CBMC 희귀질환 센터 컨퍼런스: 2020.07.15.

NGS data 해석을 위한 기본개념

Reference Genome, HGVS nomenclature

Jihoon G. Yoon, MD

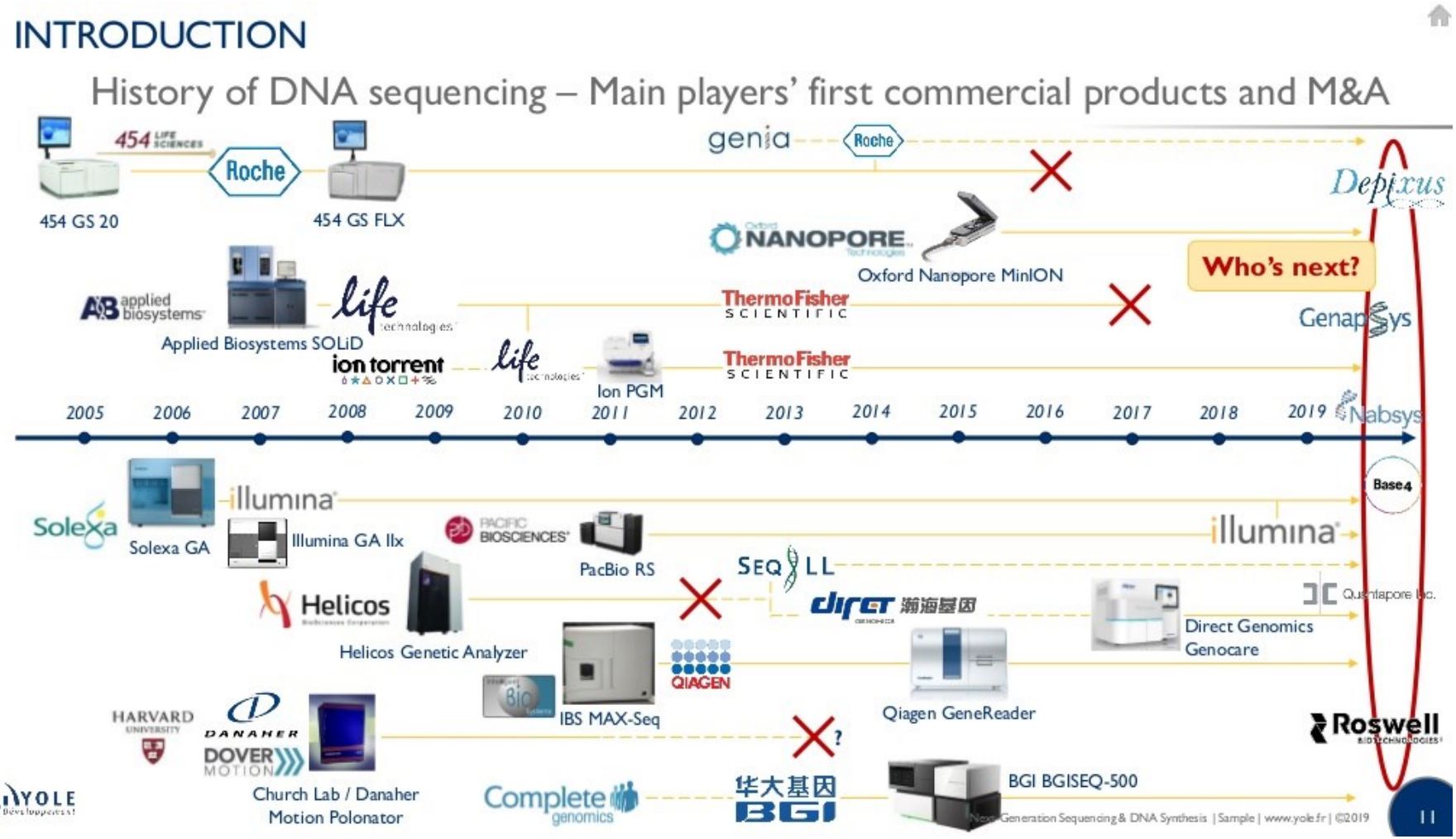
Department of Pharmacology

Yonsei University College of Medicine

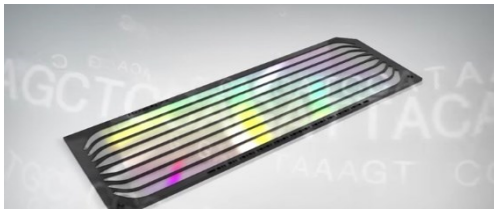
강의 내용 및 순서

- Next-generation sequencing의 기본 원리
- 시퀀싱 Data로부터 변이 정보를 얻기까지의 Process
- 변이 정보 (VCF) 해석과 임상 적용을 위한 배경 지식
  - Reference Genome
  - HGVS nomenclature
  - ACMG classification
  - Genotype-Phenotype correlation

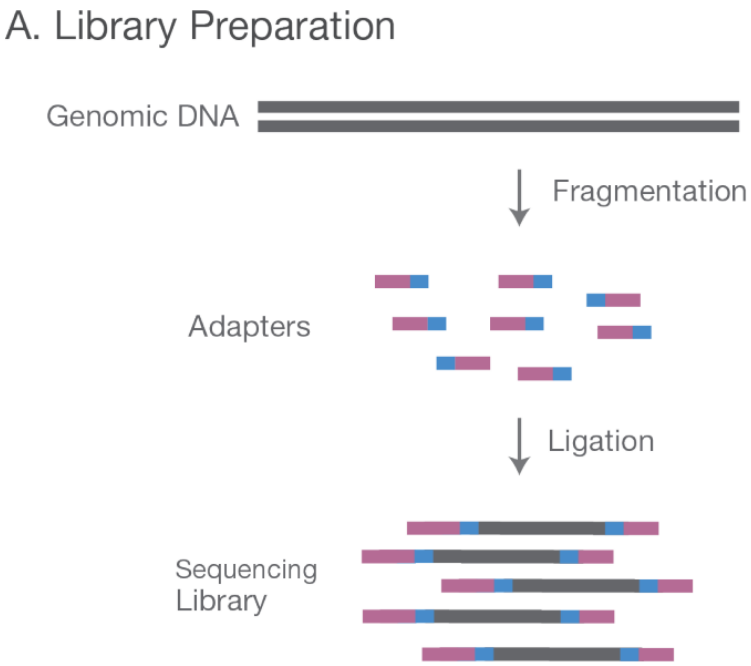
Next-generation sequencing (NGS): Overview



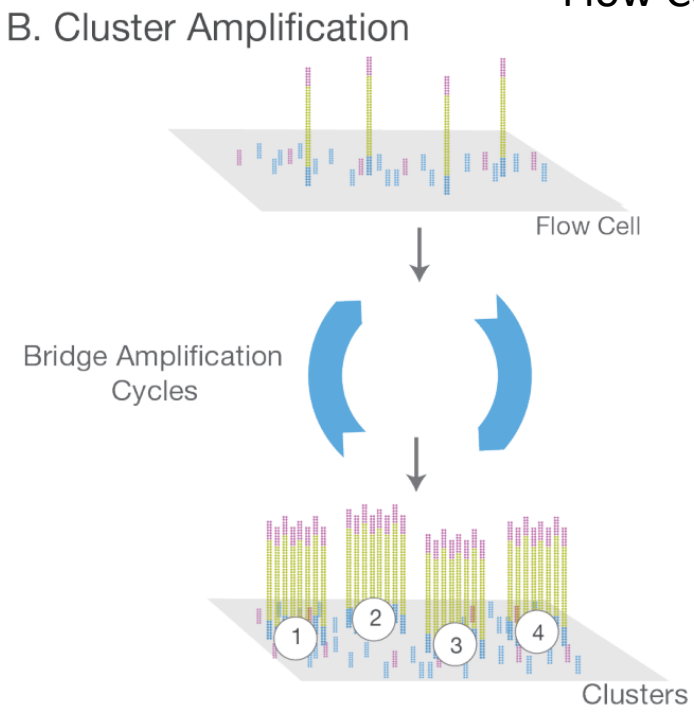
Illumina sequencing: Basic Principle



Flow Cell



NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

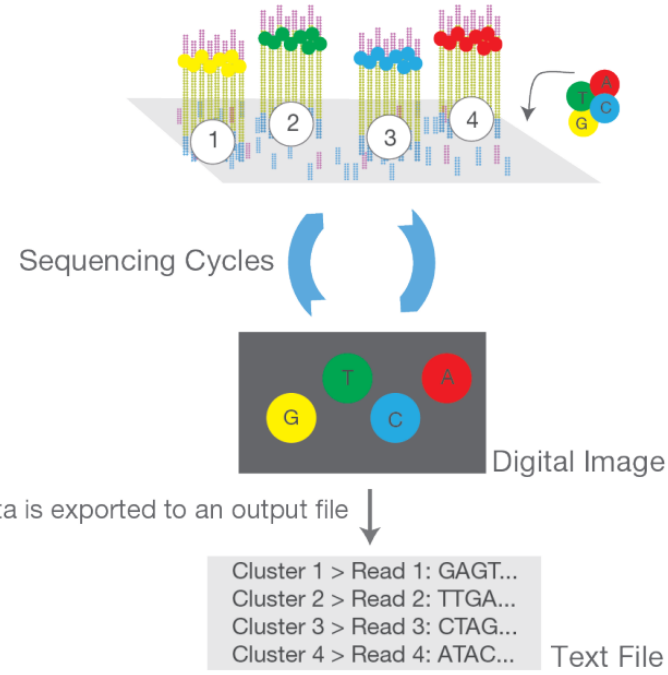


Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

https://youtu.be/fCd6B5HRaZ8

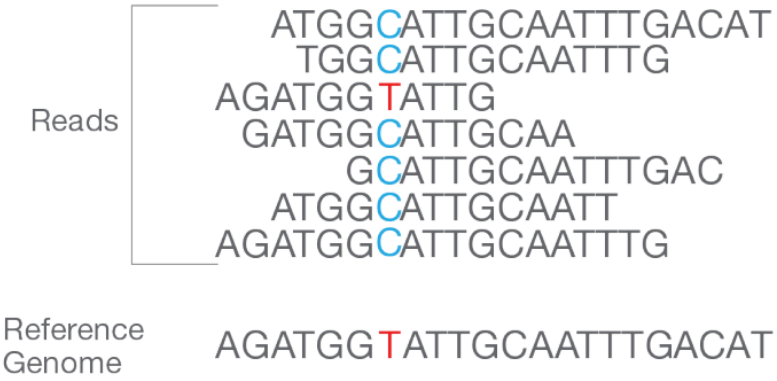
Illumina sequencing: Basic Principle

C. Sequencing



Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated "n" times to create a read length of "n" bases.

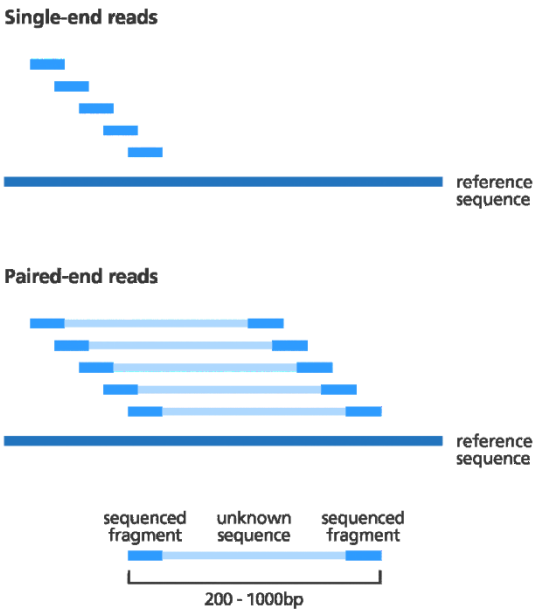
D. Alignment and Data Anaylsis

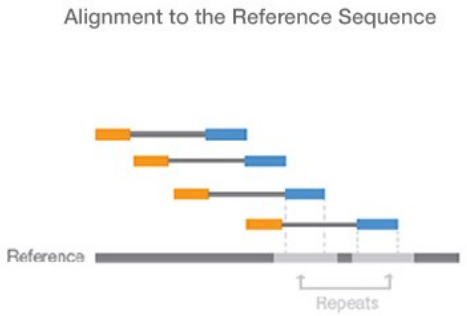


Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.

https://youtu.be/fCd6B5HRaZ8

Read structure & information: FASTQ file



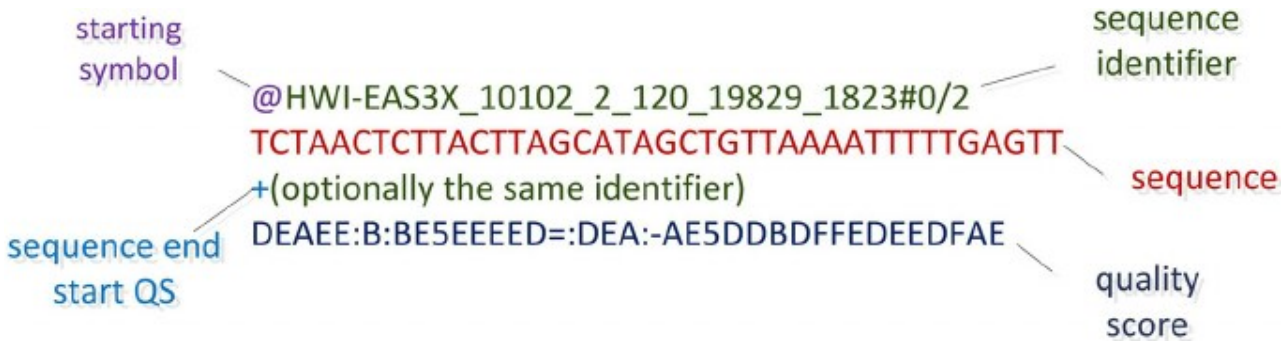


Short-read sequencing

Read length: 30-150 bp

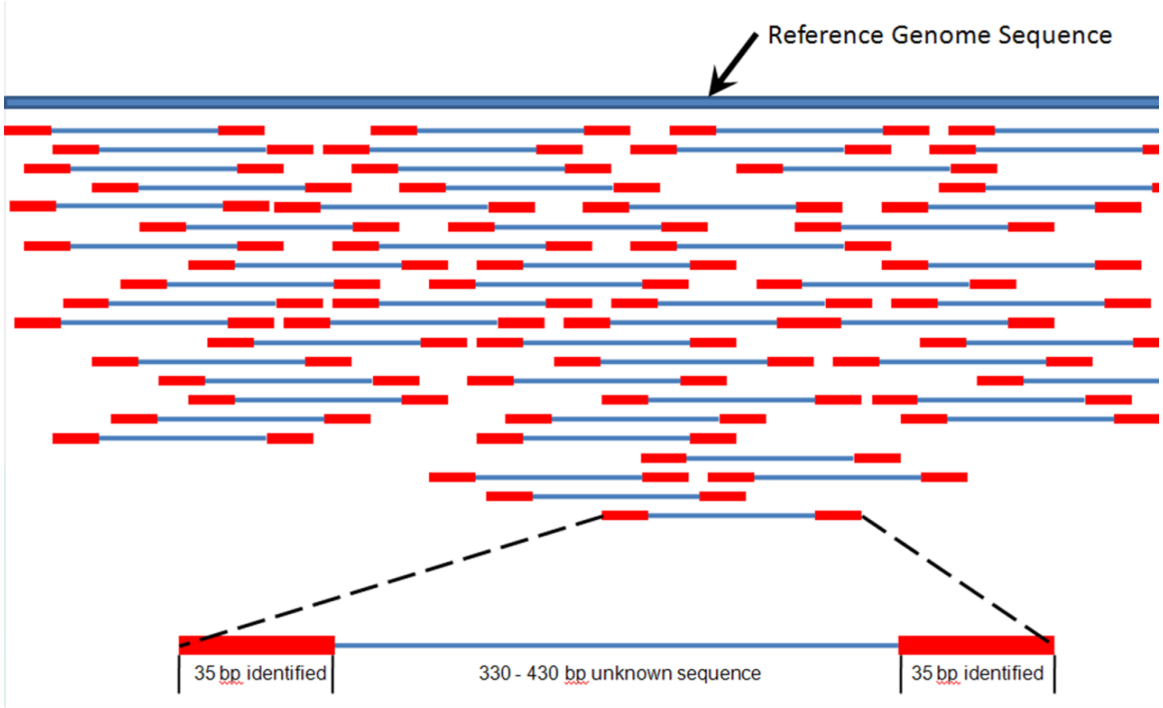
FASTQ file

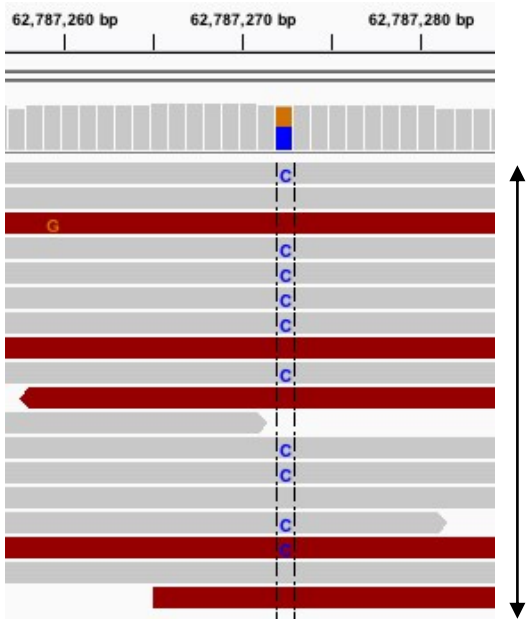
시퀀싱 장비로부터 생산되는 원시 데이터 (raw data)



Read Mapping to Reference Genome: BAM file

Read 정보에서 변이 정보를 재구성하기 위한 데이터 처리 과정





Depth of coverage = 17

GCTAGCTGATAGCTAGCTAGCTGATGAGCCCGA

Short Read (30-100 bp)

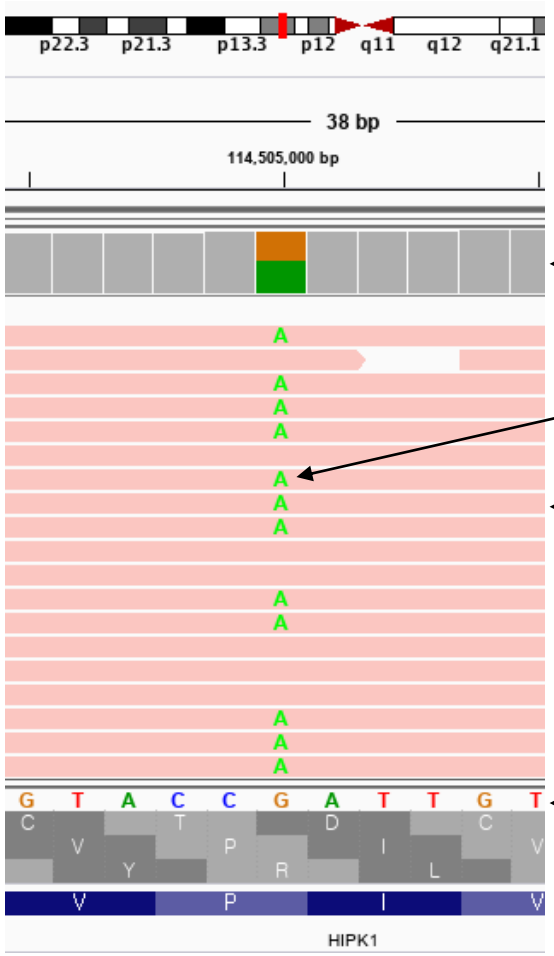
5’-ACTGGTCGATGCTAGCTGATAGCTAGCTAGCTGATGAGCCCGATCGCTGCTAGCTCGACG-3’

Reference Genome (3,000,000,000 bp)

Comparison with Reference Genome: VCF file

IGV: Bam file

정렬된 Read들로부터 변이 정보를 읽어 들이는 과정



Nucleotide distribution

Alternative sequence

Mapped reads

Reference sequence

Variant 기술의 핵심 정보 4가지

chr1:114,505,000 G A

① ② ③ ④

① Chromosome
② Position
③ Reference Allele
④ Alternative Allele

Adding interpretable information: Annotation

기본 Variant 정보로 부터, 사람이 해석 가능한 정보를 추가하는 과정 > **실제 임상 레포트에 포함된 정보**

[Variable annotation database]	**chr1:114,505,000 G A**
Cytoband	1p13.2
Gene name	HIPK1
Transcription ID	NM_198268.3
Exon, Intron information	Exon9
Coding sequence change	c.2043G>A
Amino acid change	p.Pro681=
Variant classification	Synonymous variant
Population allele frequency	25.95% (EAS)
Pathogenic report	Benign

Reference Genome (GRCh37/hg19, HGVS nomenclature)

실제 레포트 예시

유전성 암 패널
RiskCare cancer panel 검사보고서

병(의)원명	분당차병원	진료과/병동	GYC/W072/18	검체채취일	2018-10-06
수 진 자 명		의 사 명		검사의뢰일	2018-10-06
생 년 월 일		접 수 번 호	10-20181006-1069	결과보고일	2018-11-20
차 트 번 호		검 체 종 류	WB(EDTA)	기 타	A1810055892

Conclusion

Positive *BRCA2* 유전자에서 Pathogenic variant가 발견되었습니다.

Results

Identified pathogenic variants

HGVS nomenclature

Gene	Chromosomal position	Reference sequence	DNA change	Amino acid change	rs number	Zygosity	Classification
BRCA2	Chr13: 32930609	NM_000059.3	c.7480C>T	p.Arg2494Ter	rs80358972	Heterozygous	Pathogenic

Interpretation

상기 환자는 *BRCA2* 유전자에서 7480번째 염기인 C가 T로 치환되어 2494번째 아미노산인 Arginine이 종결 코돈으로 바뀌면서 premature termination되는 nonsense 변이가 관찰되었습니다. 이는 U43746 transcript의 7708C>T와 같은 변이입니다. 이 변이는 일반 인구집단에서는 거의 보고되지 않았지만 한국인과 핀란드인 유전성 유방암-난소암 증후군(HBOC) 환자에서 founder mutation으로 보고된 바 있는(PMID 9361038, 16455195, 19656164, 23199084, 22798144, 25863477) pathogenic variant입니다. 환자 및 환자 가족의 유전상담 진료가 필요합니다. 가족검사를 고려하실 경우 Family test, germline variants (Sequencing) 검사를 의뢰하시기 바랍니다.

Additional findings

Gene	Chromosomal position	Reference sequence	DNA change	Amino acid change	rs number	Zygosity	Classification
No identified variant							

유전성 암 유전자 패널 검사 결과 미분류변이(variant of uncertain significance)는 발견되지 않았습니다.

유전성 암 패널
RiskCare cancer panel 검사보고서

Test information

Test background

RiskCare 검사는 유전성 암의 위험도를 예측하기 위한 검사로서, 유방암, 난소암, 대장암 등을 비롯한 다양한 암종이 유전적인 원인에 의하여 발생한 것으로 의심되는 환자에서 시행을 고려할 수 있습니다. 특히 암이 젊은 나이에 발병하거나, 양쪽에 존재하는 장기에 각각 발생한 경우, 여러 장기에 발생한 경우, 부모 형제 자매가 같은 종류의 암이 진단된 경우 등이 유전성 암의 위험군에 해당합니다.
본 검사는 대상 유전자의 모든 coding exon 및 인접 intron (±10 bp) 부위를 차세대염기서열법으로 분석합니다. 유전성 암 증후군과 관계없이 산발적으로 암이 발생했거나, 본 검사에 포함되지 않은 유전자에 이상이 있는 경우에는 검사 결과 돌연변이가 검출되지 않을 수 있습니다. 본 검사 결과는 임상 정보와 함께 해석되어야 하고, 추가 검사가 필요할 수 있으므로 의사와의 상담이 필요합니다.

Test methods

ACMG classification

Reference Genome

접수된 검체에서 DNA를 추출하여 차세대염기서열분석법으로 대상 유전자의 염기서열을 분석하고, GRCh37/UCSC hg19를 기준으로 염기서열을 정렬하여 변이(variant)를 명명합니다. 발견된 변이는 2015 ACMG/AMP guideline (Genet Med 2015;17:405-24)에 근거하여 임상적 의미 (pathogenic, likely pathogenic, uncertain significance, likely benign, benign)를 판정하되, likely benign 및 benign으로 판정된 변이는 별도로 기술하지 않습니다. 변이의 분류와 해석은 결과 보고 시점의 관련 문헌과 public database이므로, 이러한 분류는 추후의 연구들을 통해 달라질 수 있습니다. 본 검사의 정도관리기준은 대상 유전자 영역에서 coverage 99.5% (minimum depth ≥20x)입니다. 검사 결과 pathogenic 또는 likely pathogenic variant가 발견된 경우 Sanger 염기서열 분석을 시행하여 확인합니다.

Analyzed genes

APC	*ATM*	*BARD1*	*BLM*	*BMPR1A*	*BRCA1*	*BRCA2*	*BRIP1*	*CDK4*	*CDH1*
CDKN2A	*CHEK2*	*EPCAM*	*MLH1*	*MRE11*	*MSH2*	*MSH6*	*MUTYH*	*NBN*	*PALB2*
PMS2	*PTEN*	*RAD50*	*RAD51C*	*RAD51D*	*SMAD4*	*STK11*	*TP53*	*VHL*	

Limitation

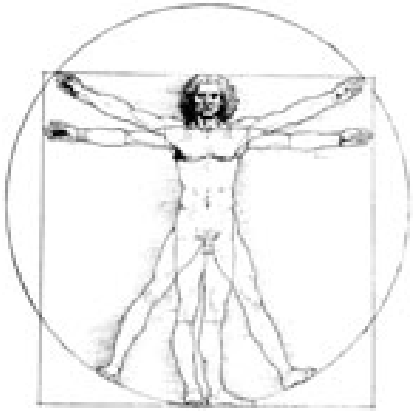
본 검사에서 deep intron에 위치하는 variant, sequence repeat 및 유전자 copy 수의 변화는 분석 대상에 포함하지 않으며, 드물게 분석 오류가 발생할 수 있습니다. 경미한 mosaicism이 존재하는 경우 본 검사의 생물정보분석 과정에서 검출되지 않을 수 있습니다. 대규모 결실/중복(large deletion/duplication)에 의한 돌연변이, 유전자재배열(rearrangement)은 이 검사법으로는 확인할 수 없습니다.

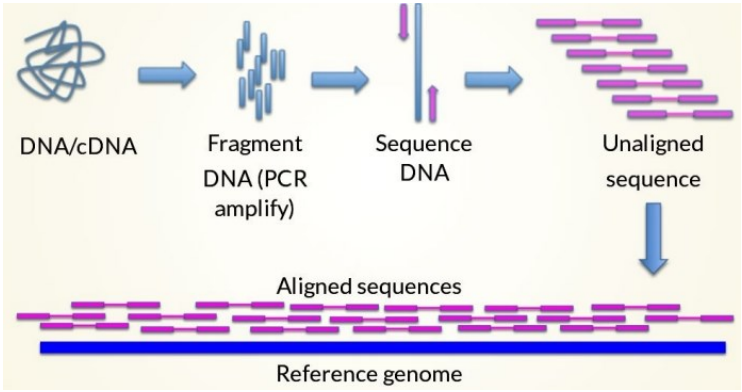
References

1. Database
Online Mendelian Inheritance in Man (OMIM®), The Human Gene Mutation Database (HGMD), Clinvar, ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles), BIC (Breast Cancer Information Core), 1000 Genomes, dbSNP, Exome aggregation consortium (ExAC), Exome Sequencing Project (ESP), Korean Reference Genome Database (KRGDB)

2. 문헌
Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics. Genet Med. 2015 May;17(5):405-24.
Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2017. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1116/

Reference Genome





Fragmented reads들로부터 variant를 불러들이는데 기본 frame이 되는 중요한 정보

Human

The human genome assembly was produced as part of the Human Genome Project (HGP). The previous assembly (NCBI36) was the last one produced by the HGP and was described in 2004 (PMID: 15496913); this was the starting point for the GRC. The assembly is based largely on assembling overlapping clone sequences.

Human assembly information

Current major assembly	GRCh38
Regions with alternate loci	178
Assembly N50	67,794,873 bp
Remaining gaps	875
Patch release version	p13
Patches released	FIX: 113, NOVEL: 72

More human assembly statistics...

***De novo assembly**

Human Reference

GRCh38 (latest version)
Genome Reference Consortium human version38

GRCh37/UCSC hg19를 기준으로
Med 2015;17:405-24)에 근거하여
하되, likely benign 및 benign으로
lic database이므로, 이러한 분류는
overage 99.5% (minimum depth
을 시행하여 확인합니다.

GRCh37
GRC human version37

UCSC hg19
UCSC human genome version19

Genome Reference Consortium https://www.ncbi.nlm.nih.gov/grc/

Reference Genome

Reference Genome이 왜 중요한가? ⟶ 엉뚱한 영역의 Genome을 확인할 수 있음.

Identified pathogenic variants

Gene	Chromosomal position	Reference sequence	DNA change	Amino acid change	rs number	Zygosity	Classification
BRCA2	Chr13: 32930609	NM_000059.3	c.7480C>T	p.Arg2494Ter	rs80358972	Heterozygous	Pathogenic

chr13:32930609

GRCh37 BRCA2 NM_000059.3: c.7480 C>T, p.R2494X

GRCh38 Intergenic 영역, Reference allele: G

유전자의 위치는 절대적인 것이 아님.
Reference Genome을 기준으로 기술된 상대적 위치 > Version에 따라 Position이 달라짐.

HGVS nomenclature

Human Genome Variation Society에서 제안한 변이 기술의 공통 약속

Identified pathogenic variants

Gene	Chromosomal position	Reference sequence	DNA change	Amino acid change	rs number	Zygosity	Classification
BRCA2	Chr13: 32930609	NM_000059.3	c.7480C>T	p.Arg2494Ter	rs80358972	Heterozygous	Pathogenic

Number mRNA (NM) > Transcription ID
"**c.**" for a coding DNA reference sequence
"**p.**" for a protein reference sequence
rs number: reference SNP number (dbSNP)

rs80358972

Current Build 154
Released April 21, 2020

Organism	*Homo sapiens*	**Clinical Significance**	Reported in ClinVar
Position	chr13:32356472 (GRCh38.p12)	**Gene : Consequence**	BRCA2 : Stop Gained
Alleles	C>A / C>T	**Publications**	14 citations LitVar 36
Variation Type	SNV Single Nucleotide Variation		
Frequency	T=0.000016 (2/125568, TOPMED) T=0.00001 (1/78700, PAGE_STUDY) T=0.0001 (1/8762, ALFA Project) (+ 2 more)	**Genomic View**	See rs on genome

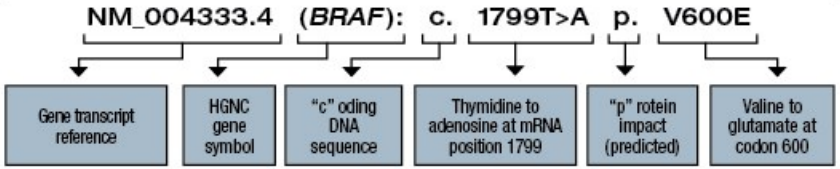
Gene: BRCA2, BRCA2 DNA repair associated (plus strand)

Molecule type	Change	Amino acid[Codon]	SO Term
BRCA2 transcript	NM_000059.4:c.7480C>A	R [CGA] > R [AGA]	Coding Sequence Variant
BRCA2 transcript	NM_000059.4:c.7480C>T	R [CGA] > * [TGA]	Coding Sequence Variant
breast cancer type 2 susceptibility protein	NP_000050.3:p.Arg2494=	R (Arg) > R (Arg)	Synonymous Variant
breast cancer type 2 susceptibility protein	NP_000050.3:p.Arg2494Ter	R (Arg) > * (Ter)	Stop Gained

Reporting mutations

- Standardized nomenclature to promote portability, enduring meaning, and accuracy
- Human Genome Variation Society (HGVS): www.hgvs.org/mutnomen/

***BRAF* mutation analysis:**
Mutation detected in codon 600, exon 15 (GTG to GAG) of the *BRAF* gene that would change the encoded amino acid from valine to glutamate (p.Val600Glu)



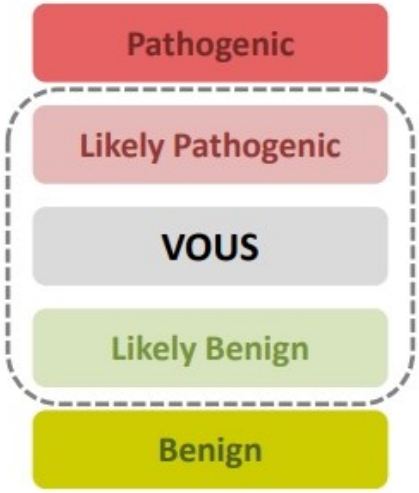
HGVS - http://varnomen.hgvs.org/recommendations/general/

dbSNP - https://www.ncbi.nlm.nih.gov/snp/

ACMG classification

American College of Medical Genetics에서 제안한 변이 해석의 기준

5-Tier system



Identified pathogenic variants

Gene	Chromosomal position	Reference sequence	DNA change	Amino acid change	rs number	Zygosity	Classification
BRCA2	Chr13: 32930609	NM_000059.3	c.7480C>T	p.Arg2494Ter	rs80358972	Heterozygous	Pathogenic

ACMG Classification - Educational use only Version: 8.3.0 Terms of use Documentation Options

Verdict
Pathogenic

Transcript NM_000059.3, canonical, protein length 3419, gene BRCA2, nonsense variant

Rules

☑ PVS1 Very Strong	☒ PS1	☐ PS2	☒ PS3	☐ PS4	☒ PM1	☑ PM2 Moderate	☐ PM3
☒ PM4	☒ PM5	☐ PM6	☐ PP1	☐ PP2	☑ PP3 Supporting	☐ PP4	☑ PP5 Very Strong
☒ BA1	☒ BS1	☒ BS2	☒ BS3	☒ BS4			
☒ BP1	☐ BP2	☒ BP3	☒ BP4	☐ BP5	☒ BP6	☒ BP7	

The verdict will update automatically if you enable or disable rules or change their strength. The blue question marks displays details about the rule, including why it was not triggered.

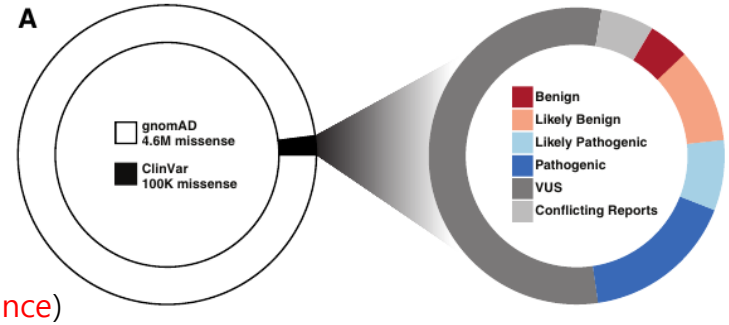
Automated criteria

Rule	Explanation
PP5 Very Strong	ClinVar classifies this variant as Pathogenic, rated 3 stars, reviewed by expert panel. Using strength Very Strong because of the evidence presented by ClinVar.
PVS1 Very Strong	Null variant (nonsense) affecting gene BRCA2, which is a known mechanism of disease (gene has 3,724 known pathogenic variants which is greater than minimum of 3), associated with Breast-ovarian cancer, familial, susceptibility to, Pancreatic cancer, susceptibility to, 2, Glioma susceptibility 3, Fanconi anemia, complementation group D1, Wilms tumor and Medulloblastoma.
PM2 Moderate	GnomAD exomes homozygous allele count = 0 is less than 3 threshold for recessive gene BRCA2 (good gnomAD exomes coverage = 92.9). Variant not found in gnomAD genomes (good gnomAD genomes coverage = 31.0).
PP3 Supporting	Pathogenic computational verdict based on 4 pathogenic predictions from DANN, EIGEN, FATHMM-MKL and MutationTaster vs no benign predictions.

https://varsome.com/variant/hg19/chr13-32930609-C-T

ACMG classification

American College of Medical Genetics에서 제안한 변이 해석의 기준



[ACMG classification 주요 고려 사항]

대부분의 변이(>50%)는 **VUS**로 구분됨. (Variant of Uncertain Significance)

- 대부분의 변이의 의미를 판독하기가 어려움.
- VUS로 분류된 변이 중 일부는 Pathogenic 또는 Benign 일수도 있음.

Genet Med 2015; 17(5):405-423.
Am J Hum Genet 2017;101(3):315-325.

[Pathogenic을 Supporting 하는 Finding]

Variant Class: Stop-gain, Frameshift, Splicing variant
Extremely rare population frequency (정상 인구 집단에서 거의 존재하지 않는 변이)
Variant occurred in highly conserved regions (종간 보존이 잘된 위치의 변이)

De novo variant (부모에게 관찰되지 않는 변이) > 가족 검사 필요
Defective in vitro Functional Assay > 변이 기능 측정 검사 필요

[Benign을 Supporting 하는 Finding]

Variant Class: Synonymous variant
Common variant: Allele frequency > 1% (정상 인구 집단에서 흔하게 존재하는 변이)
Variant occurred in variable regions

Annotation Database 활용

변이 정보에 Annotation Database를 이용하여, 원하는 정보를 추가, 다양한 Bioinformatics tools 이용

[Gene annotation DB] : Gene name, HGVS annotation data
refGene, knownGene, ensGene, UCSC_gene

[dbSNP DB] : rs number annotation data
avsnp153

[Population frequency DB] : 정상인 인구 집단에서 해당 변이의 빈도 정보
EXAC, gnomAD, esp6500, 1000G, KRGDB

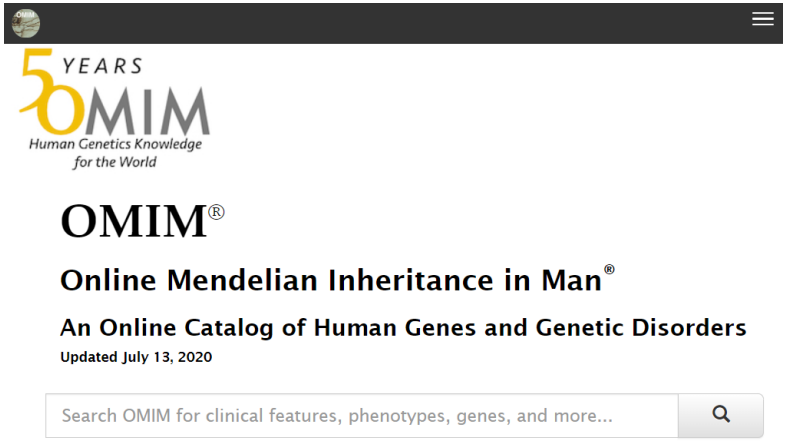
[in-silico prediction tools] : missense 변이의 기능 변화 예측 tool > score
SIFT, Polyphen-2, CADD, MutationTaster, FATHMM, DANN, GERP++, PhyloP

[Clinical information DB] : 실제 환자 case report database
ClinVar, InterVar, COSMIC, HGMD

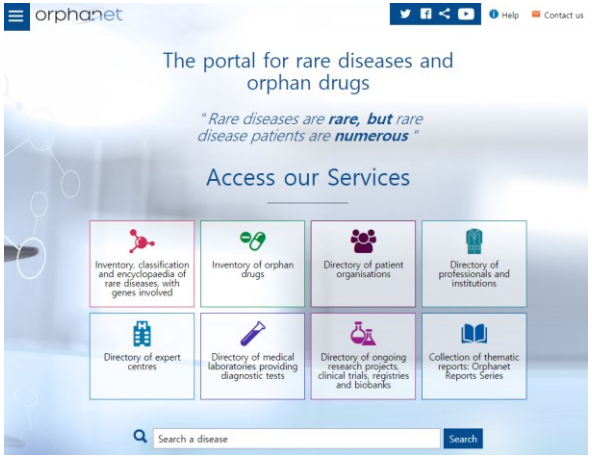
ANNOVAR: https://annovar.openbioinformatics.org/en/latest/

Genotype-phenotype correlation

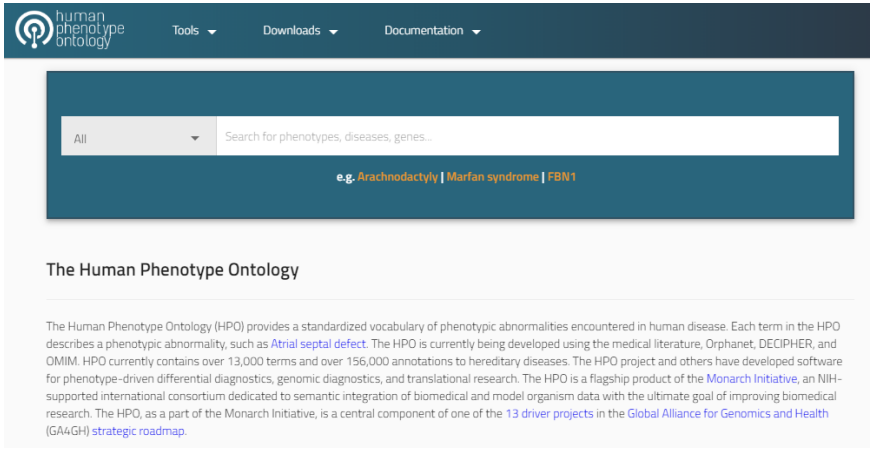
https://www.omim.org/



https://www.orpha.net/



https://hpo.jax.org/app/



https://decipher.sanger.ac.uk/

Genotype-phenotype correlation: OMIM

* 600185

BRCA2 GENE; BRCA2

Alternative titles; symbols

FANCD1 GENE; FANCD1

HGNC Approved Gene Symbol: BRCA2

Cytogenetic location: 13q13.1 Genomic coordinates (GRCh38): 13:32,315,507-32,400,267 (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
13q13.1	{Breast cancer, male, susceptibility to}	114480	AD, SMu	3
	{Breast-ovarian cancer, familial, 2}	612555	AD	3
	{Glioblastoma 3}	613029	AR	3
	{Medulloblastoma}	155255	AD, AR, SMu	3
	{Pancreatic cancer 2}	613347		3
	{Prostate cancer}	176807	AD, SMu	3
	Fanconi anemia, complementation group D1	605724	AR	3
	Wilms tumor	194070	AD, SMu	3

Other entities represented in this entry:

BREAST CANCER, FAMILIAL, SUSCEPTIBILITY TO, 2, INCLUDED
OVARIAN CANCER, FAMILIAL, SUSCEPTIBILITY TO, 2, INCLUDED

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene, MIM numb
13q13.1	{Breast-ovarian cancer, familial, 2}	612555	AD	3	BRCA2	60018

Clinical Synopsis | Phenotypic Series | PheneGene Graphics

- ▶ TEXT
- ▶ Clinical Features
- ▶ Inheritance
- ▶ Mapping
- ▶ Pathogenesis
- ▶ Clinical Management
- ▶ Molecular Genetics
- ▶ Population Genetics
- ▼ See Also:

Breast Cancer Linkage Consortium (1999)

Further Resources

- Further resources
  - Illumina sequencing 원리: https://youtu.be/fCd6B5HRaZ8
  - Genome Reference Consortium: https://www.ncbi.nlm.nih.gov/grc/
  - HGVS nomenclature: http://varnomen.hgvs.org/recommendations/general/
  - dbSNP: https://www.ncbi.nlm.nih.gov/snp/
  - ANNOVAR: https://annovar.openbioinformatics.org/en/latest/
  - OMIM: https://www.omim.org
  - Orphanet: https://www.orpha.net/
  - Human Phenotype Ontology: https://hpo.jax.org/app/
  - DECIPHER: https://decipher.sanger.ac.uk/
  - *Hum Mut 2016; 37(6):564-569.*
  - *Genet Med 2015; 17(5):405-423.*

- 개인 블로그: 두마디 정밀의료 (Towards Precision Medicine)
  https://2wordspm.com/
- E-mail: joonji422@yuhs.ac