

기관고유연구사업 최종보고서

편집순서 1 : 겉표지 (앞면)

(과제번호 : 1110520)

NGS를 이용한 위암의 암유전체 연구II

Cancer Genome study for gastric cancer using NGS II

과제책임자 : 이 승 훈

국 립 암 셴 터

편집순서 1 : 겉표지 (측면, 뒷면)

(뒷면)

(측면)

↑
5cm
↓

과
제
명

1. 이 보고서는 국립암센터 기관고유연구사업 최종보고서입니다.
2. 이 보고서 내용을 인용할 때에는 반드시 국립암센터 연구사업 결과임을 밝혀야 합니다.

(14 pont, 고딕체)

국
립
암
셴
터

↑
3cm
↓

↑
6cm
↓

제 출 문

국립암센터 원장 귀하

이 보고서를 기관고유연구사업 “NGS를 이용한 위암의 암유전체 연구 II” 과제의 최종보고서로 제출합니다.

2012. 2 . 17

국립암센터

과 제 책 임 자 : 이 승 훈

연 구 원 : 이 연 수

” : 이 건 국

” : 최 일 주

” : 김 영 우

” : 홍 동 완

” : 박 중 화

” : 김 병 철

” : 이 성 훈

목 차

< 요약 문 >

(한글)

(영문)

1. 연구의 최종목표
2. 연구의 내용 및 결과
3. 연구결과 고찰 및 결론
4. 연구성과 및 목표달성도
5. 연구결과의 활용계획
6. 참고문헌
7. 첨부서류

※ 여러개의 세부과제로 과제가 구성된 경우 위 목차와 동일하게 세부과제별로 작성함
 (I. 총괄과제, II. 제1세부과제, III. 제2세부과제.....)

편집순서 4 : 요약문 (한글)

< 요약문 >

연구분야(코드)				과제번호	1110520
과제명	NGS를 이용한 위암의 암유전체 연구 II				
연구기간/연구비 (천원)	합계	2011년 2월 1일 ~ 2011년 6월 30일		25,000	
	1차년도	2011년 2월 1일 ~ 2011년 6월 30일		25,000	
	2차년도	년 월 일 ~ 년 월 일			
	3차년도	년 월 일 ~ 년 월 일			
과제책임자	성명	이승훈	소속	대의협력실	
	전화번호	031-920-1513	전자우편	nslsh@ncc.re.kr	
색인단어	국문	차세대유전체분석, 위암, 암유전체			
	영문	NGS (Next Generation Sequencing), gastric cancer, cancer genome			

◆ 연구목표

<최종목표>

위암 환자의 sample에서 NGS를 이용하여 얻어진 data로부터, 새로운 유전자 변형 등을 발굴하여 위암의 진단 및 치료에 획기적인 기여를 할 수 있는 원천 지식을 제공하고 본격적인 암 유전체 연구를 위한 정보 분석 기반 확보

<당해연도목표>

- 위암 환자의 NGS 데이터로부터 특이적 유전자 변형 및 유전체의 특성 규명
- Small subset group에 대한 암 유전체 발현 및 후성유전체 분석, NGS와의 연관성 규명
- 분석 결과에 대한 논문 등 정보 공개

◆ 연구내용 및 방법

- Base calling, Data QC 후 SNP를 포함한 coding region에서의 sequence variant analysis (Mutation, CNV 등)
- Cancer 및 normal tissue에서의 결과를 생명정보학 기술을 통하여 비교 분석하여 위암 특이적 유전자 변형 및 유전체의 특성 규명
- Small subset group에서 microarray를 이용한 transcriptom 및 methylom 분석
 - 1) mutation과 유전자 발현과의 상관관계 분석을 통한 functional mutation 확보
 - 2) mutation과 DNA methylation 및 유전자 발현과의 상관관계 분석을 통한 tumor associate gene의 candidate 확보

-

◆ 연구성과

-정량적 성과

구분	달성치/목표치 ¹⁾	달성도(%)
SCI 논문 편수	1*/2	
IF 합	?/30	
기타 성과		

* submitted (IF 30점 이상의 journal에 논문을 제출한 상태임)

-정성적 성과

- NGS를 활용한 암 유전체 분석 체계 구축
- 위암 특이적 변이체 데이터 제공
- 위암 관련 유전자 후보군 도출

◆ 참여연구원 (최종연도 참여인원)	성명	이연수, 이건국, 최일주, 김영우, 홍동완, 박종화, 이성훈, 김병철, 이진수

※ 요약문의 총분량은 2page 이내로 제한함

Project Summary

Title of Project	Cancer Genome study for gastric cancer using NGS II
Key Words	NGS (Next Generation Sequencing), gastric cancer, cancer genome
Project Leader	Seung Hoon Lee
Associated Company	
<p>We report genome-wide mutation profiles and comparison of 14 matched diffuse type and 4 matched intestinal type tumor-blood samples for variation comparison of two cancer types in young female patients (median age 38) using ~32 fold whole genome and an additional ~100 fold exome sequences for three samples. We found 376 and 196 mutated genes in the diffuse and intestinal cancer genome pools, respectively, and just 13 (2.3%) were shared by both types.</p> <p>PIK3CA, a multi functional oncogene, had mutations only in three diffuse type tumors and MDM2 showed two and three fold gene amplification only in the diffuse type. qRT-PCR confirmed increased expression of MDM2 concordantly. On the other hand, TP53 and CDH1 tumor suppressors showed aminoacid-changing mutations in both types. Large genomic region amplification was found which contained oncogenes, such as MYC and MET, in both types. We also observed noticeable genomic region deletion that contained tumor suppressors, such as CDH1 and TP53, in both types, but the chromosome 1p-arm region containing RUNX3 and PLA2G2A tumor suppressor genes was extensively deleted only in the diffuse type. Pathway component analysis of the mutated genes revealed type specific features. In KEGG cancer pathways, the diffuse type tumors contained aberrant genes associated with cell proliferation involving PI3 kinase. In terms of characteristics, diffuse type gastric cancer resembled small cell lung cancer. Taken together, our study found gene and mutation candidates and onco-pathway characteristics specifically prevalent in diffuse type gastric cancer, suggesting that two gastric cancer types are distinct and can be described as such by genome sequencing methods.</p>	

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

1. 연구의 최종목표

(1) 최종목표 :

위암 환자의 sample에서 NGS를 이용하여 얻어진 data로부터, 새로운 유전자 변형 등을 발굴하여 위암의 진단 및 치료에 획기적인 기여를 할 수 있는 원천 지식을 제공하고 본격적인 암 유전체 연구를 위한 정보 분석 기반 확보

(2) 당해년도 목표 및 내용

구분	목표	내용 및 범위
2011년도	- Gastric cancer & normal tissue에 대한 Next Generation Sequencing 분석 결과 QC	- Sequencing Platform : Illumina GAIIx, 2x75bp - Sample 당 Sequencing Depth >x30,
	- 위암 특이적 유전자 변형 및 유전체의 특성 규명	- Mutation, CNV, SNP을 포함한 sequence variant analysis
	- small subset group에 대한 암 유전체 발현 및 후성유전체 분석, NGS와의 연관성 규명	- Small subset group에서 microarray를 이용한 transcriptom 및 methylom 분석: mutation과 DNA methylation 및 유전자 발현과의 상관관계 분석과 functional mutation 확보

2. 연구의 내용 및 결과

1. Data QC and basic curation

a. Basecalling

NGS 분석기기로부터 얻어진 sequence 이미지 데이터를 해석하여 A, G, C, T로 이루어진 염기서열과 각각의 염기서열에 대한 quality값을 얻는 과정이다.

b. Alignment

실험 목적에 맞는 referenc sequence에 실험으로 얻어진 read를 맵핑하는 과정이다.

NGS 데이터 분석에서 중요한 단계중 하나는 시퀀스 데이터를 reference에 맵핑하는 것이다. BLAST와 같은 툴은 NGS 데이터에 적용하기에는 계산 시간이 너무 긴 문제점이 있기 때문에 NGS 데이터 분석을 위해 GSnep, ELAND2, MAQ, BOWTIE, SOAP2, BWA 등 매우 다양한 분석방법들이 개발되었고, 이 중에 BWT (Burrows-Wheeler Transformation) 알고리즘을 사용하는 툴의 경우(BOWTIE, SOAP2,BWA)에는 일반적으로 BLAST에 비해 수백~수만배의 속도로 맵핑이 가능하게 되었다.

Basic Curation의 Base calling과 alignment를 통해 얻어진 데이터들을 요약하면 표 1과 같다.

표1. Sequencing statistics

Cancer	ID	total (Gb)	mapped	sequencing	mappingdepth	total reads	mapped reads
Diffuse	D-01B	86.48	82.22	30.22	28.73	960,865,581	913,524,291 (95.07%)
	D-01T	97.53	93.66	34.09	32.73	1,083,718,112	1,040,634,002 (96.02%)
	D-02B	106.83	101.08	37.33	35.33	1,186,975,847	1,123,125,683 (94.62%)
	D-02T	94.84	91.02	33.15	31.81	1,053,828,338	1,011,362,470 (95.97%)
	D-03B	108.13	89.39	37.78	31.24	1,201,490,848	993,222,812 (82.67%)
	D-03T	201.92	107.36	70.56	37.52	2,243,598,820	1,192,907,602 (53.17%)
	D-04B	93.79	89.52	32.78	31.28	1,042,157,557	994,627,921 (95.44%)
	D-04T	87.84	85.13	30.7	29.75	976,017,163	945,939,153 (96.92%)
	D-05B	81.18	77.76	28.37	27.18	901,986,574	864,049,716 (95.79%)
	D-05T	94.31	90.76	32.96	31.72	1,047,836,756	1,008,397,034 (96.24%)
	D-06B	84.29	80.52	29.46	28.14	936,578,571	894,668,495 (95.53%)
	D-06T	104.06	99.73	36.37	34.86	1,156,184,318	1,108,141,790 (95.84%)
	D-07B	71.06	47.12	24.84	16.47	789,601,782	523,507,602 (66.30%)
	D-07T	82.67	79.77	28.89	27.88	918,606,800	886,296,876 (96.48%)
	D-08B	82.26	78.74	28.75	27.52	913,984,039	874,926,613(95.73%),
	D-08T	82.05	78.72	28.68	27.51	911,697,785	874,704,379(95.94%)
	D-09B	83.33	79.95	29.12	27.94	925,855,354	888,360,238 (95.95%)
	D-09T	85.48	82.58	29.88	28.86	949,832,205	917,550,733 (96.60%)
	D-10B	91.46	86.82	31.96	30.34	1,016,223,735	964,677,731 (94.93%)
	D-10T	88.22	84.93	30.83	29.68	980,254,769	943,661,255 (96.27%)
	D-11B	92.83	88.23	32.44	30.84	1,031,497,990	980,328,676 (95.04%)
	D-11T	84.13	81.2	29.4	28.38	934,818,450	902,230,978 (96.51%)
	D-12B	95.33	90.89	33.32	31.77	1,059,184,112	1,009,897,100 (95.35%)
	D-12T	99.37	95.29	34.73	33.3	1,104,126,335	1,058,774,473 (95.89%)
D-13B	90.77	86.62	31.72	30.27	1,008,575,478	962,477,672 (95.43%)	
D-13T	97.48	87.99	34.06	30.75	1,083,142,004	977,640,141 (90.25%)	
D-14B	94.99	90.28	33.2	31.55	1,055,494,193	1,003,070,381 (95.03%)	
D-14T	96.64	93.05	33.77	32.52	1,073,743,176	1,033,856,700 (96.29%)	
Intestinal	I-01B	106.69	101.66	37.29	35.53	1,185,497,937	1,129,586,825 (95.28%)
	I-01T	84.92	81.14	29.68	28.36	943,519,000	901,546,422 (95.55%)
	I-02B	102.96	97.84	35.98	34.19	1,143,965,945	1,087,121,109 (95.03%)
	I-02T	79.24	75.88	27.69	26.52	880,405,197	843,122,159 (95.77%)
	I-03B	84.38	80.88	29.49	28.26	937,588,331	898,617,051 (95.84%)
	I-03T	86.31	82.84	30.16	28.95	958,946,893	920,459,735 (95.99%)
	I-04B	84.75	81.29	29.62	28.41	941,654,003	903,264,715 (95.92%)
I-04T	81.92	78.98	28.63	27.6	910,240,469	877,526,355 (96.41%)	

c. Variant detection & Reporting

Annotation 정보와 함께 생성된 테라급 데이터를 분석하기 위해서는 많은 스토리지가 필요하며, 생물정보학적 툴의 개발이 필요함. 시퀀싱 데이터 분석을 위해 크게 Alignment를 직접적으로 분석하면서 얻어지는 SNP, Short InDel 등 정보와 Paired-end 시퀀싱 Alignment에서 얻어지는 각

중 Long deletion 및 Structural Variation (Inversion, Translocation)을 생각할 수 있다. 또한 Coverage를 기반으로 계산이 가능한 CNV (Copy Number Variation)가 있다. 이렇게 얻어진 각종 변이(Variation) 데이터들은 기존에 이미 알려진 DB인 dbSNP, DGV, HGMD 등과 비교하거나 이미 알려진 유전자 구조 정보를 활용함으로써 좀더 유용한 정보를 얻을 수 있음. Hybridization 기술과 NGS 기술의 조합에 의한 타겟 시퀀싱의 경우에도 각종 생물정보학 알고리즘 및 도구 개발에 의해서 분석이 가능하나, 특정한 부분만을 선별해서 분석을 해야 하므로 데이터 분석에 주의하여야 한다.

분석된 위암특이적 변이들을 요약하면 아래의 그림1과 같다.

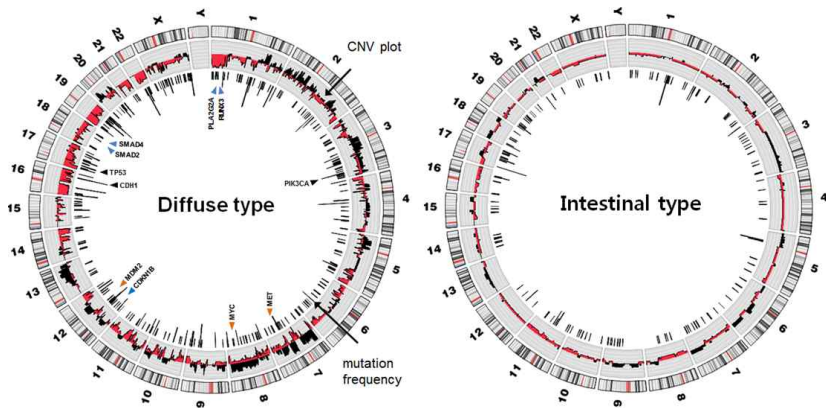


그림1. Whole genome distribution of SNVs and duplication or deletion events in diffuse type gastric cancers

2. 위암에서 나타나는 유전적 변이를 확보

위암에 영향을 주는 유전적 요인 발굴을 통한 mechanism 연구 및 진단치료 연구에 응용. 차세대 염기서열법이 발표된 이후, 모든 유전자의 exon을 sequencing 함으로써, 특히 표현형에 직접적 영향을 주는 돌연변이(mutation)를 찾는데 매우 유용한 방법이다. 암 조직의 경우는 멘델질환과는 다르게 매우 많은 돌연변이가 관여할 것으로 예상되며, 이러한 변이를 수 많은 정보에서 어떻게 정제하는가가 연구 성패의 관건일 것이며 따라서 생명정보학적 분석이 매우 중요하다. (이와 관련하여, 최근 Lung cancer에서 530개 변이를 발굴하여 발표했다, Nature. 2010 May 27;465(7297):473-477).

Mutation detection 외에 whole genome 분석을 통해 얻을 수 있는 것들은, abnormal mRNA isoform의 규명 (알려지지 않은 intronic splicing variant에 의해 발생), SNP 및 CNV들을 포함하는 polymorphism의 발굴, 위암에 특이적인 유전자 변이의 빈도 측정 (measuring the average frequency of new variants per gene across the genome) 등을 들 수 있으며 더 나아가 위암과

연관된 유전자들을 발굴해 낼 수 있었다.

Whole genome sequencing은 exome sequencing에서 missing할 수 있는 염기서열을 찾을 수 있다. exon 뿐만 아니라 intron 정도도 얻을 수 있으므로 모든 genome에서의 SNP, InDel, Structural variation, copy number variation등을 찾을 수 있습니다. 이렇게 변이 정보들을 관련된 DNA 등과 비교 분석함으로써 잠재적으로 암과 연관성이 높은 마커들을 발견할 수 있으며, 이들은 암의 조기발견 및 진단에 활용될 수 있다.

Human genome DNA 염기서열 정보로부터 얻을 수 있는 가장 핵심적인 부분은 개인차를 가장 잘 설명하는 집단 안에서 변이(variation)이며, SNP, InDel, Structural Variation, CNV 등이 있습니다. 변이 정보들을 관련된 DNA등과 비교 분석함으로써 잠재적으로 질병과 연관성이 높은 마커들을 발굴할 수 있다. 이는 분자진단을 위한 핵심 바이오 마커 발굴에도 이용할 수 있어 질병의 예측과 진단에 활용될 수 있다.

연구결과들을 sample 별로 요약하면, 그림2와 같으며, 각각을 variation 별로 상세히 살펴보면 아래와 같다.

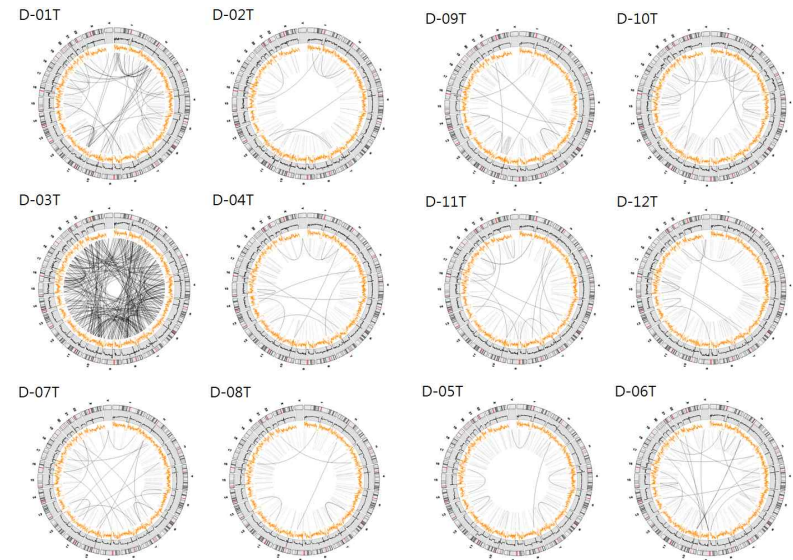


그림 2. Variation of tumor genomes. From the outer side of each ring, chromosome ideograms, mapping depths of chromosome regions (black circle), SNV density of chromosome regions (orange line), and large translocation of intra and inter chromosomal rearrangements (black lines across center).

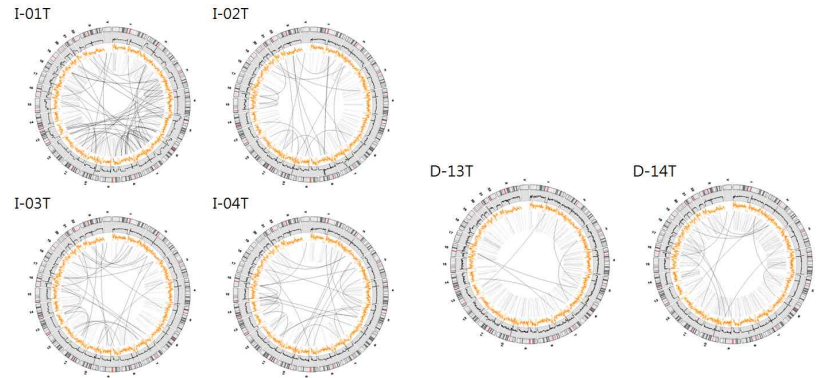


그림 2. Variation of tumor genomes. (계속)

a. SNV (Single Nucleotide Variation)

Genome의 염기 서열이 reference와 다르게 변이된 단일염기변이 (SNP)를 찾아 list up 했다. (표 2)

표2. SNVs in 18 gastric cancer genomes

Cancer	ID	total (Gb)	mapped	sequencing	mappingdepth	total reads	mapped reads
Diffuse	D-01B	86.48	82.22	30.22	28.73	960,865,581	913,524,291 (95.07%)
	D-01T	97.53	93.66	34.09	32.73	1,083,718,112	1,040,634,002 (96.02%)
	D-02B	106.83	101.08	37.33	35.33	1,186,975,847	1,123,125,683 (94.62%)
	D-02T	94.84	91.02	33.15	31.81	1,053,828,338	1,011,362,470 (95.97%)
	D-03B	108.13	89.39	37.78	31.24	1,201,490,848	993,222,812 (82.67%)
	D-03T	201.92	107.36	70.56	37.52	2,243,598,820	1,192,907,602 (53.17%)
	D-04B	93.79	89.52	32.78	31.28	1,042,157,557	994,627,921 (95.44%)
	D-04T	87.84	85.13	30.7	29.75	976,017,163	945,939,153 (96.92%)
	D-05B	81.18	77.76	28.37	27.18	901,986,574	864,049,716 (95.79%)
	D-05T	94.31	90.76	32.96	31.72	1,047,836,756	1,008,397,034 (96.24%)
	D-06B	84.29	80.52	29.46	28.14	936,578,571	894,668,495 (95.53%)
	D-06T	104.06	99.73	36.37	34.86	1,156,184,318	1,108,141,790 (95.84%)
	D-07B	71.06	47.12	24.84	16.47	789,601,782	523,507,602 (66.30%)
	D-07T	82.67	79.77	28.89	27.88	918,606,800	886,296,876 (96.48%)
	D-08B	82.26	78.74	28.75	27.52	913,984,039	874,926,613(95.73%)
	D-08T	82.05	78.72	28.68	27.51	911,697,785	874,704,379(95.94%)
	D-09B	83.33	79.95	29.12	27.94	925,855,354	888,360,238 (95.95%)
	D-09T	85.48	82.58	29.88	28.86	949,832,205	917,550,733 (96.60%)
D-10B	91.46	86.82	31.96	30.34	1,016,223,735	964,677,731 (94.93%)	
D-10T	88.22	84.93	30.83	29.68	980,254,769	943,661,255 (96.27%)	
D-11B	92.83	88.23	32.44	30.84	1,031,497,990	980,328,676 (95.04%)	
D-11T	84.13	81.2	29.4	28.38	934,818,450	902,230,978 (96.51%)	
D-12B	95.33	90.89	33.32	31.77	1,059,184,112	1,009,897,100 (95.35%)	

	D-12T	99.37	95.29	34.73	33.3	1,104,126,335	1,058,774,473 (95.89%)
	D-13B	90.77	86.62	31.72	30.27	1,008,575,478	962,477,672 (95.43%)
	D-13T	97.48	87.99	34.06	30.75	1,083,142,004	977,640,141 (90.25%)
	D-14B	94.99	90.28	33.2	31.55	1,055,494,193	1,003,070,381 (95.03%)
Intestinal	D-14T	96.64	93.05	33.77	32.52	1,073,743,176	1,033,856,700 (96.29%)
	I-01B	106.69	101.66	37.29	35.53	1,185,497,937	1,129,586,825 (95.28%)
	I-01T	84.92	81.14	29.68	28.36	943,519,000	901,546,422 (95.55%)
	I-02B	102.96	97.84	35.98	34.19	1,143,965,945	1,087,121,109 (95.03%)
	I-02T	79.24	75.88	27.69	26.52	880,405,197	843,122,159 (95.77%)
	I-03B	84.38	80.88	29.49	28.26	937,588,331	898,617,051 (95.84%)
	I-03T	86.31	82.84	30.16	28.95	958,946,893	920,459,735 (95.99%)
	I-04B	84.75	81.29	29.62	28.41	941,654,003	903,264,715 (95.92%)
	I-04T	81.92	78.98	28.63	27.6	910,240,469	877,526,355 (96.41%)

b. CNV (Copy Number Variation)

일정한 염기 서열이 하나의 genome 안에서 반복적으로 존재할 때 이를 반복서열변이 (Copy Number Variation) 라고 하며 CNV의 후보 서열과 위치를 찾아낼 수 있다.

본 연구에서는 특히, 기존에 암에서의 역할이 잘 알려진 MDM2의 copy수가 특정 위암환자들에게서 증가한 것을 발견하고 유전자 발현 수준이 증가한 것을 확인했다. (그림3)

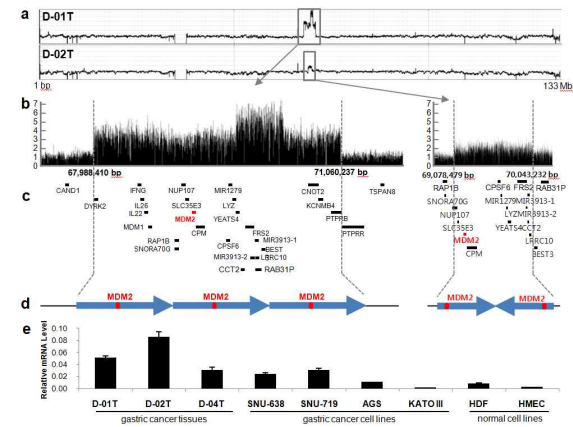


그림3. Duplication region of MDM2 gene on chromosome 12 in samples D-01T and D-02T

c. InDel

Genome의 염기 서열을 reference와 비교하여 수 base의 차이가 있는 부분을 찾아 어떠한 서열이 삽입되고 삭제되었는지 알 수 있으며, 본 연구에서 다양한 InDel mutation을 발견하였다. (표3)

표3. Small indels in 18 gastric cancer genomes

ID	Tissue	Total	PROMOTER	5UTR	CDS	SPLICING	3UTR	INTRON
D-01	Blood	643,663	24,210	535	346	6	4,903	242,448
	Tumor	704,674	27,753	626	401	7	5,353	267,355
D-02	Blood	706,185	28,351	629	373	3	5,460	269,990
	Tumor	673,305	25,336	592	318	3	5,139	264,504
D-03	Blood	707,457	27,254	617	377	5	5,362	267,019
	Tumor	814,818	32,295	735	436	4	5,985	309,664
D-04	Blood	693,592	27,546	684	409	4	5,378	261,387
	Tumor	698,650	26,768	660	388	5	5,366	262,196
D-05	Blood	682,469	26,210	632	384	3	5,221	256,107
	Tumor	683,790	27,200	656	425	3	5,299	258,600
D-06	Blood	696,127	27,140	649	393	3	5,248	263,242
	Tumor	714,221	28,022	655	401	6	5,389	271,097
D-07	Blood	863,279	34,732	736	470	7	6,384	331,841
	Tumor	699,576	26,384	583	380	4	5,346	263,170
D-08	Blood	692,190	26,588	628	372	3	5,261	260,473
	Tumor	685,302	26,258	651	378	3	5,230	257,707
D-09	Blood	682,092	26,132	635	382	2	5,182	257,097
	Tumor	699,055	26,911	650	379	3	5,335	265,059
D-10	Blood	613,344	24,099	591	360	3	4,854	232,094
	Tumor	649,740	24,440	597	359	5	5,013	245,233
D-11	Blood	661,455	26,211	628	404	5	5,174	250,682
	Tumor	638,012	23,755	576	384	5	4,887	239,469
D-12	Blood	679,860	26,638	659	382	6	5,307	257,796
	Tumor	683,625	26,220	641	384	5	5,266	258,133
D-13	Blood	656,465	26,853	692	433	8	5,110	249,970
	Tumor	662,816	25,156	648	380	6	5,112	249,681
D-14	Blood	648,771	26,165	699	395	6	5,033	245,532
	Tumor	683,598	26,781	649	376	5	5,217	258,156
I-01	Blood	707,884	28,076	639	371	6	5,470	268,646
	Tumor	666,001	25,649	573	321	4	4,958	252,809
I-02	Blood	719,737	28,511	669	388	7	5,425	272,223
	Tumor	678,271	25,709	594	353	5	5,033	253,796
I-03	Blood	681,734	25,823	602	376	5	5,218	256,529
	Tumor	704,496	26,828	623	381	5	5,393	266,116
I-04	Blood	673,267	26,534	663	394	5	5,193	253,630
	Tumor	683,877	25,800	637	374	4	5,143	256,426

d. Structural variation (Insertion / Deletion / Inversion / Rearrangement)

Genome의 염기 서열 중 일정 부분이 reference와 대비하여 삭제되거나, 다른 곳에 삽입되거나, 서

열의 방향이 뒤바뀌는 등 염기 서열 구조에 변화가 있을 때 이를 찾아낼 수 있다. 본 연구에서 밝힌 structural variation들을 정리해 보면 표4와 같다.

표4. Structural variation counts. Structural variations were detected using Breakdancer program with a score >=30.

a. Diffuse type					b. Intestinal type				
Sample	Type	Normal	Tumor	Somatic SV	Sample	Type	Normal	Tumor	Somatic SV
D-01	DEL	6171	6607	405	I-01	DEL	5648	7536	580
	INS	2585	2406	166		INS	4202	4403	443
	INV	235	357	37		INV	277	283	29
	ITX	274	440	45		ITX	288	293	35
	CTX	375	680	27		CTX	375	540	60
D-02	DEL	7551	5916	181	I-02	DEL	6681	8142	389
	INS	3173	3063	224		INS	3347	3183	229
	INV	272	374	32		INV	259	315	33
	ITX	308	285	23		ITX	273	295	33
D-03	CTX	564	461	1	CTX	414	591	7	
	DEL	8054	3119	91	I-03	DEL	8146	5860	122
	INS	4543	0	0		INS	4650	3722	165
	INV	283	589	62		INV	425	690	40
ITX	250	12425	148	ITX		296	365	56	
D-04	CTX	401	1786	373	CTX	403	473	16	
	DEL	6662	6467	75	I-04	DEL	6436	6159	55
	INS	3614	4202	318		INS	4144	4722	301
	INV	266	589	4		INV	489	512	31
ITX	313	298	20	ITX		300	308	27	
D-05	CTX	438	396	3	CTX	412	524	13	
	DEL	8156	5953	300	D-06	DEL	5799	5297	198
	INS	4870	2689	131		INS	1965	2841	103
	INV	380	484	30		INV	398	487	15
ITX	301	439	41	ITX		400	550	26	
D-07	CTX	428	359	1	CTX	570	458	0	
	DEL	3025	6109	212	D-08	DEL	5937	4648	217
	INS	0	3937	387		INS	4713	3163	260
	INV	520	913	13		INV	325	752	23
ITX	5786	300	31						
D-08	CTX	3680	609	2					

D-09	ITX	293	312	36
	CTX	427	1123	2
	DEL	6970	7412	371
	INS	4599	2819	242
	INV	538	1318	45
D-10	ITX	272	307	34
	CTX	404	703	6
	DEL	7361	6841	280
	INS	3283	3100	278
D-11	INV	253	334	18
	ITX	249	262	35
	CTX	334	390	8
	DEL	6356	6307	157
D-12	INS	4263	2830	114
	INV	243	390	19
	ITX	279	284	21
	CTX	380	372	8
D-13	DEL	6643	6338	351
	INS	3660	2815	268
	INV	253	271	16
	ITX	291	285	40
D-14	CTX	445	404	3
	DEL	6536	6375	426
	INS	2853	2675	236
	INV	241	396	53
D-14	ITX	289	327	61
	CTX	320	451	6
	DEL	6986	6373	347
	INS	2747	3652	233
D-14	INV	257	296	18
	ITX	303	325	41
	CTX	351	397	2

e. Transition and Transversion

Base가 어떻게 변하는 가는 암종마다 특이하게 나타난다고 보고되어 있으며, 이에 따라 본 연구에서도 위암 특이적 transition과 transversion을 분석했다. (그림 4)

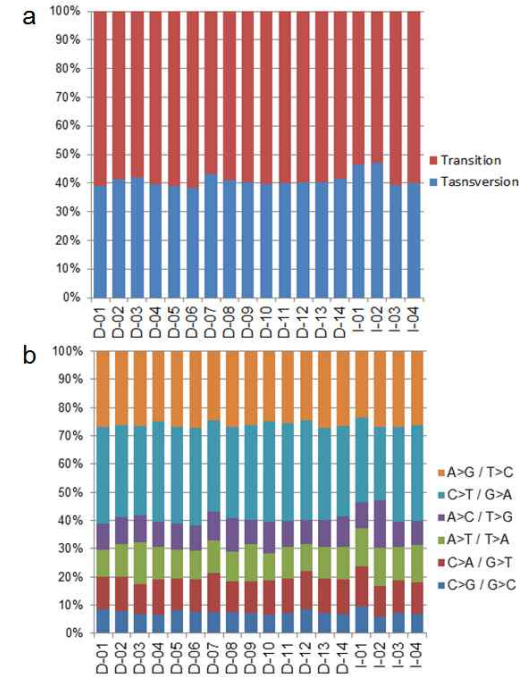


그림 4. Somatic mutation profiles. a, Transversion and transition ratio b, Nucleotide substitution ratio.

3. 연구결과 고찰 및 결론

1. NGS 연구를 위한 DNA 분리 정제 기술 확보 및 QC 기준 확립

Macrodissection을 통한 암세포 분리 및 DNA 정제를 시행하여 암세포 특이적 변이를 찾는 것을 용이하게 함. (그림 5)

시료 QC:

- UV-spectrophotometry에 의한 QC, Nanodrop 사용 (O.D.260/280 기준으로 1.6이상, 총량 15ug 이상 사용, 표5)
- DNA size fractionation by gel electrophoresis (그림 6)

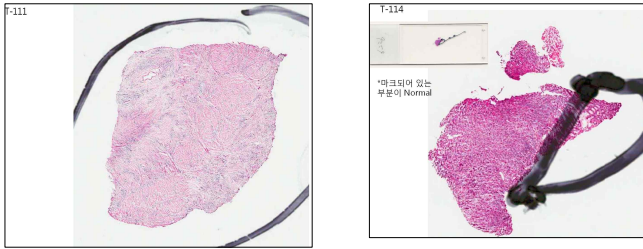


그림 5. Macrodissection을 통한 암세포 분리의 예

표5. Macrodissection을 통해 추출한 DNA의 quantity 및 quality

ID	Volume (ul)	Conc_UV	260/280	260/230	UV기준_총량(ug)
T-108	529	28.6	1.7	2.66	15.13
T-109	525	27.9	1.69	2.13	14.65
T-110	146	102.8	1.93	2.36	15.01
T-111	130	116.5	1.87	2.22	15.15

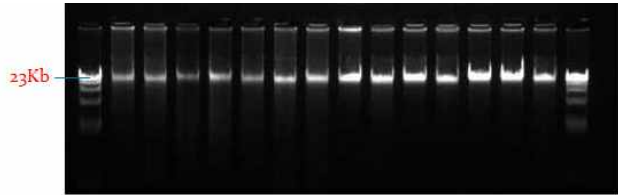


그림 6. Macrodissection을 통해 추출한 DNA의 fragmentation monitoring

2. NGS 기술을 이용한 whole genome 분석의 프로세스 확립 (그림 7)

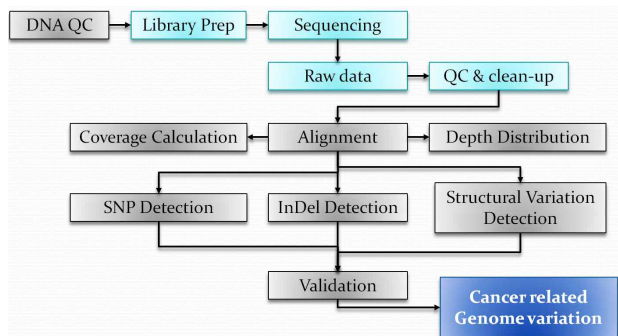


그림 7. NGS를 이용한 암 특이적 유전체 변형 발굴 프로세스 (모식도)

3. 위암, 특히 diffuse type 특이적 유전체 변형 발견 및 신규 유전자 변이 탐색

이제까지 보고된 바 없는, diffuse type Gastric Cancer의 유전체 변형들을 발굴하였으며 특히 암과의 연관성이 알려지지 않은 신규 유전자의 변이들을 다수 탐색해냈다. (그림 8 및 표6)

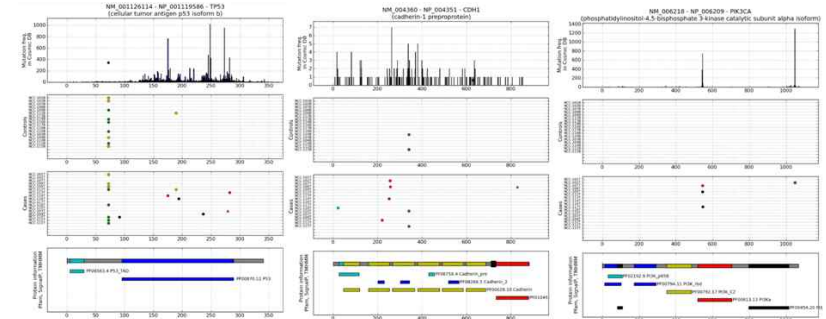


그림 8. 위암의 NGS 연구를 통해 발굴한 신규 유전자 변이의 예

표6. Sanger sequencing을 통해 확인된, 위암 특이적 유전자 변형

Sample	Gene	Chr. Nr.	Variation	Position	Allele change
I01-T	TP53	chr17	indel	7577102	-CAGGACAGGCACAAACCG
D03-T	PHF2	chr9	indel	96439004	+CCTGCCTCCACCACA
I01-T	CCDC39	chr3	indel	180337124	-TAGTTGAATTTT
D10-T	CAPN1	chr11	indel	64974107	+GTGCT
D10-T	PTEN	chr10	indel	89685314	-GTAA
D07-T	C12orf1	chr12	indel	91347581	+TCT
I01-T	SCN5A	chr3	indel	38601696	-T
I01-T	LCE2B	chr1	indel	152659422	-A
D10-T	A4GNT	chr3	indel	137843404	-G
D05-T	CDH1	chr16	indel	68867236	+T
D04-T	PIK3CA	chr3	SNV	178936091	G>A
D06-T	PIK3CA	chr3	SNV	178936091	G>A
I02-T	TP53	chr17	SNV	7577574	T>G
D10-T	TP53	chr17	SNV	7578269	T>A
D12-T	CDH1	chr16	SNV	68772218	C>T
I03-T	CDH1	chr16	SNV	68842726	A>G
D09-T	CDH1	chr16	SNV	68844167	T>G
D02-T	CDH1	chr16	SNV	68844179	A>G
D05-T	CDH1	chr16	SNV	68844181	G>A
I04-T	DEGS1	chr1	SNV	224377572	A>T
D12-T	CDK6	chr7	SNV	92404109	T>A
D12-T	SMAD2	chr18	SNV	45368211	G>C

4. 연구성과 및 목표달성도

(1) 연구성과

가. 국내 및 국제 전문학술지 논문 게재 및 신청

논문명	저자 (저자구분 ¹⁾)	저널명(IF.)	Year: Vol(No):Page	구분 ²⁾	지원과제번호 ³⁾
Genomic profile analysis of diffuse type gastric cancers by whole genome sequencing	이승훈 (교신)	Nature Genetics (36.377)	Submitted	국외 SCI	1110520

- 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

- 1) 지역 : 국내, 국외

다. 산업재산권

구분 ¹⁾	특허명	출원인	출원국	출원번호

- 1) 구분 : 발명특허, 실용신안, 의장등록 등

라. 저서

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
위암 환자의 sample에서 NGS를 이용하여 얻어진 data로부터, 새로운 유전자 변형 등을 발굴하여 위암의 진단 및 치료에 획기적인 기여를 할 수 있는 원천 지식을 제공하고 본격적인 암 유전체 연구를 위한 정보 분석 기반 확보	1차년도	Gastric cancer & normal tissue에 대한 Next Generation Sequencing 분석 결과 QC - Sequencing Platform : Illumina GAIIx, 2x75bp - Sample 당 Sequencing Depth >x30,	90	90
		위암 특이적 유전자 변형 및 유전체의 특성 규명 - Mutation, CNV, SNP을 포함한 sequence variant analysis	100	
		위암 특이적 유전자 변형 및 유전체의 특성 규명에 대한 정보를 논문 및 DB 등으로 공개 NGS data의 논문 등을 통한 공개	80	

나. 평가의 착안점에 따른 목표달성도에 대한 자체평가

평가의 착안점	자체평가
20 pair의 sample들에 대해 whole genome based analysis를 수행하여 얻은, sample 당 Sequencing Depth >x30 수준의 염기서열에 대해 QC를 마쳤는가?	NGS 분석을 모두 완료했음. 그러나 하나의 sample은 neuroendocrine으로 확인되어 제외했고, 하나의 sample은 데이터 분석 결과 MSI로 밝혀져 최종 report에서 제외했음. 논문에 report하는 것을 기준으로 한다면 90%의 완성도를 보임
Mutation, CNV, SNP을 포함한 sequence variant analysis를 수행하였는가	모두 수행하였으며, indel 및 intra-, inter chromosomal translocation과 base transition등 추가 분석까지 수행했으므로 완성도 100%로 평가
NGS data의 논문 등을 통한 공개가 이루어 졌는가?	whole genome NGS에 이어, 원래 계획에 없던 exom sequencing을 추가하여 논문을 revise함. 그에 따라 처음 계획보다는 늦어졌으나 현재 논문을 submit 한 상태이므로 80% 달성으로 평가

5. 연구결과의 활용계획

(1) 연구종료 2년후 예상 연구성과

구분	건수	비고
학술지 논문 게재	2	Genome Research, Gastroenterology 등 IF 10점 이상의 journal
산업재산권 등록		특허 등록 예상 국가, 예상 특허명 등
기타		

(2) 연구성과의 활용계획

- 본 연구를 통해 구축된, NGS를 활용한 암 유전체 분석 체계를 활용하여 다른 암종의 유전체 변이 분석 (2012년 신규 기관고유사업과제로 “역형성갑상선암의 유전체연구를 통한 신규 표적 발굴” 연구가 진행되고 있음)

- 위암 특이적 변이체 데이터 제공: 2012년 신규 기관고유사업 목적과제로 “국립암센터 호발암 특이적 변이체 데이터베이스 구축” 과제가 진행되고 있으며 이를 통해 본 과제에서 발굴한 위암 특이적 변이체 데이터베이스를 구축하고 연구자들에게 데이터를 충분히 활용할 수 있도록 지원할 예정임

- 위암 관련 유전자 후보군 도출 및 신규 타겟 발굴을 위한 데이터 제공: 본 연구를 통해 발굴된 위암관련 유전자 후보군들을 신규 항암 타겟 발굴을 위한 연구에 제공

- 추가 후속 연구의 필요성: 본 연구에서 도출된 후보군들은 large scale genome study를 통해 발굴된 것으로 clinical sample들에서의 true positive에 대한 validation 및 functional role에 대한 검증이 필요함

6. 참고문헌

○ 보고서 작성시 인용된 모든 참고문헌을 열거

7. 첨부서류

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