mutation in exon 21 (L858R) [10]. Retrospective and prospective studies have established that 70% of NSCLCs harboring 19DEL or L858R mutations respond to EGFR-TKI therapy. Unfortunately, 10-20% of NSCLCs with these mutations show primary resistance to EGFR-TKI therapy [2-5, 7]. Other mutations found in lung adenocarcinoma can also affect the sensitivity to EGFR-TKI therapy, which include *KRAS, BRAF, ALK, RET, ROS1*, and *PIK3CA*. Like *EGFR* mutations, *ALK, RET,* and *ROS1* rearrangements are more common in NSLAs [11, 12]. In contrast, *KRAS* mutations are common in smoking-related lung adenocarcinoma. In the case of *BRAF*, the V600E mutation is more common in never-smokers, but the non-V600E mutations are more common in smoking-related lung adenocarcinoma [13]. These mutations are usually mutually exclusive and show primary resistance to EGFR-TKI therapy [14]. However, *PIK3CA* mutations often coexist with another oncogenic mutation, such as *EGFR* or *KRAS* mutations. The coexistence of *PIK3CA* mutations in *EGFR*-mutant lung cancer is associated with primary resistance to EGFR-TKI therapy [15-17]. Thus, multiple genomic alteration tests may facilitate the best treatment decision.

However, several challenges persist in the implementation of multiple molecular tests to match patients with the best therapy. Although Sanger sequencing and PCR are routinely used to identify clinically relevant mutations and to select the best treatment for patients, these techniques are insensitive to alterations occurring at an allele frequency lower than 20% [18]. In addition, Sanger sequencing and PCR require a few micrograms of DNA for a single

test [19]. However, most NSCLC biopsy samples are not amenable to multiple molecular tests. Finally, multiple, separate tests result in higher costs. Thus, a more sensitive, comprehensive, and cost-effective multiplex testing platform is necessary to best utilize targeted therapy. Recent progress in next-generation sequencing (NGS) technologies and an exponential decrease in the cost of sequencing may overcome these limitations and provide more comprehensive genomic information to improve treatment decisions [20]. To investigate the clinical relevance of NGS for determining the mutational status and more effectively implementing EGFR-TKI therapy, we performed targeted sequencing using the Ion AmpliSeq™CancerPanelinsamplesfromNSLAs.

Methods

Patients and tissues

The eligible patients were NSLAs who had received gefitinib or erlotinib for their recurrent advanced NSCLC after surgical resection. All the tumor samples were collected at the time of surgical resection. Patients receiving neoadjuvant chemotherapy were excluded. The medical records of all the patients were reviewed to extract data on their clinicopathological characteristics, including age, sex, stage, histology, and smoking history. One medical oncologist (SHK), who was unaware of the mutational statuses of the tumors, reviewed all the radiologic images and used the standard Response Evaluation Criteria in Solid Tumors (RECIST, version 1.0) to classify the responses as complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). Progression-free survival (PFS) was defined as the time from the beginning of EGFR-TKI administration to confirmed disease progression or death. This research was approved by the Institutional Review Board of the National Cancer Center Hospital. All the participants provided written informed consent.

DNA extraction

Genomic DNA was extracted from 10 μm-thick sections of 10% neutral formalin-fixed, paraffin-embedded (FFPE) tumor tissue blocks using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). The concentration and purity of the extracted DNA were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA) and Qubit fluorometric quantitation (Life Technologies, Grand Island, NY, USA). The mean concentration of extracted DNA was 382.04 ng/μl (range: 10.2–2936.8 ng/μl), and the measured 260/280 purity was from 1.7 to 2.1. The extracted DNA was stored at −20°C until use.

Conventional molecular screening tests

Mutation screening tests were performed to identify well-known driver mutations in all the samples. We analyzed *EGFR* mutations at exon 18–21 using the polymerase chain reaction (PCR)-based direct DNA sequencing method [21]. Peptide nucleic acid (PNA)-mediated clamping PCR using the PNAClamp™ *EGFR* mutation detection kit (Panagene, Inc., Daejeon, Korea) was used to confirm the *EGFR* mutation status [22]. The assays for the detection of seven different *KRAS*, one *BRAF,* and 11 *PIK3CA* variants were obtained through the PNAClamp™ KRAS [23], PNAClamp™ BRAF [24], and PNAClamp™ PIK3CA Mutation Detection kits (Panagene, Inc., Daejeon, Korea), respectively (Table 1). *ALK* rearrangements were detected by fluorescent in situ hybridization (FISH) in formalin-fixed paraffin-embedded specimens using the break-apart probe for the *ALK* gene (Vysis LSI ALK Dual Color; Abbott Molecular, Abbott Park, IL) [25]. *ROS1* and *RET* fusions were detected by a combined strategy of reverse transcriptase PCR (RT-PCR) and quantitative real-time PCR [26].

Ion Torrent deep-amplicon sequencing

Ten nanograms of DNA was used for multiplex PCR of a panel covering 739 mutations in the following 46 key cancer genes: *ABL1, AKT1, ALK, APC, ATM, BRAF, CDH1, CDKN2A, CSF1R, CTNNB1, EGFR, ERBB2, ERBB4, FBXW7, FGFR1, FGFR2, FGFR3, FLT3, GNAS, HNF1A, HRAS, IDH1, JAK2, JAK3, KDR, KIT, KRAS, MET, MLH1, MPL, NOTCH1, NPM1, NRAS, PDGFRA, PIK3CA, PTEN, PTPN11, RB1, RET, SMAD4, SMARCB1, SMO, SRC, STK11, TP53,* and *VHL* (Ion AmpliSeq Cancer Panel, Life Technologies, Grand Island, NY, USA) [27]. Fragment libraries were constructed using DNA fragmentation, barcode and adaptor ligation, and library amplification, according to the manufacturer's instructions as stipulated in the Ion DNA Barcoding kit (Life Technologies,

Grand Island, NY, USA). The size distribution of the DNA fragments was analyzed on the Agilent Bioanalyzer using the High Sensitivity Kit (Agilent, Santa Clara, CA, USA). Template preparation, emulsion PCR, and Ion Sphere Particle (ISP) enrichment was performed using the Ion Xpress Template kit (Life Technologies, Grand Island, NY, USA) according to the manufacturer's instructions. The ISPs were loaded onto a 316 chip and sequenced using an Ion PGM 100 sequencing kit (Life Technologies, Grand Island, NY, USA).

Bioinformatic analysis

Ion Torrent platform-specific pipeline software (Torrent Suite v2.0) was used to separate the barcoded reads, generate a sequence alignment with the hg19 human genome reference, perform a target-region coverage analysis, and filter and remove poor signal reads. Variant calling was performed with the Variant Caller v2.0 software (Life Technologies, Grand Island, NY, USA). The variant detection required a minimum coverage of 500 reads, and at least 5% of the mutant reads were selected for the variant identification.

Mutation confirmation

The mutations detected by Ion Torrent NGS were confirmed by the PCR-based direct DNA sequencing and PNA-mediated clamping PCR.

Results

Patients

From September 2001 to June 2009, 53 NSLAs received gefitinib or erlotinib for their recurrent advanced disease after surgery. Among them, 48 patients had an adequate quality of DNA for comprehensive mutational analysis. These samples included 40 women and eight men. The median age was 65 years (range 40–81). Nine patients received an EGFR-TKI as a first-line therapy. Thirty-one and eight patients received an EGFR-TKI as a second- or third-line therapy, respectively (Table 2).

Conventional molecular screening

Conventional screening detected 25 *EGFR* mutations (ten exon-19 deletions [19DEL], nine L858R mutations, one G719X mutations, four double *EGFR* mutations, and one concomitant L858R and *PIK3CA* mutations); four *KRAS* codon-12 mutations; and three *EML4-ALK*, two *KIF5B-RET,* and two *CD74-ROS1* rearrangements. No mutations were found in 12 of the patients (Table 1). After conventional screening, we excluded 12 patients with mutations related to resistance or poor response to EGFR-TKIs. We included 31 patients harboring only common sensitive *EGFR* mutations (19DEL or L858R) or none of the above mutations for further analysis. The patients were divided into four groups: 1) responders (PR or SD) to EGFR-TKIs with only *EGFR* 19DEL or L858R mutations (n=15); 2) primary resistance (PD) to EGFR-TKIs with only *EGFR* 19DEL or L858R mutations (n=4); 3) primary resistance to EGFR-TKIs without any mutations (n=8); and 4) responders (1 PR+3 SD) to EGFR-TKIs without any mutations (n=4). The Kaplan-Meier plots of PFS for each group are displayed in Figure 1.

Ion Torrent NGS

Ion Torrent NGS was performed on the 31 samples that had only *EGFR* 19DEL or L858R mutations or no mutations detected by conventional screening. Deep sequencing of the selected regions of 46 cancer-associated genes on the Ion Torrent Personal Genome Machine achieved an average of 1681.6 mapped sequence reads with a 79-bp mean read length. Each base pair in all the amplicons was covered at least once, and at least 95% of all the base pairs were covered at least 100 times. Furthermore, the selected regions in each tumor were covered an average of at least 600 times. The molecular data of the Ion Torrent NGS are summarized in Supplementary Table 1. In group 1, all the known *EGFR* 19DEL or L858R mutations were identified with NGS. Some additional uncommon *EGFR* mutations in exon 13, 20, or 21 were also identified concurrently with the known *EGFR* mutations. In addition, certain mutations that are prone to cancer development were detected in a subset of patients (three *MLH1* V384D mutations and three *STK11* F354L mutations). Four patients in group 2 were characterized as having primary resistance to EGFR-TKIs even with an *EGFR* L858R mutation. With NGS, the known *EGFR* L858R was correctly identified in all but one case

(T2-4). One additional *PIK3CA* E542K mutation was identified concurrently with L858R (case T2-1). Eight patients in group 3 also exhibited primary resistance to EGFR-TKIs even without any resistance-related mutations detected by conventional screening. With NGS, two additional *KRAS* codon-12 point mutations (case T3-3 and T3-7) and one additional *PIK3CA* E542K mutation were detected (case T3-5). Additional concurrent *EGFR* L858R and *PIK3CA* E542K mutations were found in one patient (case T3-8). Four patients in group 4 exhibited PR (n=1) or SD (n=3) without any mutations by conventional screening. In group 4, one additional *EGFR* exon 19DEL was found in a patient who exhibited PR to gefitinib (case T4-4).

Validation

All newly detected *EGFR, KRAS,* and *PIK3CA* mutations by NGS were confirmed by PCR-based direct sequencing and PNA-mediated clamping PCR except one *PIK3CA* E524 mutation (case T3-8).

Table 1. The PNA-mediated clamping assay detects mutations of *EGFR. KRAS, BRAF* and *PIK3CA* genes.

Gene	No	Mutation	Exon	Base change
EGFR	1	G719A	18	2156G>C
	2	G719S	18	2155G>A
	3	G719C	18	2155G>T
	4	E746_A750del	19	2235_2249 del 15
	5	E746_T751>I	19	2235_2252 > AAT
	6	E746_T751del	19	2236_2253 del 18
	7	E746_T751>A	19	2237_2251 del 15
	8	E746_S752>A	19	2237_2254 del 18
	9	E746_S752>V	19	2237_2255>T
	10	E746_A750del	19	2236_2250 del 15
	11	E746_S752>D	19	2238_2255 del 18
	12	L747_A750>P	19	2238_2248 >GC
	13	L747_T751>Q	19	2238_2252 >GCA
	14	L747_E749del	19	2239_2247 del 9
	15	L747_T751del	19	2239_2253 del 15
	16	L747_S752del	19	2239_2256 del 18
	17	L747_ A750>P	19	2239_2248 TTAAGAGAAG>C
	18	L747_P753>Q	19	2239_2258 >CA
	19	L747_T751>S	19	2240_2251 del 12
	20	L747_P753>S	19	2240_2257 del 18
	21	L747_T751del	19	2240_2254 del 15
	22	L747_T751>P	19	2239_2251>C
	23	T790M	20	2369C>T
	24	S768I	20	2303G>T
	25	V769_D770insASV	20	2307_2308 ins9
	26	H773_V774insH	20	2319_2320 insCAC
	27	D770_N771insG	20	2310_2311 insGGT
	28	L858R	21	2573T>G
	29	L861Q	21	2582T>A
KRAS	1	G12D	2	35G>A
	2	G12A	2	35G>C
	3	G12V	2	35G>T

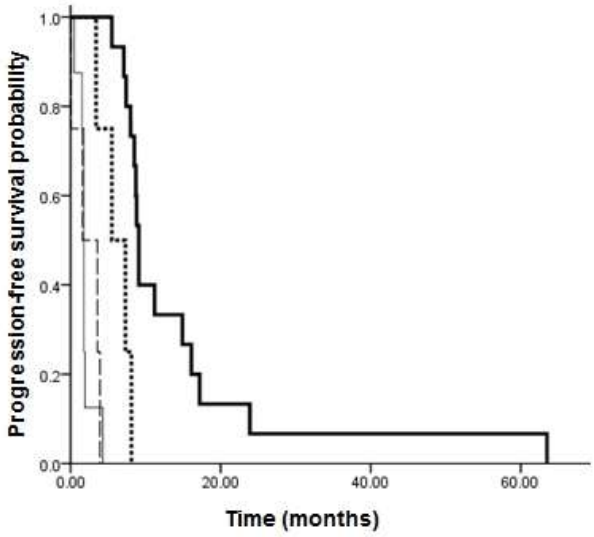
	4	G12S	2	34G>A
	5	G12R	2	34G>C
	6	G12C	2	34G>T
	7	G13D	2	38G>A
BRAF	1	V600E	15	1799T>A
PIK3CA	1	Q542K	9	1624G>A
	2	Q545K	9	1633G>A
	3	Q545G	9	1634A>G
	4	Q545D	9	1635G>T
	5	Q545E	9	1636C>G
	6	Q545K	9	1636C>A
	7	Q545P	9	1637A>C
	8	Q545R	9	1637A>G
	9	H1047Y	20	3139C>T
	10	H1047R	20	3140A>G
	11	H1047L	20	3140A>T

Table 2. Baseline patient characteristics and tumor molecular changes at screening

Characteristics		N
Age, years	Median (range)	65 (40-91)
Gender	Female	40
	Male	8
The line of EGFR-TKI	1	9
	2	31
	3	8
Molecular alterations at screening	EGFR	25 Exon 19 deletions-10 L858R-9 G719X-1 Double mutations-4 L858R & PIK3CA-1
	KRAS	4
	EML4-ALK	3
	KIF5B-RET	2
	CD74-ROS1	2
	PIK3CA	1*
	BRAF	0
	No known mutations	12

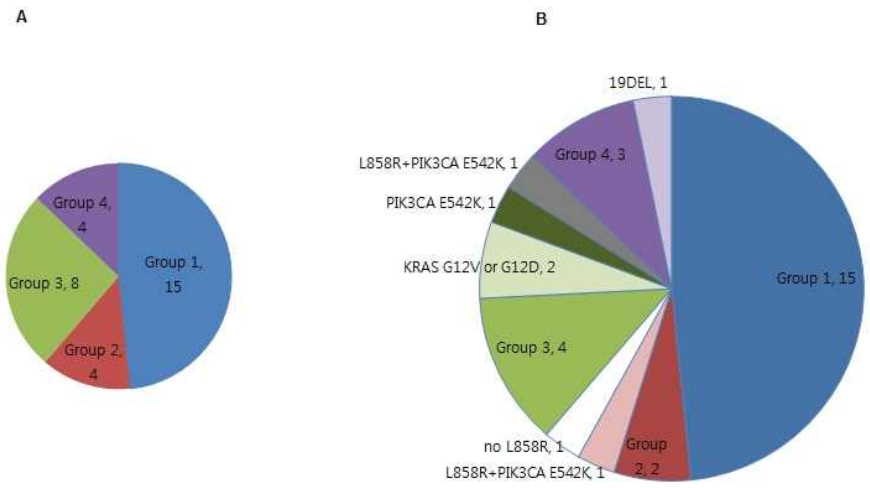
*Concomitant with EGFR L858R mutation

Figure 1. Kaplan-Meier plots of progression-free survival accoring to screening group.



	Median (months)	(95% CI)
Group 1 (n=15)	9.1	(8.6-9.6)
Group 2 (n=4)	1.6	(0.0-5.0)
Group 3 (n=8)	1.8	(1.6-2.0)
Group 4 (n=4)	5.5	(1.7-9.3)
P	<.0001	

Figure 2. Genomic alterations found by conventional screening (A) and targeted sequening (B).



◆ **Whole-genome analysis of a patient with early stage small cell lung cancer**

Introduction

Small-cell lung cancer (SCLC) is characterized by rapid proliferation and early development of widespread dissemination. Thus most patients diagnosed with advanced staged diseases. Although response rates to the initial therapy are impressive, most patients experience relapse within the first two years and die from systemic metastasis (Masters, et al. 2003). Despite the collaborative efforts to improve treatment, survival of SCLC has not been changes over the past 25 years (Lally, et al. 2007). Thus, more comprehensive understating of SCLC biology is needed to develop more effective treatment for this devastating disease. However, the lack of availability of surgical materials frequently hampers to study of SCLC biology. Recent advances in next generation sequencing technologies provide a means of discovering mutational processes across the whole genome (Pleasance, et al. 2010). Although a study has provided comprehensive genomic view of SCLC using massively parallel sequencing, it sequenced a SCLC cell line derived from a bone marrow metastasis of a SCLC patient (Pleasance, et al. 2010). Other recently published studies reported novel somatic driver mutations of SCLC by integrated analyses of various data sets generated by next generation sequencing method (Peifer, et al. 2012;Rudin, et al. 2012) . However, the complexity of SCLC and ethnic differences of somatic mutations for Lung cancer propose the possibility of other variations associated with SCLC of Korean ethnic group (Seo, et al. 2012). The importance of early detection of SCLC was also other point to focus our study for early stage SCLC.

Here, we analyzed whole genomes of matched normal-tumor samples from a patient underwent curative resection for stage IA SCLC in National Cancer Center, Korea. This paired normal-tumor comparison identified a comprehensive mutational profile of early stage SCLC. The data reported here increases our understanding of the pathogenesis of SCLC and will allow the development of more targeted therapies for SCLC.

Methods

Patient and specimen collection

Tumor and normal tissue samples were obtained from a patient with SCLC underwent curative resection. After the pathological examination, the tumor and normal tissue samples were snap frozen and maintained in liquid nitrogen until genomic DNA extraction. This study was conducted under the approval of the ethical review boards of National Cancer Center, Goyang, Gyeonggi, Korea (Assurance Number FWA00005026) and the guidelines for good clinical practice. Written informed consent was obtained from the patient for publication of this research article.

Genomic DNA preparation

The frozen tumor sample was micro-dissected and lightly stained with hematoxylin to identify the portion consisting of 80% or more cancer cells. The genomic DNA was extracted with MagAttract DNA Blood Midi Kit (Qiagen, Inc.) according to the manufacturer's protocols. The DNA quality was assessed with the use of F200 spectrophotometer (Tecan). A260/280 value greater than 1.7 was accepted for further analysis. The DNA quantity was assessed with the use of Qubit fluorometer (Invitrogen). Control DNA from matched normal tissue was processed in the same manner. The same frozen tumor samples were used for total RNA extraction using a QIGEN RNeasy Mini Kit (Qiagen, Inc). Quality of total RNA was assessed with lab-on-a-chip on the Agilent 2100 Bioanalyzer (Aglient Technologies, USA).

Whole genome sequencing

5 μg of genomic DNA was sheared using Covaris S series (Covaris, MS, USA). The fragment of sheared DNA was end-repaired, A-tailed, and ligated to pair end adapters according to

manufacturer's protocol (Pair End Library Preparation Kit, Illumina, CA, USA). Adapter ligated fragments were purified and dissolved in 30 μl of elution buffer, and 1 μl of the mixture was used as a template for 12 cycles of PCR amplification. The PCR product was gel purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Library quality and concentration was determined using Agilent 2100 BioAnalyzer (Agilent, CA, USA). Libraries were quantified using a SYBR green qPCR protocol on the LightCycler 480 (Roche, IN, USA) according to Illumina's library quantification protocol. Based on the qPCR quantification, libraries were normalized to 2nM and then denatured using 0.1 N NaOH. Cluster amplification of denatured templates occurred in flow cells, according to manufacturer's protocol (Illumina). Flow cells were paired-end sequenced (2x100bp) on Illumina HiSeq 2000 using HiSeq Sequencing kits. A base-calling pipeline (Sequencing Control Software, SCS; Illumina) was used to process the raw fluorescent images and the called sequences.

Reads alignment and variation detection

90bp paired-end sequence reads with ~300bp insert size were aligned to hg19 human reference genome (NCBI build 37) with a BWA algorithm ver. 0.5.9 (Li and Durbin 2009). Two mismatches were permitted in 45bp seed sequence. To remove PCR duplicates of sequence reads, which can be generated during the library construction process, we used the "rmdup" command of Samtools (Li, et al. 2009). Aligned reads were realigned at putative indel positions with the GATK IndelRealigner algorithm to enhance mapping quality. Base quality scores were recalibrated using the TableRecalibration algorithm of GATK (McKenna, et al. 2010) .

SNP and small insertion/deletion (Indel) analysis

Putative SNVs were called and filtered using UnifiedGenotyper and VariantFiltration commands in GATK (McKenna, et al. 2010). The options used for SNP calling were a minimum five to a maximum 200 read mapping depth with consensus quality 20, and the prior likelihood for heterozygosity value 0.001. To obtain somatic mutations in cancer genomes, SNVs from cancer genomes were filtered using the SNVs from normal tissue genomes. Remaining SNVs were filtered again using normal tissue genomes' mapping status. At each remaining tumor SNV position, if minimum mapping depth was at least three and SNV nucleotide ratio was at least 0.2 in normal tissue genomes, the tumor SNV was discarded. To obtain somatic small indels, IndelGenotyperV2 paired sample mode of GATK was used (McKenna, et al. 2010). Window size 300 was used and other default options were used. All somatic mutations altering amino acid were checked by expert lab personnel using the tview command of Samtools (Li, et al. 2009). If an SNV was in a low quality region or germ-line mutation, it was discarded. 21 novel indels in frameshift was validated using Samtools Tview with depth<5, depth rate<0.3, continuous mapped region<2 in both 10bp sides, and continuous mapped region rate<0.1 to remove false-positive results (Additional file 5: Table S5).

Annotation of variations

Predicted SNVs were compared with NCBI dbSNP version 131 (http://www.ncbi.nlm.nih.gov/projects/SNP/) to annotate known SNP information (Wheeler, et al. 2008). Each SNV was mapped on the genomic features of the UCSC gene table such as coding region, UTR, and intron. Non-synonymous SNV information was extracted by comparing UCSC (http://genome.ucsc.edu/) reference gene information. KEGG (http://www.genome.jp/kegg/) and Biocarta (http://www.biocarta.com/) pathways were used to analyze altered protein sets. Information on cancer-related mutations was obtained from the Cosmic cancer information database (http://www.sanger.ac.uk/genetics/CGP/cosmic/).

Identification of copy number variation regions

Due to the heterogeneity of cancer samples, a new method was developed for identifying copy number variations based on the differences of sequencing depths between normal and

cancer samples. The program defined the CNV regions containing the borders which presented significant differences by considering each pair of samples. Final CNV regions were defined through merging adjacent CNV regions of which copy numbers are similar to each other. To calculate frequencies of duplication or deletion events, the number of CNV regions in tumor samples which showed duplication or deletion was counted for all 23 chromosomes.

Genome-wide SNP analysis

SNP genotyping was performed using an Axiom genotyping solution including Axiom Genome-Wide ASI 1 Array Plate and reagent kit according to manufacturer's protocol (Affymetrix, CA, USA). Briefly, total genomic DNA (200 ng) was treated with 20 μl of Denaturation buffer and 40 μl Neutralization buffer, followed by amplification for 23 hr using 320 μl of Axiom amplification mix. Amplified DNA was randomly fragmented into 25 to 125 base pair (bp) size with 57 μl of Axiom fragmentation mix at 37 ºC for 30 min, followed by DNA precipitation for DNA clean up and recovery. DNA pellets were dried and resuspended with 80 μl of hybridization master mix. 3 μl of suspended sample was kept for sample qualification. A Hybridization ready sample was denaturated using a PCR machine at 95 ºC for 20 min and 48 ºC for 3 min. Denatured DNA was transferred to a Hybridization tray and loaded to a GeneTitan MC with Axiom ASI array plate (Affytmerix). Hybridization continued on the GeneTitan for 24 hr, followed by loading ligation, staining, and stabilization reagent trays into the instrument. GeneTitan was controlled by an Affymetrix GeneChip Command Console GeneTitan Control (Affymetrix). The chip image was scanned with the GeneTitan and the resulting data, a dat file, was automatically transformed to a cel file as a final intensity file. To genotype call, the cel intensity file was normalized, and genotype calling was done using Genotyping Console 4.1 with Axiom GT1 algorithms according to manufacturer's manual. The cut-off values for data quality control were DISHQC >= 0.82 for hybridization, and the QC call rate >= 97%.

Validation of SNVs by Sanger sequencing

SNVs were validated by conventional Sanger sequencing using dye-terminator chemistry analyzed with an automatic sequencer ABI 3730 (Applied Biosystems, CA, USA). The target regions were amplified by PCR followed by direct sequencing, or cloned into TA vectors. At least 20 TA vector clones were sequenced, because mutations with low purity are difficult to detect by Sanger sequencing.

Structural variants (SVs) and gene fusion analysis

SVs are analyzed using breakdancer (Chen, et al. 2009). An SV not found in normal tissue samples is defined to somatic SV. Gene fusions are analyzed using the SVs by defining fusion signals which inform fusion point and fusion direction. We excluded fusion signals which found from not only mate normal tissue same but also other 19 samples. We decided two fusion signals are equal, if breakpoints are distant smaller than 1kbp. When we scan gene fusion event, fusion signals not locate in two genes, and gene fusions making opposite transcription of a constitute gene are excluded. Genes having SV breakpoint are decided to suffer gene breakage event. When define the gene breakage, SVs locate in an intron are also excluded.

Results

A 57 year-old Korean man with former smoking history (average 20 cigarettes per day for 15 years) presented to our hospital with an asymptomatic screening detected early stage lung cancer in June 2006. Ultimately, stage IA (by AJCC, the 6thedition)SCLCoftherightupperlobewithoutregionallymphnodeinvolvementwasdiagnosed.Hewastreatedwithrightupperlobelobectomyandthefinalpathologicstagewas T2N0M0.Herecei vedfourcyclesofadjuvantchemotherapywithirinotecanandcisplatinaftersurgery.Sofar,thepatientisalivewithoutotherrecurre

ncesofhisdisease(January 2013).

On average, 99 gigabases per sample were produced at ~33X sequencing depth, and they were mapped to the reference genome (NCBI build 37, HG19) at an over 95% mapping rate (Additional file 1: Table S1). Among the mapped reads, the reads inside of the mean/standard deviation range (properly mapped reads) were over 93% and the reads that didn't make it into the contigs (singletons) were less than 5%. Therefore, we got enough amounts of qualified sequencing reads to cover whole genome. Using the final properly mapped reads, we constructed a genomic profile database for detecting single nuclear variations (SNVs), copy number variations (CNVs), structural variations (SVs) and fusion genes.

Structural variation analysis showed that genome duplications and deletions occurred randomly (Figure 1A). In tumor genome, there are 989 deletions, 910 insertions, 45 inversions, 89 intra-chromosomal translocations, and 24 inter-chromosomal translocations of breakdancer score ³ 80 (Additional file: Table S2). In overall, deletion events occurred twenty times more in length than duplication, suggesting that genome-wide damage by deletion has a higher impact than amplification in SCLC. The structural variants in genic regions were further analyzed by comparing with the cancer-related genes in COSMIC (Catalog Of Somatic Mutations In Cancer, http://www.sanger.ac.uk/genetics/CGP/cosmic/) database. Total six genes showed more than 300 bp of large structural variations including 4 deletions and 2 inversions (Additional file: Table S3). The copy number variation (CNV) analysis revealed that Chromosomes 4p, 5q, 13q, 15q, 17p, and 22q contained many blocks of deletion. In contrast, chromosomes 5p showed notably increased duplicated blocks (Figure 1A). When the exact loci of tumor suppressors and oncogenes were analyzed, we found that tumor suppressors, *RB1* at chromosome 13q and *TP53* at chromosome 17p were in highly deleted chromosomal regions. In contrast, the copy number of oncogene, *hTERT* at chromosome 5p was increased. It is reported that telomerase gene amplification appears to increase both *hTERT* mRNA expression and telomerase activity, which are necessary during the early phase of carcinogenesis. Increased *hTERT* mRNA and telomerase activity are frequently observed in SCLC (Nishio, et al. 2007). These finding suggest that increased telomerase activity by *hTERT* amplification may occur early in tumorigenic transformation and initiate SCLC.

We also identified about 0.5 million short insertions and deletions (indel) in normal and tumor tissues. By subtracting indels found in the normal genome, 6,430 somatic indels specific for SCLC were identified (Additional file: Table S4). Among them, 21 novel indels from 21 genes resulted in frameshifts (Additional file: Table S5), which most likely damage protein function. Interestingly, they are not previously reported as cancer-related genes.

By sequencing normal and tumor genomes, we identified 3.6 million SNVs. The results of NGS were also examined with Axiom genome-wide genotyping microarrays (Affymetrix, USA). The genotyping data from microarrays showed 99.8% of concordance with the NGS data. By subtracting SNVs found in the normal genome, we were able to identify 62,763 somatic SNVs specific for SCLC (Table 1). Among them, there were 43,339 novel SNVs including 116 nsSNVs from 108 genes. A part of nsSNVs were validated and confirmed by conventional Sanger sequencing (Table 2). Additionally, we analyzed the pattern of single nucleotide substitutions in total somatic SNVs. The result shows that G>A/C>T (28%) and A>G/T>C (24%) transitions were more common than G>T/C>A (19%) transversions (Figure 1B). The mutation patterns in this study are somewhat different from that of heavy-smoking related lung cancer. Only three types of mutations showed consistent differences in relation with tobacco smoking (Govindan, et al. 2012;Pleasance, et al. 2010). Both G>T/C>A transversions and A>G/T>C transitions elevated in ever smokers compared with never smokers, while G>A/C>T transitions are decreased in a progressive manner with cumulative exposure to tobacco. In our study, the patient was a former smoker with a 15 pack-year smoking history and quitted smoking 6 years ago before the diagnosis of SCLC. Thus the relatively higher frequency of G>A/C>T transitions (29%) compared with G>T/C>A transversions (19%) in Figure 1B could be caused by mild smoking habit of the patient.

Supporting Information Table S1: Sequencing Statistics.

	Normal	Tumor
Total bases (Gb)	101.46	97.24
Sequencing depth (fold)	35.46	33.98
Mapped bases (Gb)	97.3	94.18
Mapping depth (fold)	34	32.91
Total reads	1,127,335,789	1,080,412,496
Mapped reads	1,081,083,269 (95.90%)[a]	1,046,418,608 (96.85%)[a]
Properly paired reads	1,057,740,768 (93.83%)[b]	1,026,271,891 (94.99%)[b]

[a]Percentage of the mapped reads to the total reads.

[b]Percentage of the properly paired reads to the total reads.

[c]Percentage of the singletons to the total reads.

Supporting Information Table S2: Summary of genomic rearrangements.

Genomic Rearrangements	Count	Percentage
Deletion	989	48.1%
Insertion	910	44.2%
Inversion	45	2.2%
Intra-chromosomal translocation	89	4.3%
Inter-chromosomal translocation	24	1.2%

Supporting Information Table S3: List of genomic rearrangements of known cancer-related genes.

Gene	Chromosome position	NM ID[a]	NP ID[b]	Structural variation type
ALPK1	4q25	NT_016354.19	NP_001095876.1 NP_001240813.1 NP_079420.3	Deletion
CAMK2A	5q32	NT_029289.11	NP_057065.2 NP_741960.1	Deletion
MAP4K3	2p22.1	NT_022184.15	NP_003609.2	Inversion
RPS6KA2	6q27	NT_025741.15	NP_001006933.1 NP_066958.2	Inversion
PRKCB	16p11.2	NT_010393.16	NP_002729.2 NP_997700.1	Deletion
TRIP3	20p13-p12.2	NT_011387.8	NP_066081.2	Deletion

[a]Nucleotide sequence accession number. [b]Protein sequence accession number.

Supporting Information Table S4: Summary of short somatic insertions and deletions.

Variant type	Total
Indels	6,430
coding region	2,570
frameshift	21
x3	8
3′ UTR	28
5′ UTR	8
INTRON	2,179
PROMOTER	326
non-coding region	158
3′ UTR	5
5′ UTR	2
INTRON	108
PROMOTER	43
Intergenic region	3,702

Supporting Information Table S5: Gene list resulted in frameshift.

Gene	Chromosome[a]	Start position[b]	End position[b]	Type	Validation using Tview	Novelty[c]
CR1L	chr1	207870892	207870893	-G	validated	Yes
SIK3	chr11	116733010	116733010	+C	validated	Yes
OR11G2	chr14	20666521	20666521	+G	validated	Yes
C14orf4	chr14	77492131	77492132	-C	validated	Yes
SPEG	chr2	220349692	220349692	+G	depth<5	Yes
PLA2G4E	chr15	42302337	42302338	-A	depth<5	Yes
FAM22A	chr10	88988495	88988495	+G	depth rate<0.3	Yes
HOXB6	chr17	46675492	46675493	-T	depth rate<0.3	Yes
TAF10	chr11	6633018	6633018	+G	continuous mapped region<2	Yes
KRT4	chr12	53207602	53207602	+G	continuous mapped region<2	Yes
GOLGA3	chr12	133363087	133363088	-C	continuous mapped region<2	Yes
RASGRF1	chr15	79298649	79298649	+G	continuous mapped region<2	Yes
KIF7	chr15	90176150	90176152	-AG	continuous mapped region<2	Yes
NTN1	chr17	9143044	9143044	+G	continuous mapped region<2	Yes
ACCN1	chr17	31618659	31618659	+G	continuous mapped region<2	Yes
IGFBP4	chr17	38600213	38600213	+C	continuous mapped region<2	Yes
PPM1E	chr17	56833424	56833424	+C	continuous mapped region<2	Yes
SMCHD1	chr18	2656085	2656085	+G	continuous mapped region<2	Yes
TRIM67	chr1	231298883	231298883	+C	continuous mapped region<2	Yes
PID1	chr2	230135746	230135746	+C	continuous mapped region<2	Yes
EOMES	chr3	27763111	27763111	+C	continuous mapped region<2	Yes

[a]Chromosome on which the variation was located. [b]Nucleotide position of the variant allele in the human reference genome sequence version 19/build 36. [c]Novelty means that results not included in dbSNP.

Figure 1. A. Variations of tumor genome. From the outer side of each ring, chromosome numbers and mapping depths of chromosome regions were indicated by numbers. Large translocation of intra and inter chromosomal rearrangements were indicated by color lines across center. **B. Single nucleotide substitution pattern in somatic mutations.**

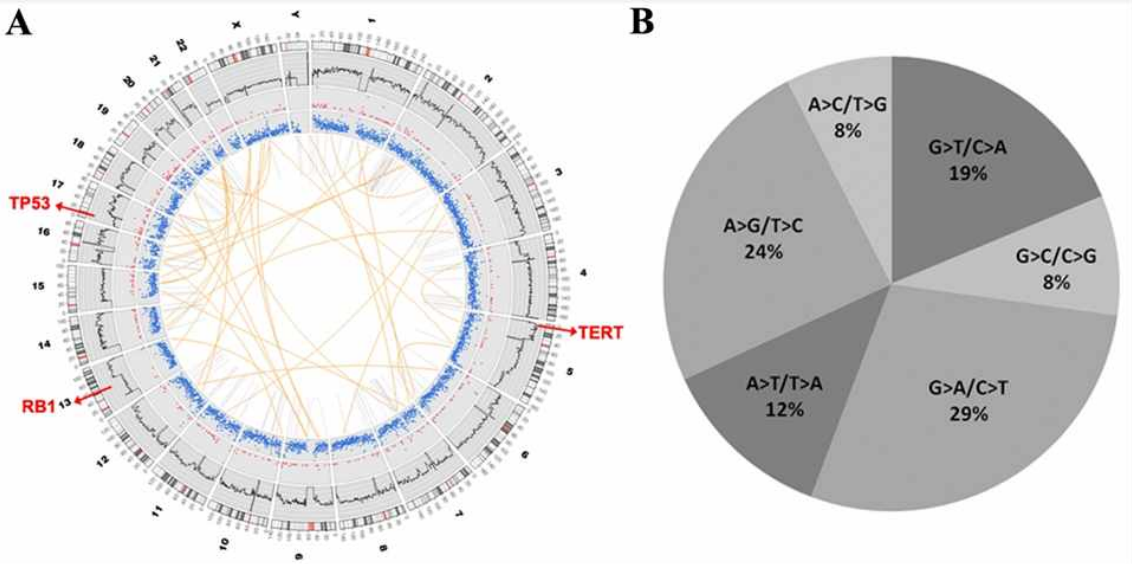
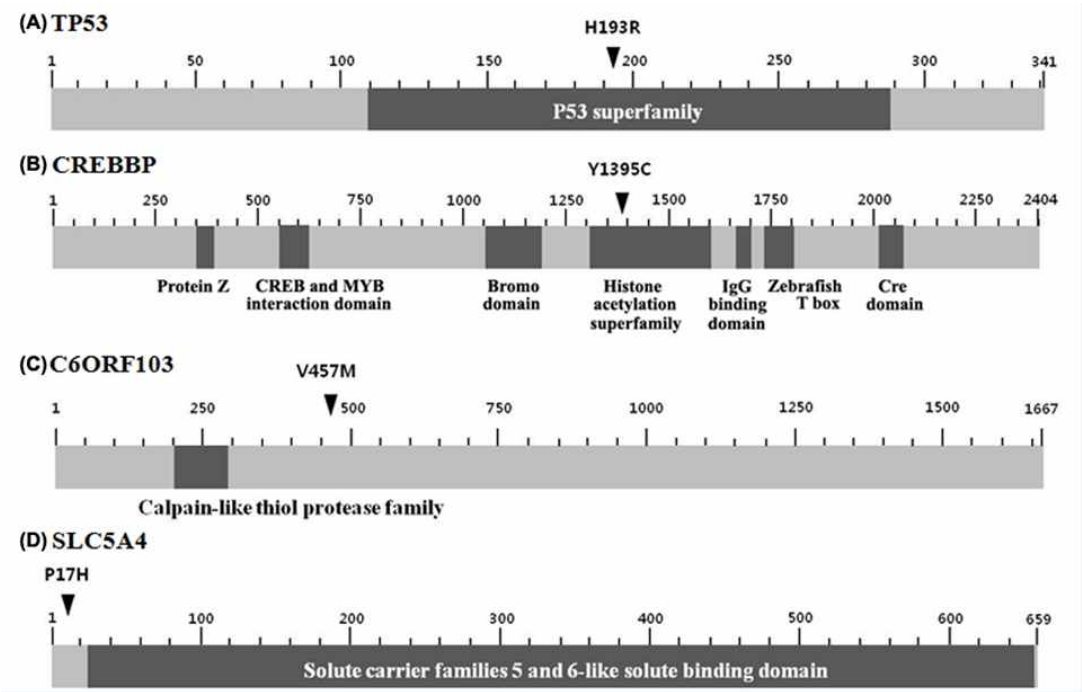


Figure 2. Schematic of TP53, CREBBP, C6ORF103, SLC5A4 protein illustrating functional domains with the location of their mutations.



◆ **Whole exome sequencing (WES) and RNA sequencing of never-smoking lung adenocarcinoma without known driver mutations**

<WES 분석결과>

1) DNA QC

- 엑솜시퀀싱을 위한 DNA QC 통과 기준

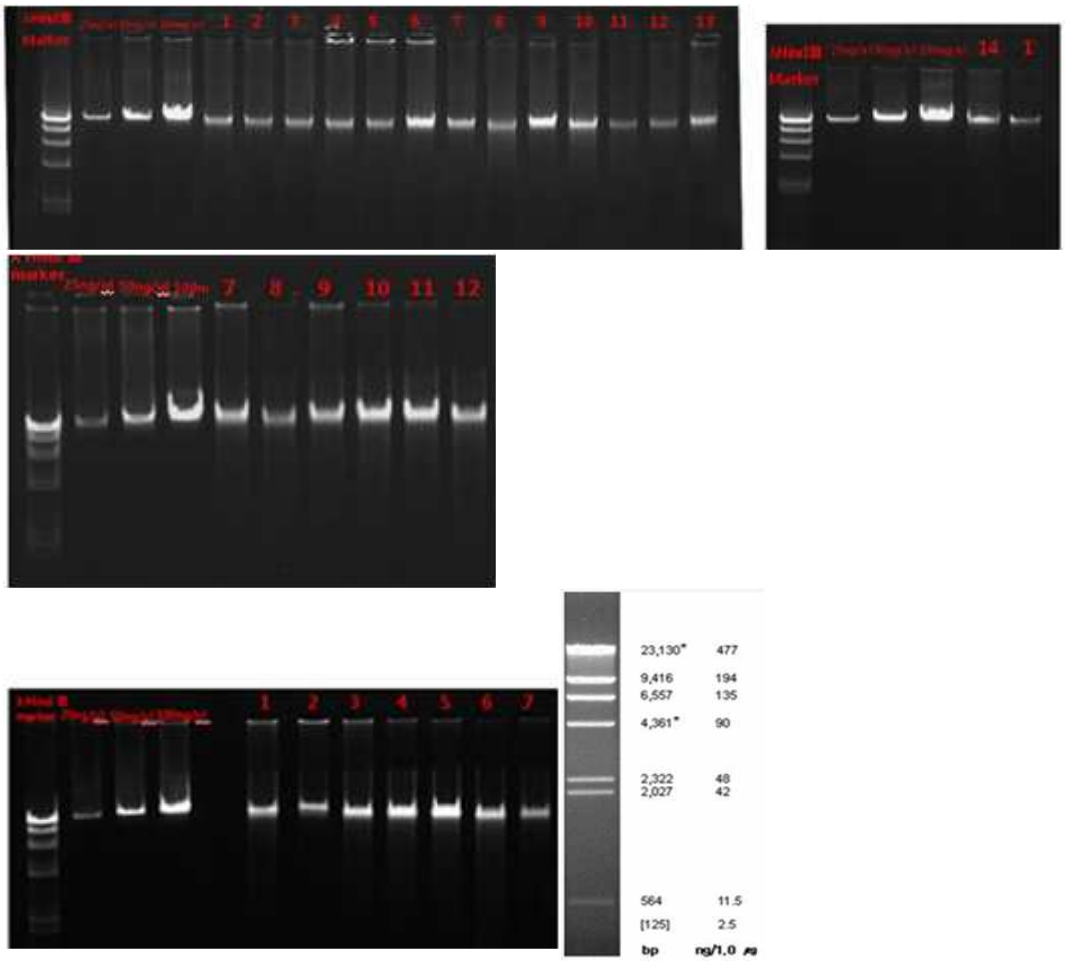
분석 항목	정성 분석		정량 분석
	UV 농도 측정	전기영동 (DNA degradation 정도)	형광 농도 측정 (dsDNA양 기준)
측정 기준	○ 적합도 범위 - OD260/280: 1.6 ~ 2.3 - OD260/230 : > 1.6	○ 0.7% agarose gel 상에서 DNA degradation유무 확인	○적합도 범위 - ≥ 100 ng/ul

- 농도 및 총량 QC 결과

Delivery ID	Tumor_ID	Conc_ Quant-iT (ng/ul)	Volume_ 잔량 (ul)	Fluorescence 기준_ 총량 (ug)	Conc_ UV (ng/ul)	260/280	260/230
S07-5165T	T1301D0131	146	40	5.84	231	1.8	2.04
S05-9795T	T1301D0132	126	40	5.04	206.6	1.79	2.12
S05-15444T	T1301D0133	155	40	6.2	231	1.81	2.26
S04-13604T	T1301D0134	295	40	11.8	284.7	1.84	1.99
S04-63T	T1301D0135	150	40	6	247	1.84	2.05
S04-3413T	T1301D0136	256	40	10.24	389.4	1.81	2.16
S06-7551T	T1301D0137	184	40	7.36	273	1.82	2.21
S07-2404T	T1301D0138	236	40	9.44	374.6	1.79	1.83
S01-2235T	T1301D0139	251	40	10.04	347.3	1.81	2.19
S02-380T	T1301D0140	229	40	9.16	336.7	1.83	2.21
S04-4460T	T1301D0141	110	40	4.4	184.7	1.72	1.38
S05-4996	T1301D0142	112	40	4.48	171.1	1.78	2.1
S03-1963T	T1301D0143	259	40	10.36	376.1	1.82	2.23
S08-1314T	T1301D0144	211	40	8.44	292	1.82	2.17
S07-5165B	T1212D1699	16.2	100	1.62	27.5	1.71	1.14
S05-9795B	T1212D1695	55.8	65	3.63	85.9	1.83	1.42

S05-15444B	T1212D1696	43.2	85	3.67	73.1	1.84	1.27
S04-13604B	T1212D1688	34.1	105	3.58	62.7	1.77	1.32
S04-63B	T1212D1691	56.6	60	3.40	95.9	1.83	1.53
S04-3413B	T1212D1692	50.4	90	4.54	77.5	1.88	1.59
S06-7551B	T1212D1697	63.8	65	4.15	95.2	1.82	1.43
S07-2404B	T1212D1698	62.7	60	3.76	95.2	1.77	1.39
S01-2235B	T1303D1144	57.3	17	0.97	85	1.8	1.39
S02-380B	T1303D1145	44.0	22	0.97	65.1	1.76	1.16
S04-4460B	T1303D1147	64.9	12	0.78	104.9	1.85	1.62
S05-499B	T1303D1148	13.5	70	0.95	30.5	1.82	1.01
S03-1963B	T1303D1146	61.9	18	1.11	88.8	1.82	1.3
S08-1314B	T1302D0822	66.2	45	2.98	98.3	1.81	1.72

<사진> 전기영동 결과



- Agarose 농도: 0.7%
- 전기영동 시간: 30분
- DNA size marker: λHindⅢ marker

2) 엑솜시퀀싱 데이터 기본분석 결과

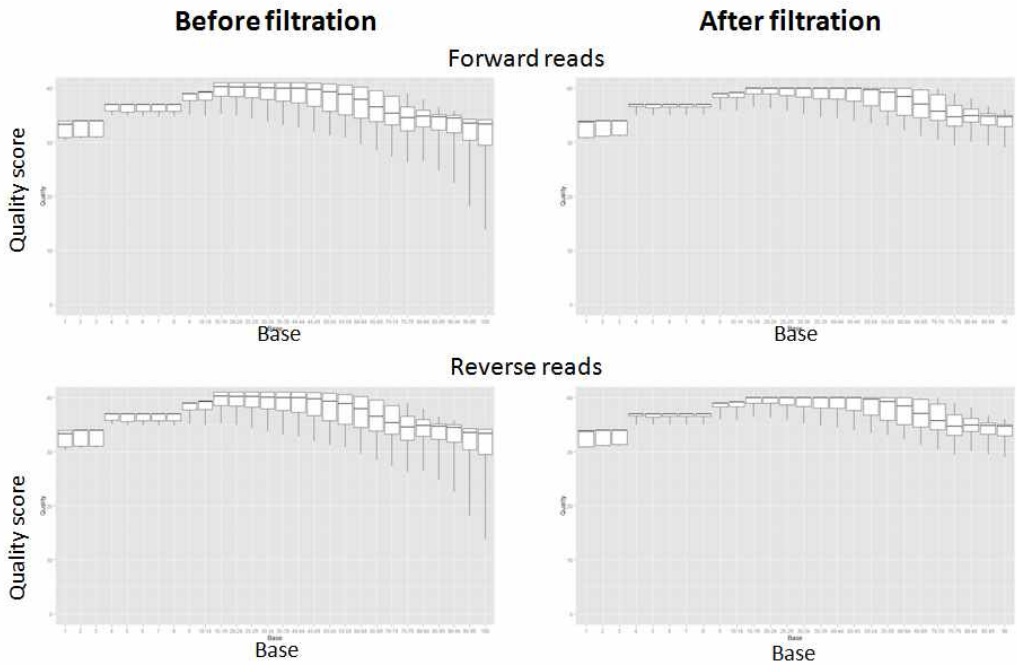
2-1) 시퀀싱 데이터 필터링 수행

2-1-1) 다음과 같은 조건과 함께 low-quality sequence들은 제거됨

·Read의 3'-end로부터 10-bp trimming 시킴
·한 read에서'N'개수가 10% 이상일 경우
·한 read의 평균 quality score (QS)가 20 미만일 경우
·한 read에서 QS 20보다 작은 nucleotides이 5% 이상일 경우
·Paired reads이 동일 할 경우

2-2) 시퀀싱 데이터 퀄리티 결과

2-2-1) 시퀀싱 데이터 필터링 전/후 결과 비교: 필터링 전/후 paired-end sequence들에서 관찰된 base quality의 상태는 비교적으로 비슷한 퀄리티 스코어를 나타냄 (원시데이터 자체가 기본적으로 높은 퀄리티를 가지고 있음을 의미)



<그림> 엑솜 시퀀싱 데이터 필터링 전/후 퀄리티 결과 비교

2-3) 시퀀싱 데이터 생산량 및 필터링 전/후 퀄리티 통계 결과

Sample	TotalReads	MappedReads	MappedRate	ProperlyPairedReads	ProperlyPairedRate
T1212D1688	63918902	63608509	99.51%	63434594	99.24%
T1212D1691	59156302	58727287	99.27%	58560818	98.99%
T1212D1692	53877774	53487976	99.28%	53339148	99.00%
T1212D1695	73108406	72750466	99.51%	72435656	99.08%
T1212D1696	73768479	73360551	99.45%	73048730	99.02%
T1212D1697	50618433	50308870	99.39%	50178306	99.13%
T1212D1698	60351154	60076931	99.55%	59940838	99.32%
T1212D1699	67987909	67641791	99.49%	67224264	98.88%
T1301D0131	141916270	141163491	99.47%	140736914	99.17%
T1301D0132	78021379	77601728	99.46%	77306272	99.08%
T1301D0133	136661182	135897625	99.44%	135527824	99.17%
T1301D0135	130268753	129560795	99.46%	129129332	99.13%
T1301D0136	98792196	98030072	99.23%	97650002	98.84%
T1301D0137	147459591	146296447	99.21%	145808746	98.88%
T1301D0138	156687544	155884126	99.49%	155322864	99.13%
T1301D0139	86328572	85781581	99.37%	84203008	97.54%
T1301D0140	77214151	76761221	99.41%	75256832	97.47%
T1301D0141	100686421	99939922	99.26%	98993234	98.32%
T1301D0142	93487108	92974209	99.45%	91451490	97.82%
T1301D0143	127141139	126482817	99.48%	125593108	98.78%
T1301D0144	145495675	144848986	99.56%	143916638	98.91%
T1302D0822	53436525	53125282	99.42%	52711446	98.64%
T1303D1144	72746117	72399711	99.52%	72048928	99.04%
T1303D1145	39451783	39245233	99.48%	38947378	98.72%
T1303D1146	37456735	37290293	99.56%	37033006	98.87%
T1303D1147	70852811	70507376	99.51%	70142544	99.00%
T1303D1148	56765558	56512651	99.55%	56206244	99.01%

Sample	WithItselfAndMateMappedReads	WithItselfAndMateMappedRate	SingletonReads	SingletonRate
T1212D1688	63498756	99.34%	109753	0.17%

T1212D1691	58620820	99.09%	106467	0.18%
T1212D1692	53390787	99.10%	97189	0.18%
T1212D1695	72613749	99.32%	136717	0.19%
T1212D1696	73223427	99.26%	137124	0.19%
T1212D1697	50220844	99.21%	88026	0.17%
T1212D1698	59996511	99.41%	80420	0.13%
T1212D1699	67541092	99.34%	100699	0.15%
T1301D0131	140929079	99.30%	234412	0.17%
T1301D0132	77462056	99.28%	139672	0.18%
T1301D0133	135681738	99.28%	215887	0.16%
T1301D0135	129344036	99.29%	216759	0.17%
T1301D0136	97852759	99.05%	177313	0.18%
T1301D0137	145986279	99.00%	310168	0.21%
T1301D0138	155591066	99.30%	293060	0.19%
T1301D0139	85559344	99.11%	222237	0.26%
T1301D0140	76558690	99.15%	202531	0.26%
T1301D0141	99672656	98.99%	267266	0.27%
T1301D0142	92745840	99.21%	228369	0.24%
T1301D0143	126270355	99.32%	212462	0.17%
T1301D0144	144617704	99.40%	231282	0.16%
T1302D0822	53043576	99.26%	81706	0.15%
T1303D1144	72286019	99.37%	113692	0.16%
T1303D1145	39174370	99.30%	70863	0.18%
T1303D1146	37230779	99.40%	59514	0.16%
T1303D1147	70399614	99.36%	107762	0.15%
T1303D1148	56423709	99.40%	88942	0.16%

3) On-target 해당 엑솜 시퀀스 선발

3-1) 선발과정

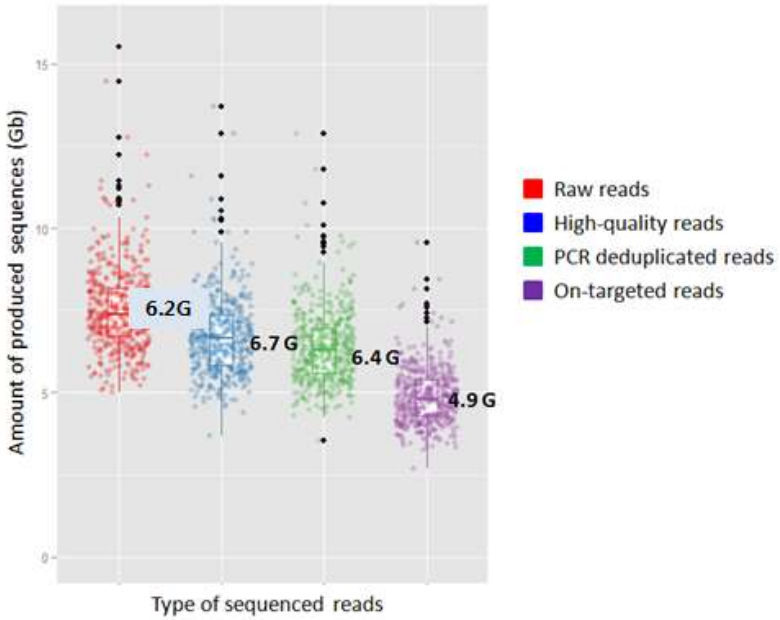
- 원시 시퀀싱 데이터로부터 필터링 과정을 거쳐 선발된 high-quality 데이터는 Burrows-Wheeler Aligner (BWA)를 이용하여 human genome reference (hg19)에 맵핑됨

- PCR duplicates들이 제거됨

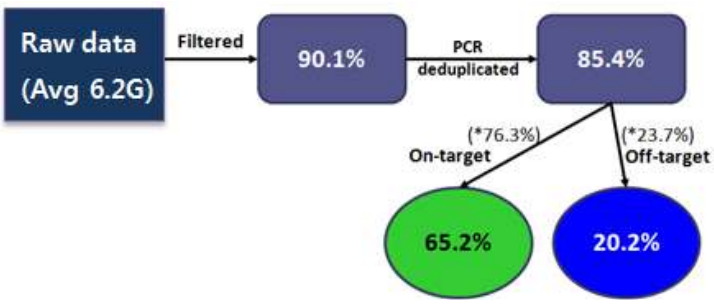
- 엑솜 타겟 지역들(TruSeq kit에 포함된 타겟 지역들에 해당)을 토대로 on-target 지역에 걸쳐있는 데이터들을 선발함 (off-target 지역: 타겟 지역으로부터 벗어난 지역으로, 이 지역에 맵핑된 시퀀스들은 배제됨)

- 엑솜 시퀀싱 데이터의 생산량 결과 요약 (아래 그림 참조)

-시퀀싱된 샘플(500 샘플)의 평균 생산 시퀀스의 개수는 약 62.2 million (±1.2)으로 약 6.2 Gb의 데이터를 생산함 (원시데이터 생산에 근거)
- High-quality를 가진 시퀀스가 선발됨 (약 5.9 Gb의 데이터 생산)
- PCR deduplication을 통해 약 5.6 Gb의 데이터를 생산함 (high-quality 시퀀스들 중 4.7%가 제거됨. 따라서 원시데이터 중 85.4%의 데이터가 선발됨)
- 5.6 Gb 중에서 온타겟 데이터는 약 4.2 Gb (85.4% 중에서 65.2%), 오프타겟 데이터는 약 1.4 Gb (85.4% 중에서 20.2%)로 평가됨. 본 과제에서는 온타겟 데이터에 한정하여 분석이 수행됨



<그림> 엑솜 시퀀싱 데이터의 생산량 예시



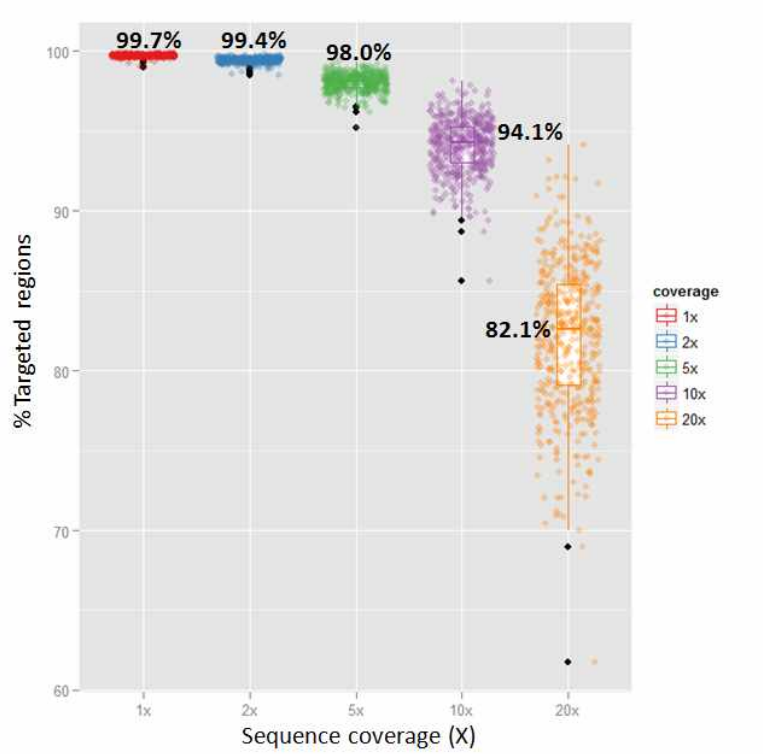
<그림> 엑솜 시퀀싱 데이터 생산 모식도

4) 온타겟 지역에서의 시퀀싱 상태

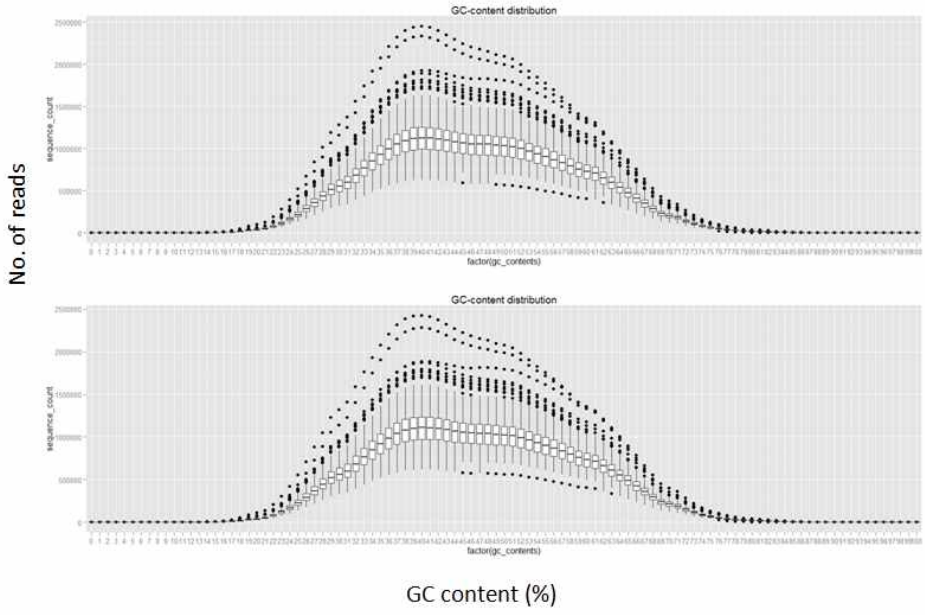
4-1) 원시 시퀀싱

- Sequence depth: 온타겟 지역에서 각 샘플 당 평균 42.4X의 sequence depth를 나타냄. 또한 주로 40X에서 45X depth 사이에서 높은 분포를 보임

- Sequence coverage (X) 비교: 시퀀싱된 전체 샘플들을 대상으로 온타겟 지역에서 sequence coverage (1X, 2X, 5X, 10X, 20X 일 때)를 비교함. 중간값을 토대로 분석하였을 때, 5X 이상에 대해서는 98%, 10X 이상에 대해서는 94.1%, 20X 이상에 대해서는 82.1%가 온타겟 지역으로 평가됨.

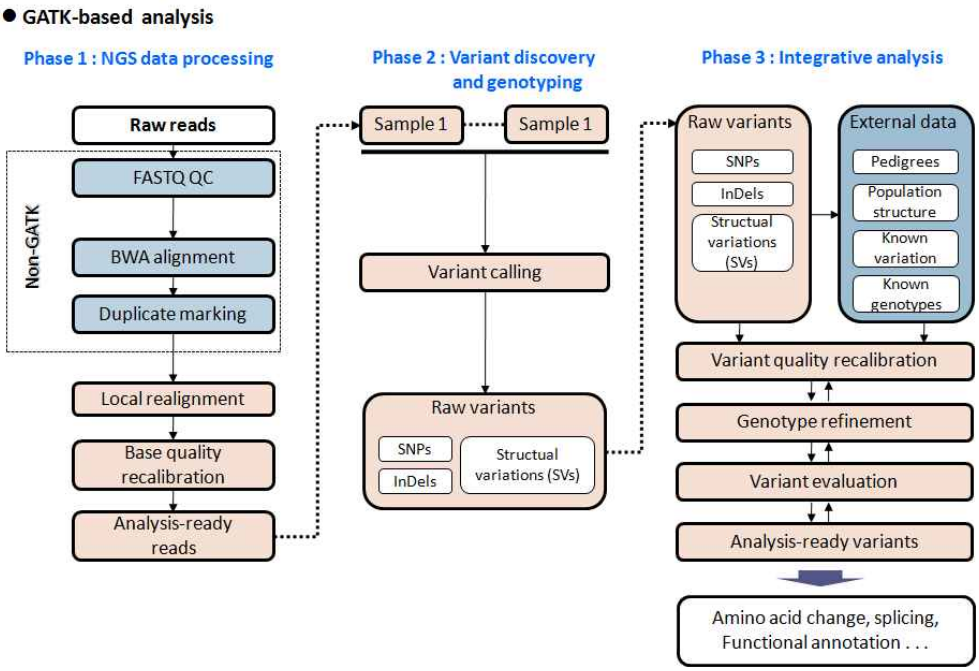


- GC content: 온타겟에 해당하는 forward/reverse 시퀀스들을 상대로 GC content(%)를 조사하였을 때, 약 40% 주위로 높은 분포가 확인됨. 이것은 온타겟 지역이 잘 캡쳐 되었음을 의미함



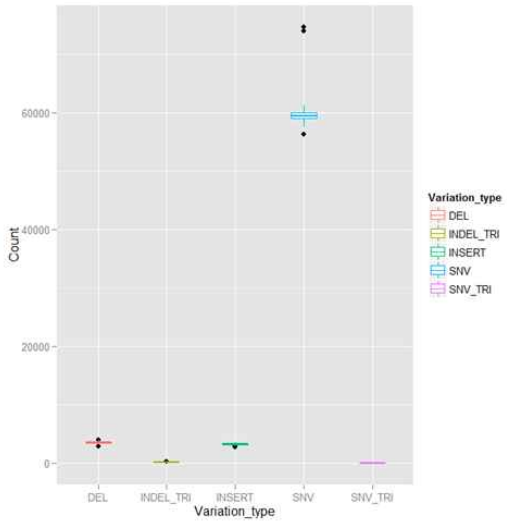
5) 각 샘플에서의 variant calling

5-1) GATK를 이용하여 각 엑솜 시퀀싱된 샘플들로부터 변이 탐색 (기본 분석)



5-2) Variant calling 요약

- 각 샘플당 평균 SNV는 46958개, deletion은 2444개, Insertion은 2159개로 확인됨 (아래 그림 참조).



시퀀싱된 샘플에서 확인된 variant들의 개수

6) Somatic 및 Germline 변이 선발

6-1) Tumor, normal pair 샘플을 이용한 Somatic mutation calling

- Mutect 툴을 이용하여 tumor와 normal 샘플로부터 Somatic과 Germline 변이를 선발하였음.

- 선발된 Somatic mutation 중, CDS 영역에 있는 평균 SNV 숫자는 65개, UTR 영역에 있는 평균 SNV 숫자는 50개였음.

- 선발된 Germline mutation 중, CDS 영역에 있는 평균 SNV 숫자는 29684개, UTR 영역에 있는 평균 SNV 숫자는 34501개였음.

6-2) EGFR driver mutation 분석

- EGFR driver mutation의 경우, 기존에 진행하였던 Targeted sequencing(Ion Torrent cancer panel) 결과와 비교 분석하였음.

- Cancer panel sequencing의 경우, 평균 해독배수가 1000~2000X이기 때문에, sensitivity가 엑솜시퀀싱보다 훨씬 높음.

- EGFR driver mutation(L858R, exon19 deletion) 분석결과

	Delivery_ID	Theragen_ID	WES	Cancer panel
Group 1	T4 (S04-13604T)	T1301D0134	L858R	L858R
	T10 (S04-63T)	T1301D0135	Nonframeshift exon19 deletion	L747-S752del
	T12 (S04-3413T)	T1301D0136	Nonframeshift exon19 deletion	L747-S752del
	T25 (S06-7551T)	T1301D0137	L858R	L858R

	T28 (S07-2404T)	T1301D0138	L858R	L858R
Group 2	S05-15444T	T1301D0133	L858R	L858R
	S05-9795T	T1301D0132	L858R	L858R
Group 3	S07-5165T	T1301D0131	ND	ND
	S01-2235T	T1301D0139	ND	ND
	S02-380T	T1301D0140	ND	ND
	S04-4460T	T1301D0141	ND	ND
	S05-4996	T1301D0142	ND	ND
Group 4	S03-1963T	T1301D0143	ND	ND
	S08-1314T	T1301D0144	ND	ND

6-3) 기타 somatic mutation 분석결과

- Known Somatic mutatin target SNP list(Total depth 이상이며, 아미노산 변화를 일으켜, protein function에 영향을 미치는 변이 중 cosmic DB에 등록된 변이): 34개

Sample	Chr	Start	End	Ref	Alt	Gene	cosmic number
T4 (S04-13604T)	7	55,259,515	55,259,515	T	G	EGFR	COSM6224
T4 (S04-13604T)	15	88,679,150	88,679,150	C	A	NTRK3	COSM124879
T4 (S04-13604T)	17	7,578,534	7,578,534	C	G	TP53	COSM301221, COSM301222, COSM43963,COSM301220
T10 (S04-63T)	17	45,663,749	45,663,749	G	A	NPEPPS	COSM980459
T12 (S04-3413T)	12	52,452,615	52,452,615	C	A	NR4A1	COSM468526
T12 (S04-3413T)	17	45,663,749	45,663,749	G	A	NPEPPS	COSM980459
T25 (S06-7551T)	7	55,259,515	55,259,515	T	G	EGFR	COSM6224
T25 (S06-7551T)	12	31,250,875	31,250,875	G	C	DDX11	COSM304699
T25 (S06-7551T)	15	28,230,247	28,230,247	C	T	OCA2	COSM307620)
T25 (S06-7551T)	17	45,234,327	45,234,327	C	T	CDC27	COSM1179393 ,COSM222498
T28 (S07-2404T)	7	55,241,677	55,241,677	G	A	EGFR	COSM12988
T28 (S07-2404T)	7	55,259,515	55,259,515	T	G	EGFR	COSM6224
T28 (S07-2404T)	12	113,515,335	113,515,335	T	G	DTX1	COSM356740

T28 (S07-2404T)	17	45,663,749	45,663,749	G	A	NPEPPS	COSM980459
T28 (S07-2404T)	17	7,578,528	7,578,528	A	C	TP53	COSM11319;
T19(S05-9795T)	7	55,259,515	55,259,515	T	G	EGFR	COSM6224
T22(S05-15444T)	7	55,259,515	55,259,515	T	G	EGFR	COSM6224
T22(S05-15444T)	17	45,663,749	45,663,749	G	A	NPEPPS	COSM980459
T01(S01-2235T)	19	11,144,113	11,144,113	G	T	SMARCA4	COSM255878
T06(S02-380T)	12	106,820,987	106,820,987	C	T	POLR3B	ICOSM246715
T14(S04-4460T)	9	136,340,607	136,340,607	G	T	SLC2A6	COSM608183
T18(S05-4996T)	13	20,006,639	20,006,639	T	A	TPTE2	COSM1128406,COSM1128405
T29(S07-5165T)	1	25,891,685	25,891,685	G	A	LDLRAP1	COSM414511
T29(S07-5165T)	2	172,641,873	172,641,873	C	T	SLC25A12	COSM126218
T29(S07-5165T)	7	55,259,524	55,259,524	T	A	EGFR	COSM6213
T29(S07-5165T)	9	120,474,743	120,474,743	C	A	TLR4	COSM95359
T08(S03-1963T)	2	209,195,248	209,195,248	C	A	PIKFYVE	COSM1179148
T08(S03-1963T)	20	42,088,521	42,088,521	T	A	SRSF6	COSM166813
T08(S03-1963T)	22	26,853,885	26,853,885	G	T	HPS4	COSM478827, COSM478828state)
T30(S08-1314T)	11	125,476,271	125,476,271	C	T	STT3A	COSM466531
T30(S08-1314T)	11	66,032,690	66,032,690	G	T	KLC2	COSM1127678
T30(S08-1314T)	12	52,452,615	52,452,615	C	A	NR4A1	COSM468526
T30(S08-1314T)	15	72,873,082	72,873,082	C	A	ARIH1	COSM471033
T30(S08-1314T)	X	142,718,314	142,718,314	C	T	SLITRK4	COSM205441

- Unknown Somatic mutatin target SNP list(Total depth 이상이며, 아미노산 변화를 일으켜, protein function에 영향을 미치는 변이 중 상위 15개)

Sample	Chr	Start	End	Ref	Alt	Gene
T10 (S04-63T)	19	56,190,084	56,190,084	G	T	EPN1
T12 (S04-3413T)	11	4,615,654	4,615,654	C	A	OR52I1
T12 (S04-3413T)	12	83,251,082	83,251,082	A	C	TMTC2
T12 (S04-3413T)	19	39,371,504	39,371,504	C	G	SIRT2
T25 (S06-7551T)	8	33,406,949	33,406,949	C	T	RNF122
T25 (S06-7551T)	X	151,366,116	151,366,116	C	T	GABRA3
T19(S05-9795T)	7	126,173,751	126,173,751	G	T	GRM8
T01(S01-2235T)	17	59,479,224	59,479,224	G	T	TBX2

T06(S02-380T)	X	118,985,541	118,985,541	C	T	UPF3B
T14(S04-4460T)	14	70,419,015	70,419,015	G	T	SMOC1
T14(S04-4460T)	17	11,827,247	11,827,247	G	T	DNAH9
T18(S05-4996T)	12	81,568,606	81,568,606	C	T	ACSS3
T29(S07-5165T)	1	186,106,707	186,106,707	C	T	HMCN1
T29(S07-5165T)	13	110,853,209	110,853,209	C	T	COL4A1
T29(S07-5165T)	21	33,296,849	33,296,849	G	C	HUNK

- Known Somatic mutatin target INDEL list(Total depth 이상이며, 아미노산 변화를 일으켜, protein function에 영향을 미치는 변이 중 cosmic DB에 등록된 변이): 11개

Sample	Chr	Start	End	Ref	Alt	Gene	cosmic64
T4 (S04-13604T)	17	7,496,121	7,496,121	-	G	FXR2	COSM297879
T4 (S04-13604T)	19	36,224,005	36,224,005	-	C	MLL4	COSM302664
T25 (S06-7551T)	17	20,370,782	20,370,782	-	TC	LGALS9B	COSM302550
T19(S05-9795T)	19	36,224,005	36,224,005	-	C	MLL4	COSM302664
T01(S01-2235T)	17	46,128,958	46,128,958	-	C	NFE2L1	COSM307741
T06(S02-380T)	4	3,015,469	3,015,469	-	A	GRK4	COSM1179565
T06(S02-380T)	22	31,487,798	31,487,798	-	C	SMTN	COSM112101
T14(S04-4460T)	6	158,508,008	158,508,008	-	C	SYNJ2	COSM1162364
T14(S04-4460T)	19	46,393,971	46,393,971	-	G	MYPOP	COSM297312
T18(S05-4996T)	6	146,275,963	146,275,963	-	T	SHPRH	COSM87919
T18(S05-4996T)	17	46,128,958	46,128,958	-	C	NFE2L1	COSM307741

- Unknown Somatic mutation target INDEL list(Total depth 이상이며, 아미노산 변화를 일으켜, protein function에 영향을 미치는 변이): 6099개(리스트는 엑셀표 첨부)

<RNA sequencing 분석결과>

1) RNA QC 결과: T1301R0165 샘플을 제외하고 모두 기준에 미달이었으나, 시퀀싱 진행

Delivery ID	Theragen_ID	BA _Conc. (ng/ul)	RIN value	28s/18s ratio	결과
S07-5165T	T1301R0153	572	4.5	1.6	불합격
S05-15444T	TN1305R2485	1770	5.9	0.9	불합격

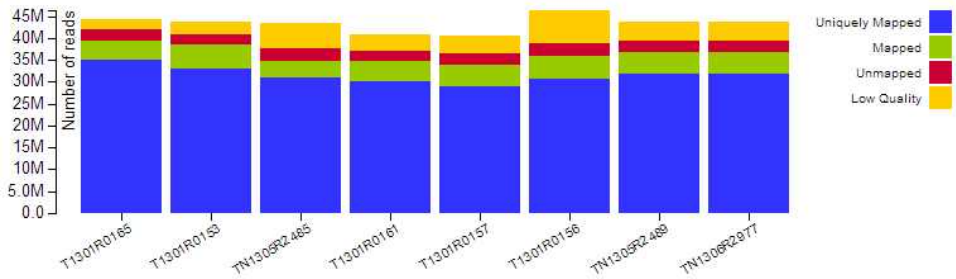
S04-13604T	T1301R0156	1430	5.1	0.4	불합격
S04-63T	T1301R0157	390	5.6	1	불합격
S01-2235T	T1301R0161	615	4.7	2.3	불합격
S02-380T	TN1305R2489	860	4.2	1.2	불합격
S05-4996	TN1306R2977	840	4.7	1.2	불합격
S03-1963T	T1301R0165	282	7.8	1.3	합격

- RNA 시퀀싱을 위한 RNA QC 통과 기준

분석 항목	측정항목	
	RIN value	28s/18s ratio
측정 기준	RIN : > 6.0	28s/18s ratio: > 6.0

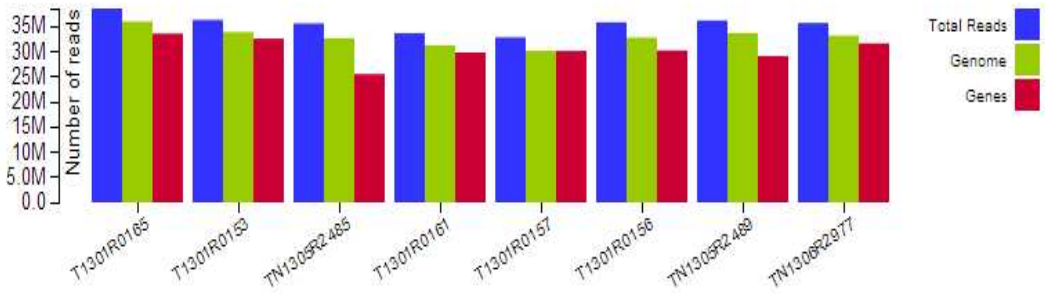
2) RNA 시퀀싱 데이터 기본분석 결과

2-1) 시퀀싱 데이터 생산량 및 필터링 전/후 퀄리티 통계 결과

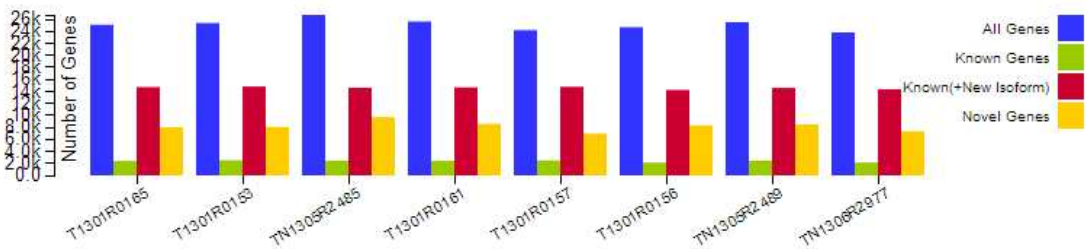


No	Name	Raw	Clean	Mapped	Uniquely Mapped	READ 1 / READ 2	Strand(+) / Strand(-)	Splice
1	T1301R0165	41,075,882	38,628,722 (94.0%)	36,016,833 (93.2%)	34,976,623 (90.5%)	17,535,588 17,441,035	17,492,270 17,484,353	8,360,722 (21.6%)
2	T1301R0153	39,271,578	36,358,108 (92.6%)	33,926,961 (93.3%)	33,003,796 (90.8%)	16,542,451 16,461,345	16,507,870 16,495,926	9,512,567 (26.2%)
3	TN1305R2485	41,419,988	35,649,618 (86.1%)	32,721,494 (91.8%)	30,883,203 (86.6%)	15,469,746 15,413,457	15,455,498 15,427,705	3,629,349 (10.2%)
4	T1301R0161	37,377,794	33,697,484 (90.2%)	31,311,210 (92.9%)	30,236,983 (89.7%)	15,176,974 15,060,009	15,128,204 15,108,779	6,640,710 (19.7%)
5	T1301R0157	36,900,422	32,876,016 (89.1%)	30,194,081 (91.8%)	28,958,642 (88.1%)	14,532,845 14,425,797	14,496,541 14,462,101	8,376,388 (25.5%)
6	T1301R0156	43,376,036	35,856,070 (82.7%)	32,831,444 (91.6%)	30,755,299 (85.8%)	15,425,959 15,329,340	15,385,215 15,370,084	6,039,463 (16.8%)
7	TN1305R2489	40,479,800	36,239,452 (89.5%)	33,711,332 (93.0%)	31,740,686 (87.6%)	15,907,938 15,832,748	15,886,656 15,854,030	7,290,081 (20.1%)
8	TN1306R2977	40,036,596	35,741,662 (89.3%)	33,216,332 (92.9%)	31,898,008 (89.2%)	16,015,417 15,882,591	15,966,069 15,931,939	7,590,973 (21.2%)

2-2) Sequence read alignment 결과



3) Gene expression level 결과



ID	Name	Gene					Transcript			
		SUM	KNOWN	Novel	MOD	UNEXP	SUM	KNOWN	Novel	UNEXP
S0001	T1301R0165	25,108	2,395	8,001	14,712	6,687	230,073	148,276	81,797	0
S0002	T1301R0153	25,386	2,510	8,070	14,806	6,409	186,668	114,219	72,449	43,405
S0003	TN1305R2485	26,750	2,443	9,709	14,598	5,045	188,063	115,099	72,964	42,010
S0004	T1301R0161	25,644	2,404	8,587	14,653	6,151	167,418	94,774	72,644	62,655
S0005	T1301R0157	24,203	2,488	6,955	14,760	7,592	179,430	106,951	72,479	50,643
S0006	T1301R0156	24,694	2,134	8,321	14,239	7,101	182,717	111,304	71,413	47,356
S0007	TN1305R2489	25,535	2,465	8,495	14,575	6,260	170,194	99,427	70,767	59,879
S0008	TN1306R2977	23,827	2,136	7,334	14,357	7,968	167,277	95,875	71,402	62,796

4) Fusion gene 분석결과

4-1) Known fusion gene: 0개

4-2) Unknown fusion gene: 69개

- Unknown fusion gene 중 3개 이상의 샘플에서 detection 된 gene lists: 8개

#gene1	gene2	Detect 된 샘플 수	Detect된 Sample ID
FAM178B	SEMA4C	3	T1301R0165,T1301R0165,T1301R0165
AC007405.7	Y_RNA.703	3	T1301R0153,T1301R0161,TN1305R2485
RP11-504G3.1	UEVLD	4	T1301R0153,T1301R0156,TN1305R2485,TN1305R2489
CIITA	DEXI	5	T1301R0153,T1301R0153,T1301R0157,T1301R0157,T1301R0165
AC138035.1	AL645608.1	3	T1301R0153,T1301R0161,TN1305R2489
CTC-431G16.2	CTC-431G16.3-001	3	T1301R0153,T1301R0153,TN1305R2485
SYT8	TNNI2	3	T1301R0156,T1301R0165,TN1306R2977
NRG1	SLC3A2	4	TN1306R2977,TN1306R2977,TN1306R2977,TN1306R2977

3. 연구결과 고찰 및 결론

1) Clinical utility of targeted next-generation sequencing for predicting the responsiveness to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in never-smokers with lung adenocarcinoma

Discussion & Conclusion

Currently, molecular screening for sensitive EGFR mutations is essential for selecting patients who are best suited for EGFR-TKI therapy. However, the heterogeneous nature of lung cancers can lead to inaccurate molecular classification and therapeutic resistance. Thus, more comprehensive genetic information is necessary to improve the efficacy of EGFR-TKI therapy for lung cancer patients. With advances in NGS technology, the regions of interest can be sequenced at an extremely high depth, thus improving the chances of identifying rare variants. In the present study, we investigated the clinical usefulness of targeted sequencing as a molecular screening test for predicting the responsiveness to EGFR-TKI therapy in NSLA. With targeted sequencing, we were able to identify additional EGFR, PIK3CA, and KRAS mutations that were not identified by Sanger sequencing or PNA-clamp methods. These results may support the use of targeted sequencing with improved sensitivity to identify mutations of low frequency. Furthermore, the additional mutations detected by targeted sequencing were helpful to understand the clinical course of some of the patients in each group.

The cancer-related PIK3CA mutations mostly cluster in exon 9 (E542K and E545K), which codes for the helical domain, or exon 20 (H1047R or L), which codes for the catalytic domain. Mutations in the helical binding domains interfere with p85 binding and enable PI3K activation. The mutations in the catalytic subunit are thought to increase kinase activity [28]. The PIK3CA mutations are associated with a poor response to anti-EGFR therapy in colorectal cancer and anti-HER2 therapy in breast cancer [29, 30]. The PIK3CA mutations are observed in 1-4% of NSCLCs, with a higher prevalence in squamous cell carcinoma. In lung adenocarcinomas, 70% of PIK3CA mutations occur concurrently with EGFR, KRAS, and ALK mutations and are associated with a poor prognosis of lung adenocarcinoma [31]. In our study, four PIK3CA E542K mutations were found, and three occurred concurrently with L858R. Moreover, all the patients with PIK3CA mutations displayed primary resistance to the EGFR-TKI therapy. The higher frequency of these mutations overlapping with EGFR mutations and the negative impact of EGFR-TKI therapy on response rate and PFS supports the implementation of more comprehensive genotyping including PIK3CA for lung adenocarcinoma.

The prevalence of KRAS mutations in lung adenocarcinoma is 10-30%, and a lower frequency is observed in Asians compared with whites [32]. KRAS mutations usually occur in EGFR-wild type tumors. Although EGFR-TKIs are approved for EGFR-wild type patients as a second-line therapy, the presence of KRAS mutations has a negative impact on both the response rate and PFS in patients treated with EGFR-TKI therapy [14]. Thus, the presence of EGFR and KRAS genotyping information may help to determine the appropriate second-line therapy. In our study, all the patients were never-smokers; however, four patients harboring KRAS mutations were found during the screening. Additionally, two more KRAS mutations were found via NGS. All the patients had KRAS codon-12 mutations and exhibited primary resistance to EGFR-TKIs. The median PFS was 1.8 months. Interestingly, one additional KRAS Q61P mutation was detected in a patient in group 4, who exhibited SD with 2nd-lineuseoferlotinibfor5.5months(caseT4-2).

The most frequent and clinically significant EGFR mutations that predict higher response rates and a longer PFS to EGFR-TKI therapy are 19DEL and L858R [10]. Thus, we selected NSLAs harboring these two classic mutations. During the screening, we confirmed these mutations using two methods, namely, Sanger sequencing and PNA-clamp based PCR. With NGS, we also confirmed all known EGFR 19DEL or L858R mutations except one L858R in group 2, which suggests that the sample provided for target sequencing might not have contained tumor cells harboring the L858R mutation. For this study, we conducted multiple tests on each sample, which might have caused variability in the samples.

In the present investigation, we demonstrated that targeted deep sequencing using FFPE tumor samples was feasible and provided more accurate genotype information through multiplex testing with an improved sensitivity. Thus, the clinical application of NGS technology

as a molecular screening test might favorably alter the clinical outcomes of lung cancer patients. However, the targeted sequencing did not explain the biological features of all the patients. Although targeted sequencing can yield a much higher coverage of genomic regions of interest, it is limited when identifying novel genetic alterations supporting the biological phenotype of an individual cancer. The current challenges in the successful transition of NGS to the clinic include the complexities of the data analysis and interpretation [33]. In addition, the sequencing errors remain significant. Such errors cannot be distinguished from genetic variance and may lead to false-positive results [34]. Thus, the development of physician-friendly computational analysis tools and training programs are critically needed. Further investigation of the correlation between genomic data generated by NGS technology and the biological behavior of individual cancers would facilitate the implementation of NGS in cancer clinics.

2) Whole-genome analysis of a patient with early stage small cell lung cancer

Discussion

SCLC cells commonly contain somatic mutations in tumor suppressor genes, including *TP53*, *RB1*, *p16*, *RASSF1A*, *CREBBP* and *FHIT* genes (Sanchez-Cespedes 2009). In this study, we found variations in *TP53*, *RB1* and *CREBBP* genes, but not in *p16*, *RASSF1A*, and *FHIT* genes.

In details, we found a somatic mutation in *TP53* gene, which resulted in H193R amino acid change (Figure 2A). Together with the identification of loss of heterozygosity (LOH) in chromosomal regions covering *TP53* gene, the function of TP53 may defected influencing in carcinogenesis of this patient. Especially, a *TP53* mutation was resulted in H193R amino acid change, which resides in the DNA binding domain of TP53. It was reported that TP53 H193R mutant protein weakly binds to and transactivates *p21* gene, which results in abnormal regulation of DNA synthesis, cell cycle and apoptosis (el-Deiry, et al. 1993;Harper, et al. 1993;Noda, et al. 1994). Loss of TP53 and RB1 function possibly triggers cell-type specific carcinogenesis in SCLC (Sozzi, et al. 1995;Sutherland, et al. 2011). Together with the identification of loss of heterozygosity (LOH) in chromosomal regions covering *TP53* and *RB1* genes, possible loss of TP53 function by H193R mutation may involve in early steps of carcinogenesis of SCLC. A heterozygous mutation in *CREBBP* gene was also found. This mutation changed Tyrosine at 1395 amino acid position to Cysteine (Y1395C), which could affect the function of CREBBP transcription factor by interupting the HAT domain (Figure 2B). This mutation could be a functional mutation affecting the role of CREBBP, a transcriptional cofactor that acetylates proteins including TP53 (Karamouzis, et al. 2007). The dominant mutations of *CREBBP* gene were known to cause Rubinstein-Taybi syndrome and increased sensitivity to tumorigenesis (Janknecht 2002;Miller and Rubinstein 1995). Also, gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300 (Yao, et al. 1998). Truncated CREBBP protein leads to classical Rubinstein-Taybi syndrome phenotypes in mice: implications for a dominant-negative mechanism (Oike, et al. 1999). In addition, the cancer genome of NCI-H209 cell line harbors the fusion of the first two exons in *CREBBP* gene to *BTBD12* gene and homozygous *CREBBP* deletion was observed in two SCLC cell lines (Kishimoto, et al. 2005;Pleasance, et al. 2010). As recent SCLC genome and transcriptome study reported that somatic mutations in *CREBBP* were clustered around the HAT domain encoding sequence, we also found a nsSNV at HAT domain as heterozygous mutation in *CREBBP* (Peifer, et al. 2012). The coexistance of *TP53* and *CREBBP* mutations suggests that somatic mutations in those genes may affect multiple cancer-related pathways, since TP53 and CREBBP have overlapping roles in several different pathways.

Many other nsSNVs were found in number of genes, which are not well functionally analyzed in cancer. In addition to SNVs in hot target genes, we found two novel recurrent nsSNVs. For example, we identified a somatic nsSNV in *C6orf103* (Calpain-7 like protein) gene which was also reported by whole genome sequencing of a SCLC cell line (Pleasance, et al. 2010). This mutation changed Valine at 457 amino acid position to Methionine (Figure 2C). A somatic nsSNV in *C6orf103* gene was recurrent as reported by whole genome sequencing of a SCLC cell line (Pleasance, et al. 2010). C6orf103 (Calpain-7 like protein) is a member of calpain protein family which is a conserved family of cysteine proteinases that catalyze the controlled proteolysis of many specific substrates. Calpain activity is implicated in several

fundamental physiological processes, including cytoskeletal remodelling, cellular signaling, apoptosis and cell survival (Suzuki, et al. 2004). Calpain expression is increased in tumor cells (Lee, et al. 2007;Lee, et al. 2008). Calpains are important for proteolysis of numerous substrates in tumor pathogenesis, such as inhibitors of nuclear factor-κB (IκB), focal adhesion kinases and talins and cMYC (Chan, et al. 2010;Franco, et al. 2004;Han, et al. 1999;Small, et al. 2002). Therefore, C6orf103 protein could be interpreted as a tumor suppressor. A novel nsSNV was also found in *SLC5A4* gene, which encodes a family member of low affinity sodium-glucose cotransporters. This mutation changed Phenylalanine at 17 amino acid position to Histidine (Figure 2D). One more nsSNV changing Glycine at 43 amino acid position to Glutamic acid was observed by analyzing the exonic regions of *SLC5A4* gene in 23 additional SCLC samples, showing together 8.3% (2/24) frequency of nsSNV in *SLC5A4* gene. SLC5A4 is a unique member of low affinity sodium-glucose cotransporters (SLC5 family). Among glucose transporters, SLC5A4 can not transport sugar, but act as a sugar sensor (Diez-Sampedro, et al. 2003). Therefore, damaging changes in SLC5A4 may result in the loss of glucose sensing which is essential to the control of cell growth and proliferation. As cancer cells require energy to feed uncontrolled proliferation, they normally over-express the glucose transporter family members. Therefore, loss of function variations in glucose sensor may contribute to lack of negative feedback regulation of glucose transporter gene expression in cancer cells. Our finding could give a clue on a novel cancer pathway for SCLC tumorigenesis.

Mutational profiling in cancers has revealed the somatic variations classifying the different stages and types of cancer, which play important roles in the development of cancer. Recently, pathway-level analysis of cancer mutations has provided physicians with the insight into the molecular mechanism of cancer progression and a clue to the treatment of cancer patients. To analyze the influence of somatic variations in the tumorigenesis at the pathway-level, SNVs in components of specific signaling pathways were defined through functionally annotated KEGG database (Kanehisa, et al. 2012). The 116 novel, somatic nsSNVs and 21 novel, somatic indels were searched against KEGG pathway maps, and significantly affected pathways ($p < 0.01$) were shown in table 3. In this study, we did not find well known oncogenic variations in SCLC like *c-MYC* and *SOX2* gene amplification (Peifer, et al. 2012;Rudin, et al. 2012). Interestingly, only *TP53* gene has a somatic mutation among previously known 64 components of small cell lung cancer pathway in KEGG database. Together with lack of the mutation pattern frequently found in heavy smokers, the mutation landscape of this case is different from that of other SCLCs, mostly reported from advanced stage SCLC.

Total 15 KEGG pathways including Notch and Wnt signaling pathways showed statistically significant association with somatic mutations of SCLC (p-value < 0.01. Several genes have overlapping roles in different pathways. For example, CREBBP involved at least in 4 pathways, including, Long-term potentiation, Notch signaling pathway, Wnt signaling pathway and melanogenesis. This indicates that somatic mutations in those genes may affect multiple cancer-related pathways. As the pathway-level evidence, Notch and Wnt signaling components had more possibly damaging amino acid changes. Notch and Wnt signal pathways play essential roles in lung development and also have a crucial role in tumorigenesis when reactivated in adult tissues through somatic mutations (Daniel, et al. 2006). Loss- or gain-of-function mutations damaging these two signaling pathways in this study may abnormally modulate Notch and Wnt signals and results in dramatic proliferative effects in early-stage cancer cells. Our finding supports the synergistic effects between Notch and Wnt signals favorable for the further accumulation of oncogenic mutations. It can be hypothesized that aberrant Notch and Wnt activation results in hyperplastic conditions representing a preneoplastic state of lung cells, in which the chances of secondary mutagenic events that can drive full-fledged malignancy are increased. Therefore, our results suggests that, activation of oncogenic Notch and Wnt pathways may be important for the early events of small cell lung cancer development.

Conclusions

As SCLC is characterized by rapid proliferation and early dissemination, most cases present with advanced stage diseases, which hamper accessing early stage SCLC. In this study, we could analyzed the genomic profile of early stage SCLC by WGS of matched normal-tumor samples from a patient with stage IA SCLC underwent curative resection. We identified 43,339 novel somatic SNVs and found their base substitution pattern different from heavy smoker's. Many somatic SNVs and copy number variations has been already reported as either oncogenes or tumor suppressors in COSMIC database, including *TP53* and *RB1* genes. This cancer genome had few mutated genes in SCLC pathway, but showed statistically meaningful

genetic changes in the Notch and WNT signaling pathways. Taken together, comprehensive analysis of whole genome from a patient with early stage SCLC provided a distinct genomic profile, which may give insights into the molecular classification of early stage SCLC patients for personalized diagnostics and treatment.

3) WES & RNA seq. 분석 결과 고찰

A. EGFR driver mutation 결과

- EGFR driver mutation(L858R, exon19 deletion) 분석결과, 기존에 진행하였던 cancer panel 데이터와 일치하였으며, 엑솜시퀀싱 데이터의 allele frequency 및 RNA 시퀀싱 데이터의 발현량 정도(RPKM value, reads per kilobase of exon model per milion mapped reads, 즉, 전체 read 수를 백만 개라고 가정했을 때 exon 1Kb 당 붙은 reads 수)는 아래 표와 같다.

	Delivery_ID	Mutation	WES	RNA sequencing (RPKM)
Group 1	T4 (S04-13604T)	L858R	47.1%	78.9
	T10 (S04-63T)	Exon19 deletion	5.5%	11.1
	T12 (S04-3413T)	Exon19 deletion	13.8%	
	T25 (S06-7551T)	L858R	47.1%	
	T28 (S07-2404T)	L858R	46.09%	
Group 2	S05-15444T	L858R	17.4%	11.4
	S05-9795T	L858R	47.1%	

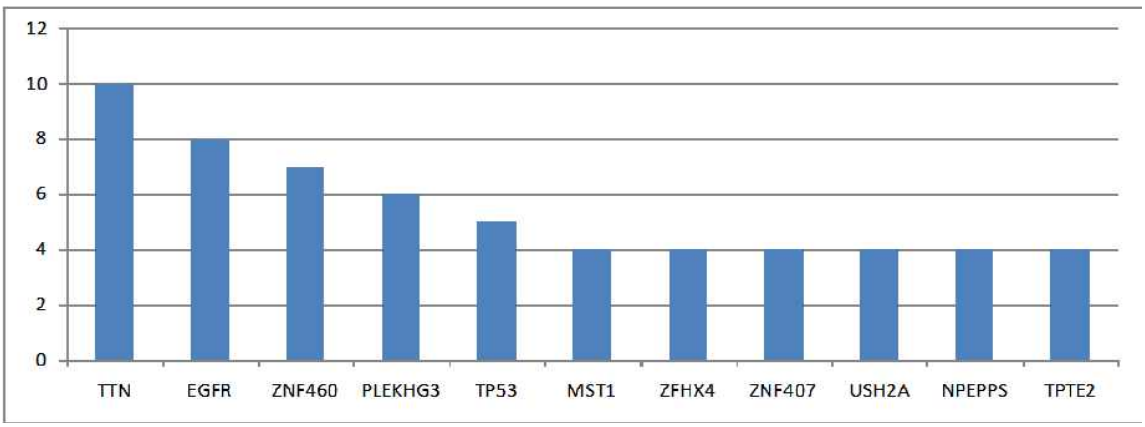
B. WES vs RNA-seq 비교결과

	S07-5165T		**S05-15444T**		**S04-13604T**		**S04-63T**	
Gene	WES coverage (alt/total)	RNA_seq (RPKM)	WES coverage (alt/total)	RNA_seq (RPKM)	WES coverage (alt/total)	RNA_seq (RPKM)	WES coverage (alt/total)	RNA_seq (RPKM)
NCF2	ND	22.6296	17/138	15.5984	ND	39.3875	ND	17.0417
SPTBN1	ND	69.7535	ND	18.9931	22/138	81.1851	ND	52.7373
FN1	ND	674.158	14/148	187.272	ND	162.41	ND	1106.27
SETD2	ND	23.8571	ND	32.0808	ND	52.1072	22/184	30.2227
DHX15	ND	81.2462	ND	35.0649	ND	104.016	12/119	51.3739
KIAA1109	30/196	22.394	ND	5.33064	ND	38.432	ND	18.688
UFSP2	20/148	15.653	ND	9.24584	ND	8.1577	ND	14.8197
PCDHGA1	ND	28.3384	ND	15.1921	ND	27.3471	22/121	30.3584
PCDHGA3	ND	28.3384	ND	15.1921	ND	27.3471	31/131	30.3584
RANBP17	22/196	16.0015	ND	13.908	ND	16.9464	ND	3.2272
HLA-B	ND	2486.52	ND	2206.17	18/110	1136.3	ND	2829.93

HLA-DRB1	ND	3416.2	ND	1057.37	44/150	24.4508	ND	1698.68
EGFR	ND	22.7948	38/180	11.4586	531/621	78.9859	ND	11.05
COL1A2	ND	155.279	ND	96.6807	22/172	46.4193	ND	558.634
STK17A	ND	35.1097	ND	18.6579	ND	32.5953	14/107	25.2989
EGFR	29/162	22.7948	38/180	11.4586	531/621	78.9859	ND	11.05
GNAT3	23/213	15.9232	ND	1.71724	ND	0.202836	ND	6.89509
COL1A2	17/132	155.279	ND	96.6807	ND	46.4193	ND	558.634
PCMTD1	59/408	12.6918	ND	8.35541	ND	8.50286	42/337	8.12855
FAM160A2	11/100	48.9368	ND	57.4141	ND	132.163	ND	26.3997
HNRNPUL2	15/136	107.454	ND	396.146	ND	222.03	ND	118.697
DDX11	ND	31.0563	ND	38.3341	ND	101.875	21/128	25.1816
SLC38A2	26/162	71.3548	ND	25.2668	ND	26.6553	ND	43.5536
RB1	24/110	12.0436	ND	3.67348	ND	1.9226	ND	14.9219
SUPT16H	ND	62.405	ND	44.7014	15/136	60.37	ND	47.7494
HERC2	11/95	29.1214	10/88	9.5258	ND	72.205	17/143	20.4433
SRRM2	ND	1342.96	ND	1589.43	14/109	665.338	ND	368.811
CHST5	ND	23.181	47/187	56.5651	ND	13.5919	ND	41.1349
EID2	32/107	11.5888	ND	4.71501	ND	6.32099	ND	10.073
BRWD1	19/128	39.7074	ND	23.4968	ND	31.9403	ND	35.8307
RBM10	ND	28.0544	25/126	62.1003	ND	141.733	ND	23.1762
PORCN	15/100	81.0656	ND	154.107	ND	150.955	ND	63.9142

	S01-2235T		**S02-380T**		**S05-4996T**		**S03-1963T**	
Gene	WES coverage (alt/total)	RNA_seq (RPKM)	WES coverage (alt/total)	RNA_seq (RPKM)	WES coverage (alt/total)	RNA_seq (RPKM)	WES coverage (alt/total)	RNA_seq (RPKM)
GORAB	ND	28.2624	ND	1.22585	ND	12.0763	67/315	24.6136
HLA-B	ND	1231.21	13/70	3240.23	ND	2385.33	ND	468.669
HLA-DRB1	ND	535.403	ND	2225.93	ND	61.2673	28/222	241.028
PCMTD1	ND	18.9769	ND	2.56204	ND	8.30279	35/250	22.4138
EPPK1	ND	2.92755	ND	0.828172	ND	3.32532	98/591	17.005
HERC2	29/108	30.7688	19/77	11.8599	23/93	13.4319	51/157	35.6723
MYO18A	ND	86.0242	ND	54.0709	ND	40.6899	55/147	29.6519
SRSF6	ND	146.647	ND	59.5452	ND	77.9256	64/160	94.1402

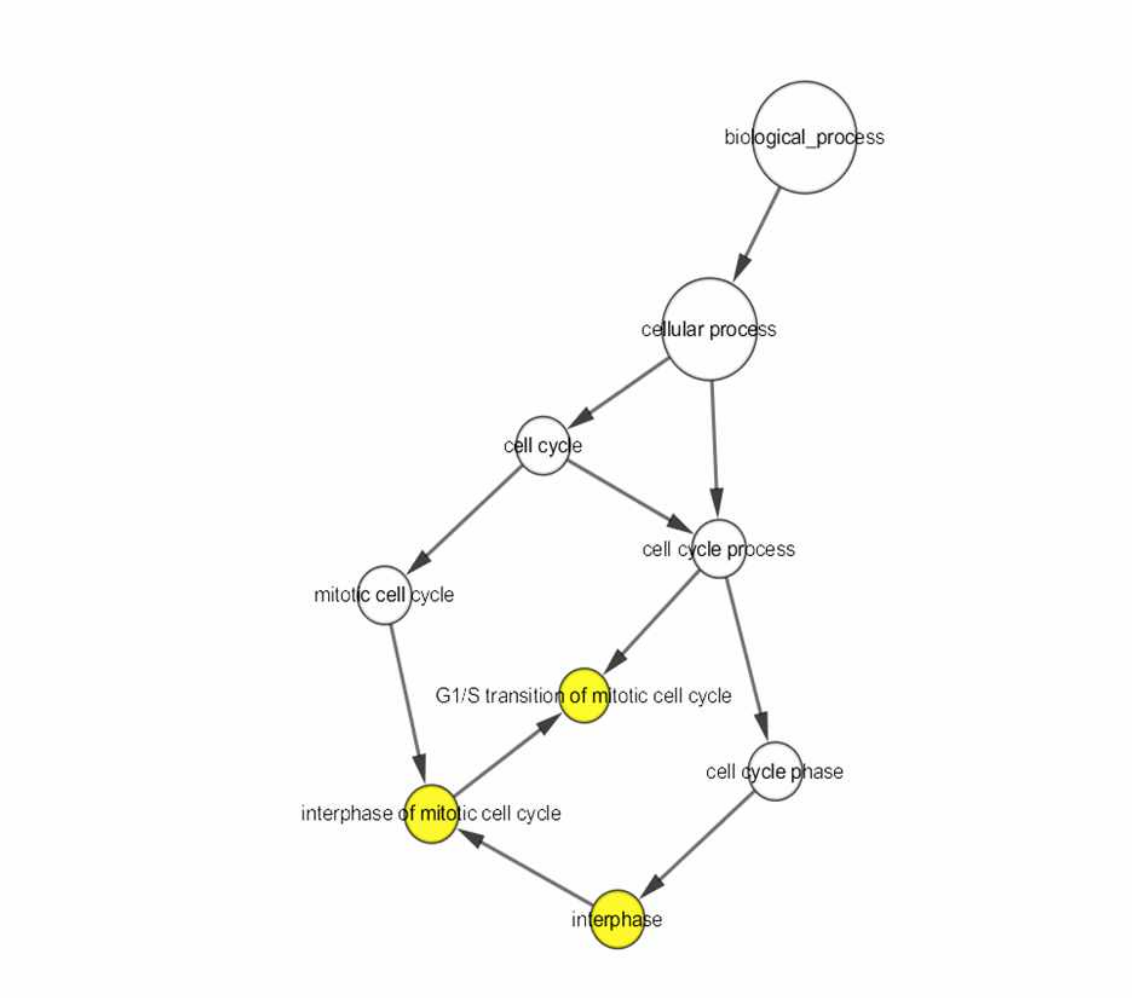
- 다수의 샘플에서 검출된 상위 유전자 11개 리스트



C. Gene Ontology 분석결과

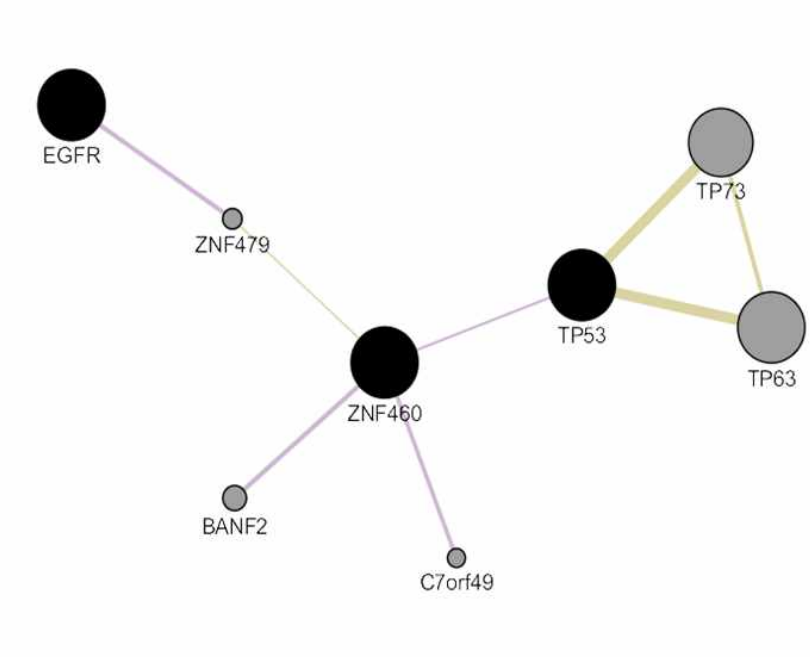
- 엑솜 분석 및 RNA 분석 결과의 타겟 유전자들을 이용하여 Biological을 분석을 실시하였다. (BiNGO 툴 이용)

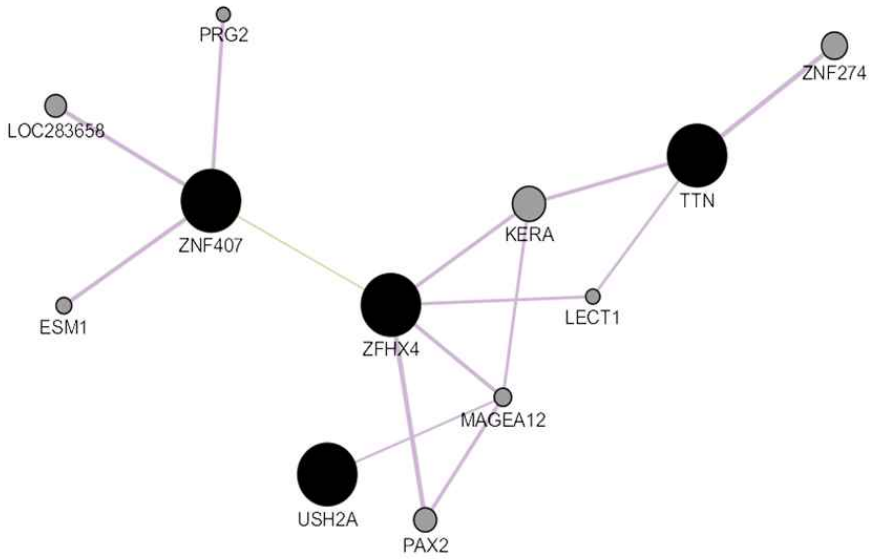
- Biological 분석 결과, cell-cycle과 관계있는 Go term이 유의있게 나타났다. 이는 최근에 발표된 논문과도 어느정도 일치하는 결과를 보였다. (Mutational landscape and significance across 12 major cancer types, Nature 502, 333–339)

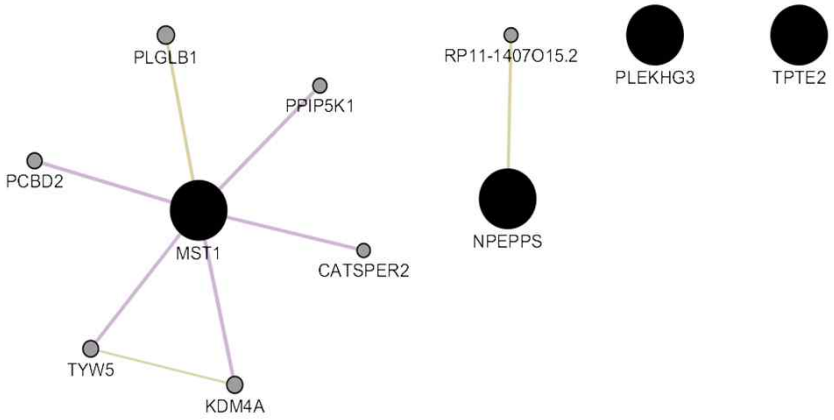


- 각 샘플의 somatic mutation 중, 각 유전자마다 변이의 수를 count 하였다. 그 중 4 샘플 이상에서 변이가 일어난 유전자를 이용하여 gene interaction 분석을 진행하였다(GeneMania 툴 이용). 그 결과, 관계있는 유전자별로 연결이 되었다.

- EGFR, ZNF460, TP53은 폐암과 관계있는 것으로 알려져 있다. 이를 이용하여 주위에 interaction하는 다른 유전자들의 심도있는 조사가 필요할 것으로 사료된다.







4. 연구성과 및 목표달성도

(1) 연구성과

가. 국내 및 국제 전문학술지 논문 게재 및 신청

논문명	저자 (저자구분[1])	저널명(I.F.)	Year; Vol(No):Page	구분[2]	지원과제번호[3]
Whole-genome analysis of a patient with early stage small cell lung cancer	Han JY et al. (교신&제1저자)	Genes Chromosome and Cancer (3.546)	under review	국외 SCI	1110100
Comparison of targeted next-generation sequencing with conventional sequencing for predicting the responsiveness to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in never-smokers with lung adenocarcinoma	Han JY et al. (교신&제1저자)	Cancer (5.201)	under review	국외 SCI	1110100
A genome-wide association study of survival in small-cell lung cancer patients treated with irinotecan plus cisplatin chemotherapy	Han JY et al. (교신&제1저자)	Pharmacogenomics J (5.134)	2013 (Epub ahead of print)	국외 SCIE	없음
Impact of EGFR tyrosine kinase inhibitors versus chemotherapy on the development of leptomeningeal metastasis in never smokers with advanced adenocarcinoma of the	Lee Y, Han JY, Kim HT, Yun T, Lee GK, Kim	J Neuro-oncology (3.115)	2013 (Epub ahead of print)	국외 SCI	없음

lung.	HY, Lee JS. (공동저자)				
Analysis of treatment outcomes of intraventricular chemotherapy in 105 patients for leptomeningeal carcinomatosis from non-small-cell lung cancer.	Gwak HS, Joo J, Kim S, Yoo H, Shin SH, Han JY, Kim HT:Lee JS, Lee SH. (공동저자)	J Thorac Oncol (4.473)	2013;8:599-605	국외 SCI	없음
Prognostic Value of Gross Tumor Volume for Definitive Radiation Therapy in Patients With Locoregionally Recurrent Non-Small-Cell Lung Cancer After Surgical Resection	Lee NK, Moon SH, Kim TH, Han JY, Yun T, Kim HT, Lee HS, Kim MS, Lee JM, Cho KH, Lee JS (공동저자)	Clin Lung Cancer (2.038)	2013;14:399-406	국외 SCIE	없음
First-SIGNAL: First-Line Single-Agent Iressa Versus Gemcitabine and Cisplatin Trial in Never-Smokers With Adenocarcinoma of the Lung.	Han JY et al. (제1저자)	J Clin Oncol (18.038)	2012;30 :1122-8	국외 SCI	없음
Prognostic implications of genetic variants in advanced non-small cell lung cancer: a genome-wide association study	Lee YJ, Yoon KA, Joo JN, Lee DH, Bae K, Han JY, Lee JS (공동저자)	Carcinogenesis (5.632)	2012;34 :307-13	국외 SCI	없음
A phase II study of sunitinib in patients with relapsed or refractory small cell lung cancer	Han JY et al. (교신&제1저자)	Lung Cancer	2012;79 :137-42	국외 SCI	없음
A randomized phase II study of irinotecan plus cisplatin versus irinotecan plus capecitabine with or without isosorbide-5-mononitrate in advanced non-small cell lung cancer	Han JY et al. (제1저자)	Ann Oncol	2012;23 :2925-30	국외 SCI	없음
A genome-wide association study for irinotecan-related severe toxicities in patients with advanced non-small cell lung cancer	Han JY et al. (교신&제1저자)	Pharmacogenomics J	2012	국외 SCIE	없음
Comparison of clinical outcome of patients with non-small-cell lung cancer harbouring epidermal growth factor receptor exon 19 or exon 21 mutations.	Han JY et al. (교신저자)	J Clin Pathol	2011;64 947-52:	국외 SCI	없음
Association between plasma hepatocyte growth factor and gefitinib resistance in patients with advanced non-small cell lung cancer	Han JY et al. (교신&제1저자)	Lung Cancer	2011;74 :293-9	국외 SCI	없음
Simvastatin enhances irinotecan-induced apoptosis in human non-small cell lung cancer cells by inhibition of proteasome	Han JY et al. (교신저자)	Invest New Drugs.	2011;29 :883-90	국외 SCI	없음

activity.					
Circulating cell-free DNA in plasma of never smokers with advanced lung adenocarcinoma receiving gefitinib or standard chemotherapy as first-line therapy.	Yoon KA, Lee YJ, Han JY, Kim HT, Lee GK, Kim HY, Lee JS (공동저자)	Clin Cancer Res	2011;17 :5179-87	국외 SCI	없음
DNA repair gene polymorphisms and benefit from gefitinib in never-smokers with lung adenocarcinoma	Han JY et al. (교신&제1저자)	Cancer	2011;117 :3201-8	국외 SCI	없음
A phase II study of irinotecan, cisplatin and simvastatin for untreated extensive-disease small-cell lung cancer	Han JY et al. (교신&제1저자)	Cancer	2011;117 :2178-85	국외 SCI	없음
A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population.	Yoon KA, Park JH, Han J, Park S, Lee GK, Han JY, Zo JI, Kim J, Lee JE, Takahashi A, Kubo M, Nakamura Y, Lee JS. (공동저자)	Hum Mol Genet	2010;19 :4948-54	국외 SCI	없음
Association of SUMO-1 and UBC9 genotypes with tumor response in non-small cell lung cancer treated with irinotecan-based chemotherapy	Han JY et al. (교신&제1저자)	Pharmacogenomics J.	2010;10 :86-93	국외 SCIE	없음
Lovastatin overcomes gefitinib resistance in human non-small cell lung cancer cells with K-Ras mutations.	Han JY et al. (교신저자)	Invest New Drugs	2010;28 :791-9	국외 SCI	없음

1) 저자구분 : 교신, 제1, 공동
2) 구분 : 국내, 국내 SCI, 국내 SCIE, 국외, 국외SCI, 국외SCIE 등
3) 지원과제번호(Acknowledgement)
- 과제번호를 연차 표시(-1, -2, -3 등)를 생략하고 7자리로 기재하고, 과제와 관련성은 있으나 불가피하게 Acknowledgement가 누락된 경우에는 '없음'으로 기재

나. 국내 및 국제 학술대회 논문 발표

논문명	저자	학술대회명	지역[1)]	지원과제번호
Identification of novel oncogenes in a stage I small cell lung cancer through whole-genome sequencing	Han JY et al.	AACR	국외	1110100
Clinical application of targeted deep sequencing as a molecular screening for epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in never-smoking lung adenocarcinoma	Han JY et al.	World Conference of Lung Cancer	국외	1110100

1) 지역 : 국내, 국외

다. 산업재산권

구분[1)]	특허명	출원인	출원국	출원번호

1) 구분 : 발명특허, 실용신안, 의장등록 등

라. 저 서

저서명	저자	발행기관(발행국, 도시)	쪽수	Chapter 제목, 쪽수 (공저일 경우)

마. 연구성과의 정부정책 기여

보고서명	정부정책	기여내용

바. 기타연구성과

(2) 목표달성도

가. 연구목표의 달성도

○ 사업목표에 대한 달성내용 및 관련분야 기술발전에의 공헌도 등을 기술
○ 달성도(%)는 연차별목표대비 당해연도 달성도 및 최종목표대비 당해연도까지의 누적 달성도를 반드시 기입

최종목표	연차별목표		달성내용	달성도(%)	
				연차	최종
- 최신의 암 유전체 분석 기법인 NGS를 이용하여 폐암의 암 유전체 연구를 수행. - 새로운 유전자 변형 등을 발굴하여 폐암	1차년도	비흡연 폐선암 환자에서 EGFR-TKI 내성에 관여하는 새로운 유전체 변형을 발굴, EGFR-TKI 일차적 내성기전 제시	-분석 대상 환자 선정 EGFR, KRAS, PIK3CA, BRAF 변이 분석 및 EML4-ALKfusion gene 분석을 통하여 NGS 분석을 위한 최종 대상 샘플 선정 및 소그룹 분류	100	100
	2차년도	EGFR-TKI 일차내성 관련 기전 새로운 기전 분석	IonTorrent PGM을 이용한 targeted sequencing 으로 내성 기	100	100

의 진단 및 치료에 기여 할 수 있는 원천 지식을 제공하고 암 유전체 연구를 위한 기반 확보			전 확립		
		소세포폐암의 발생기전 분석	Whole genome sequencing 을 l용하여 1례의 초시 소세포폐암의 전장유전자변이 분석 및 새로운 유전자 변이 발굴		
	3차년도	비흡연 폐선암 환자에서 EGFR-TKI 내성에 관여하는 새로운 유전체 변형을 발굴하여 EGFR-TKI 일차적 내성기전 제시	Sanger sequencing, PNA-clamp PCR 분석 등의 고식적인 분석법으로 발견되지 못한 내성 관련 유전자 변이 발견 고식적 분석법과 NGS 기반의 유전자변이 발견 정도 비교 및 임상적 유용성 평가	90	90
		비흡연 폐선암의 새로운 driver oncogene 발굴	Targetd sequencing 으로 내성 기전 발견하지 못한 샘플을 대상으로 새로운 유전자 변이 발굴을 위하여 whole exone seq. & RNA se. 실시함.		

나. 평가의 착안점에 따른 목표달성도에 대한 자체평가

평가의 착안점	자 체 평 가
현재 임상적으로 폐선암환자에서 EGFR-TKI의 반응을 예측할 수 있는 예측인자는 EGFR 유전자 변이임. 그러나 EGFR유전자 변이에도 불구하고 약 10%의 환자에서 일차 내성을 보임. 치료 전 일차내성의 가능성을 평가할 수 있는 방버의 개발이 시급함.	- NGS 기반의 Targeted sequencing의 임상적 사용 가능성 확립. - NGS 기반의 분석법을 통하여 EGFR 일차내성 기전 발굴함. -향후 NGS기반의 분석법이 기존s의 고식적 분석방법에 비하여 보다 정확하게 임상적 예후를 대변할 가능성 확립함.
소세포폐암의 발생기전 분석	- 아주 드문 제1기의 소세포폐암의 전장유전자변이분석을 수행하여 소세포폐암의 유전자 변이 특성 및 새로운 유전자변이를 발굴함. - 증례가 1례라는 제한점 있음. -예후가 불량하고 현재까지 뚜렷한 치료의 발전을 이루지 못하고 있는 질환의 새로운 치료 가능성을 제시함.
현재까지 치료 표적이 알려지지 않은 비흡연 폐선암의 새로운 driver oncogene 발굴	Targeted seq.으로 유전자 변이를 발굴하지 못한 샘플을 대상으로 Whole Exone Sequencing 과 RNA sequencing을 실시하고 결과 분석 중임. 새로운 driver oncogene 발견 시 폐선암의 새로운 치료제 개발에 기여 가능함.

5. 연구결과의 활용계획

(1) 연구종료 2년후 예상 연구성과

○ 연구종료 2년후까지 연구사업 결과로 발생할 것으로 예상되는 성과, 즉 학술지 게재, 산업재산권 등을 단계별로 다음의 양식에 의거하여 작성함. 학술지 게재는 게재 예상 학술지 명과 Impact Factor 등을 기재함

○ 연구사업의 내용이 논문이나 산업재산권과 연결되기 힘든 과제의 경우, 자유 형식으로 예상연구성과 및 활용정도를 기재하되 최대한 계량화할 것

예) DB 몇 건 구축완료. OOO 시스템 구축 및 OO사업 완료

구 분	건 수	비 고
학술지 논문 게재	1	5
산업재산권 등록		특허 등록 예상 국가, 예상 특허명 등
기 타		

(2) 연구성과의 활용계획

○ 연구성과물의 활용분야 및 활용방법, 활용범위 등을 구체적(특히 시간적 구체성, 예를 들어 몇 년 안에 치료기술 실용화 등)으로 기술하되, 참여기업이 포함되어 있는 과제의 경우 기업과 연계한 활용방안에 대해서도 기술함

○ 추가 후속연구의 필요성에 대해서도 간략하게 기술함

- **EGFR 변이를 동반하고도 EGFR-TKI치료에 일차 내성이 발생할 수 있는 경우를 미리 선별하는 검사방법으로 활용하여 보다 개선된 폐암환자의 맞춤치료전략으로 사용할 예정임.**
- **임상적으로 NGS 기반의 유전자 변이 분석 프로토콜을 확립함**
- **임상유전체분석을 이용한 폐암의 맞춤치료 실현을 위하여 진행성 폐암환자 전반의 NGS 기반 유전자 분석 연구의 후속 조치가 필요함.**

6. 참고문헌

Clinical utility of targeted next-generation sequencing for predicting the responsiveness to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in never-smokers with lung adenocarcinoma

1. Shepherd FA, Rodrigues Pereira J, Ciuleanu T et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005; 353: 123-132.
2. Maemondo M, Inoue A, Kobayashi K et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med 2010; 362: 2380-2388.
3. Rosell R, Carcereny E, Gervais R et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012; 13: 239-246.
4. Zhou C, Wu YL, Chen G et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011; 12: 735-742.
5. Mitsudomi T, Morita S, Yatabe Y et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol 2010; 11: 121-128.
6. Keedy VL, Temin S, Somerfield MR et al. American Society of Clinical Oncology provisional clinical opinion: epidermal growth factor receptor (EGFR) Mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. J Clin Oncol 2011; 29: 2121-2127.
7. Lindeman NI, Cagle PT, Beasley MB et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. Arch Pathol Lab Med 2013; 137: 828-860.
8. Lindeman NI, Cagle PT, Beasley MB et al. Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors: Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Thorac Oncol 2013; 8: 823-859.
9. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. Nat Rev Cancer 2007; 7: 778-790.
10. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 2007; 7: 169-181.
11. Gainor JF, Shaw AT. Novel Targets in Non-Small Cell Lung Cancer: ROS1 and RET Fusions. Oncologist 2013; 18: 865-875.
12. Takeuchi K, Soda M, Togashi Y et al. RET, ROS1 and ALK fusions in lung cancer. Nat Med 2012; 18: 378-381.
13. Paik PK, Arcila ME, Fara M et al. Clinical characteristics of patients with lung adenocarcinomas harboring BRAF mutations. J Clin Oncol 2011; 29: 2046-2051.
14. Linardou H, Dahabreh IJ, Bafaloukos D et al. Somatic EGFR mutations and efficacy of tyrosine kinase inhibitors in NSCLC. Nat Rev Clin Oncol 2009; 6: 352-366.
15. Oxnard GR, Binder A, Janne PA. New targetable oncogenes in non-small-cell lung cancer. J Clin Oncol 2013; 31: 1097-1104.
16. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. Lancet Oncol 2011; 12: 175-180.
17. Ludovini V, Bianconi F, Pistola L et al. Phosphoinositide-3-kinase catalytic alpha and KRAS mutations are important predictors of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in patients with advanced non-small cell lung cancer. J Thorac Oncol 2011; 6: 707-715.
18. MacConaill LE. Existing and emerging technologies for tumor genomic profiling. J Clin Oncol 2013; 31: 1815-1824.
19. Desai AN, Jere A. Next-generation sequencing: ready for the clinics? Clin Genet 2012; 81: 503-510.
20. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 2010; 11: 685-696.
21. Paez JG, Janne PA, Lee JC et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004; 304: 1497-1500.
22. Kim HJ, Lee KY, Kim YC et al. Detection and comparison of peptide nucleic

acid-mediated real-time polymerase chain reaction clamping and direct gene sequencing for epidermal growth factor receptor mutations in patients with non-small cell lung cancer. Lung Cancer 2012; 75: 321-325.
23. Araki T, Shimizu K, Nakamura K et al. Usefulness of peptide nucleic acid (PNA)-clamp smart amplification process version 2 (SmartAmp2) for clinical diagnosis of KRAS codon 12 mutations in lung adenocarcinoma: comparison of PNA-clamp SmartAmp2 and PCR-related methods. J Mol Diagn 2010; 12: 118-124.
24. Jeong D, Jeong Y, Park JH et al. BRAF (V600E) mutation analysis in papillary thyroid carcinomas by peptide nucleic acid clamp real-time PCR. Ann Surg Oncol 2013; 20: 759-766.
25. Yi ES, Chung JH, Kulig K, Kerr KM. Detection of anaplastic lymphoma kinase (ALK) gene rearrangement in non-small cell lung cancer and related issues in ALK inhibitor therapy: a literature review. Mol Diagn Ther 2012; 16: 143-150.
26. Suehara Y, Arcila M, Wang L et al. Identification of KIF5B-RET and GOPC-ROS1 fusions in lung adenocarcinomas through a comprehensive mRNA-based screen for tyrosine kinase fusions. Clin Cancer Res 2012; 18: 6599-6608.
27. Singh RR, Patel KP, Routbort MJ et al. Clinical Validation of a Next-Generation Sequencing Screen for Mutational Hotspots in 46 Cancer-Related Genes. J Mol Diagn 2013.
28. Kang S, Bader AG, Vogt PK. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. Proc Natl Acad Sci U S A 2005; 102: 802-807.
29. Ikenoue T, Kanai F, Hikiba Y et al. Functional analysis of PIK3CA gene mutations in human colorectal cancer. Cancer Res 2005; 65: 4562-4567.
30. Chandarlapaty S, Sakr RA, Giri D et al. Frequent mutational activation of the PI3K-AKT pathway in trastuzumab-resistant breast cancer. Clin Cancer Res 2012; 18: 6784-6791.
31. Chaft JE, Arcila ME, Paik PK et al. Coexistence of PIK3CA and other oncogene mutations in lung adenocarcinoma-rationale for comprehensive mutation profiling. Mol Cancer Ther 2012; 11: 485-491.
32. Roberts PJ, Stinchcombe TE. KRAS mutation: should we test for it, and does it matter? J Clin Oncol 2013; 31: 1112-1121.
33. Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. Biol Proced Online 2013; 15: 4.
34. Xuan J, Yu Y, Qing T et al. Next-generation sequencing in the clinic: Promises and challenges. Cancer Lett 2012.

Whole-genome analysis of a patient with early stage small cell lung cancer

Chan KT, Bennin DA, Huttenlocher A. 2010. Regulation of adhesion dynamics by calpain-mediated proteolysis of focal adhesion kinase (FAK). J Biol Chem 285(15):11418-11426.

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 6(9):677-681.

Daniel VC, Peacock CD, Watkins DN. 2006. Developmental signalling pathways in lung cancer. Respirology 11(3):234-240.

Diez-Sampedro A, Hirayama BA, Osswald C, Gorboulev V, Baumgarten K, Volk C, Wright EM, Koepsell H. 2003. A glucose sensor hiding in a family of transporters. Proc Natl Acad Sci U S A 100(20):11753-11758.

el-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, Lin D, Mercer WE, Kinzler KW, Vogelstein B. 1993. WAF1, a potential mediator of p53 tumor suppression. Cell 75(4):817-825.

Franco SJ, Rodgers MA, Perrin BJ, Han J, Bennin DA, Critchley DR, Huttenlocher A. 2004. Calpain-mediated proteolysis of talin regulates adhesion dynamics. Nat Cell Biol 6(10):977-983.

Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, Chen K, Walker J, McDonald S, Bose R, Ornitz D, Xiong D, You M,

Dooling DJ, Watson M, Mardis ER, Wilson RK. 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 150(6):1121-1134.

Han Y, Weinman S, Boldogh I, Walker RK, Brasier AR. 1999. Tumor necrosis factor-alpha-inducible IkappaBalpha proteolysis mediated by cytosolic m-calpain. A mechanism parallel to the ubiquitin-proteasome pathway for nuclear factor-kappab activation. J Biol Chem 274(2):787-794.

Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. 1993. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell 75(4):805-816.

Janknecht R. 2002. The versatile functions of the transcriptional coactivators p300 and CBP and their roles in disease. Histol Histopathol 17(2):657-668.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40(Database issue):D109-114.

Karamouzis MV, Konstantinopoulos PA, Papavassiliou AG. 2007. Roles of CREB-binding protein (CBP)/p300 in respiratory epithelium tumorigenesis. Cell Res 17(4):324-332.

Kishimoto M, Kohno T, Okudela K, Otsuka A, Sasaki H, Tanabe C, Sakiyama T, Hirama C, Kitabayashi I, Minna JD, Takenoshita S, Yokota J. 2005. Mutations and deletions of the CBP gene in human lung cancer. Clin Cancer Res 11(2 Pt 1):512-519.

Lally BE, Urbanic JJ, Blackstock AW, Miller AA, Perry MC. 2007. Small cell lung cancer: have we made any progress over the last 25 years? Oncologist 12(9):1096-1104.

Lee SJ, Choi YL, Lee EJ, Kim BG, Bae DS, Ahn GH, Lee JH. 2007. Increased expression of calpain 6 in uterine sarcomas and carcinosarcomas: an immunohistochemical analysis. Int J Gynecol Cancer 17(1):248-253.

Lee SJ, Kim BG, Choi YL, Lee JW. 2008. Increased expression of calpain 6 during the progression of uterine cervical neoplasia: immunohistochemical analysis. Oncol Rep 19(4):859-863.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-2079.

Masters GA, Declerck L, Blanke C, Sandler A, DeVore R, Miller K, Johnson D. 2003. Phase II trial of gemcitabine in refractory or relapsed small-cell lung cancer: Eastern Cooperative Oncology Group Trial 1597. J Clin Oncol 21(8):1550-1555.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297-1303.

Miller RW, Rubinstein JH. 1995. Tumors in Rubinstein-Taybi syndrome. Am J Med Genet 56(1):112-115.

Nishio Y, Nakanishi K, Ozeki Y, Jiang SX, Kameya T, Hebisawa A, Mukai M, Travis WD, Franks TJ, Kawai T. 2007. Telomere length, telomerase activity, and expressions of human telomerase mRNA component (hTERC) and human telomerase reverse transcriptase (hTERT) mRNA in pulmonary neuroendocrine tumors. Jpn J Clin Oncol 37(1):16-22.

Noda A, Ning Y, Venable SF, Pereira-Smith OM, Smith JR. 1994. Cloning of senescent cell-derived inhibitors of DNA synthesis using an expression screen. Exp Cell Res 211(1):90-98.

Oike Y, Hata A, Mamiya T, Kaname T, Noda Y, Suzuki M, Yasue H, Nabeshima T, Araki K,

Yamamura K. 1999. Truncated CBP protein leads to classical Rubinstein-Taybi syndrome phenotypes in mice: implications for a dominant-negative mechanism. Hum Mol Genet 8(3):387-396.

Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, Koker M, Dahmen I, Muller C, Di Cerbo V, Schildhaus HU, Altmuller J, Baessmann I, Becker C, de Wilde B, Vandesompele J, Bohm D, Ansen S, Gabler F, Wilkening I, Heynck S, Heuckmann JM, Lu X, Carter SL, Cibulskis K, Banerji S, Getz G, Park KS, Rauh D, Grutter C, Fischer M, Pasqualucci L, Wright G, Wainer Z, Russell P, Petersen I, Chen Y, Stoelben E, Ludwig C, Schnabel P, Hoffmann H, Muley T, Brockmann M, Engel-Riedel W, Muscarella LA, Fazio VM, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman DA, Snijders PJ, Cappuzzo F, Ligorio C, Damiani S, Field J, Solberg S, Brustugun OT, Lund-Iversen M, Sanger J, Clement JH, Soltermann A, Moch H, Weder W, Solomon B, Soria JC, Validire P, Besse B, Brambilla E, Brambilla C, Lantuejoul S, Lorimier P, Schneider PM, Hallek M, Pao W, Meyerson M, Sage J, Shendure J, Schneider R, Buttner R, Wolf J, Nurnberg P, Perner S, Heukamp LC, Brindle PK, Haas S, Thomas RK. 2012. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet 44(10):1104-1110.

Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordonez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ. 2010. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 463(7278):184-190.

Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, Bergbower EA, Guan Y, Shin J, Guillory J, Rivers CS, Foo CK, Bhatt D, Stinson J, Gnad F, Haverty PM, Gentleman R, Chaudhuri S, Janakiraman V, Jaiswal BS, Parikh C, Yuan W, Zhang Z, Koeppen H, Wu TD, Stern HM, Yauch RL, Huffman KE, Paskulin DD, Illei PB, Varella-Garcia M, Gazdar AF, de Sauvage FJ, Bourgon R, Minna JD, Brock MV, Seshagiri S. 2012. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat Genet 44(10):1111-1116.

Sanchez-Cespedes M. 2009. Lung cancer biology: a genetic and genomic perspective. Clin Transl Oncol 11(5):263-269.

Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, Lee J, Jung YJ, Kim JO, Shin JY, Yu SB, Kim J, Lee ER, Kang CH, Park IK, Rhee H, Lee SH, Kim JI, Kang JH, Kim YT. 2012. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res 22(11):2109-2119.

Small GW, Chou TY, Dang CV, Orlowski RZ. 2002. Evidence for involvement of calpain in c-Myc proteolysis in vivo. Arch Biochem Biophys 400(2):151-161.

Sozzi G, Miozzo M, Pastorino U, Pilotti S, Donghi R, Giarola M, De Gregorio L, Manenti G, Radice P, Minoletti F, et al. 1995. Genetic evidence for an independent origin of multiple preneoplastic and neoplastic lung lesions. Cancer Res 55(1):135-140.

Sutherland KD, Proost N, Brouns I, Adriaensen D, Song JY, Berns A. 2011. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. Cancer Cell 19(6):754-764.

Suzuki K, Hata S, Kawabata Y, Sorimachi H. 2004. Structure, activation, and biology of calpain. Diabetes 53 Suppl 1:S12-18.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Ostell J, Pruitt KD, Schuler GD, Shumway M, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. 2008. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 36(Database issue):D13-21.

Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM,

Eckner R. 1998. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. Cell 93(3):361-372.

7. 첨부서류

○ 본 연구의 성과로 논문, 저서, 산업재산권, 정책정책 기여 등이 있을 경우 관련 증빙자료를 첨부토록 함