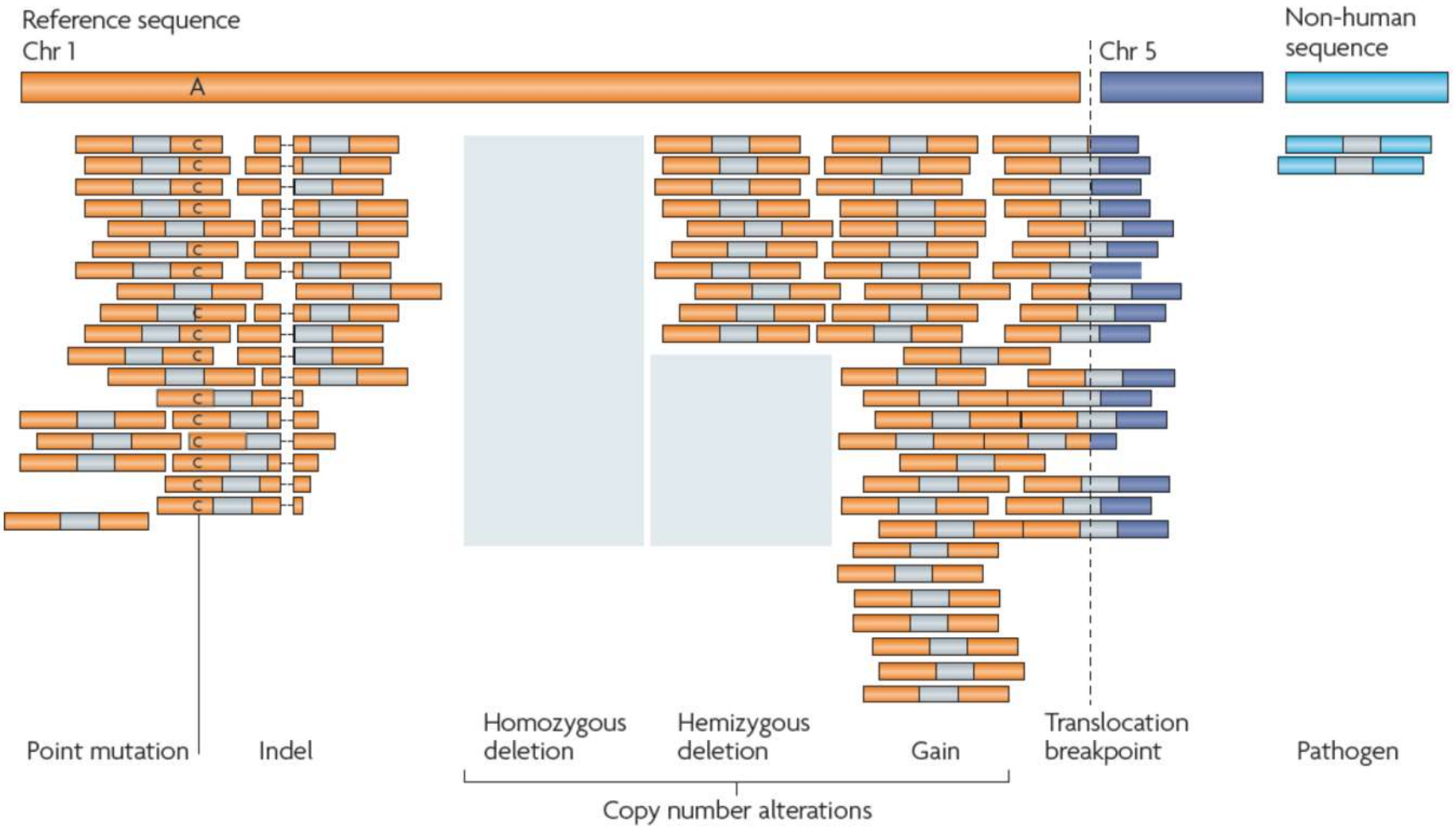
가톨릭대학교
바이오 데이터 엔지니어 양성사업

Practice of targeted sequencing analysis

(Mutation and Copy number alternations)

Summary


Reference sequence
Chr 1
A
Chr 5
Non-human sequence
Point mutation
Indel
Homozygous deletion
Hemizygous deletion
Gain
Translocation breakpoint
Pathogen
Copy number alterations

Terms

Library: A collection of DNA fragments with adapters ligated to each end (reads)

FASTQ: format used for storing both a nucleotide sequence and its corresponding quality scores

Quality score (Q-score): Quality scores measure the probability that a base is called incorrectly, Higher quality score indicates a smaller probability of error

Q20/Q30 Bases (%): percenatge of bases withquality score more than 20/30

Coverage (= read depth): The average number of sequencing reads that align to each base within the sample DNA to the target region

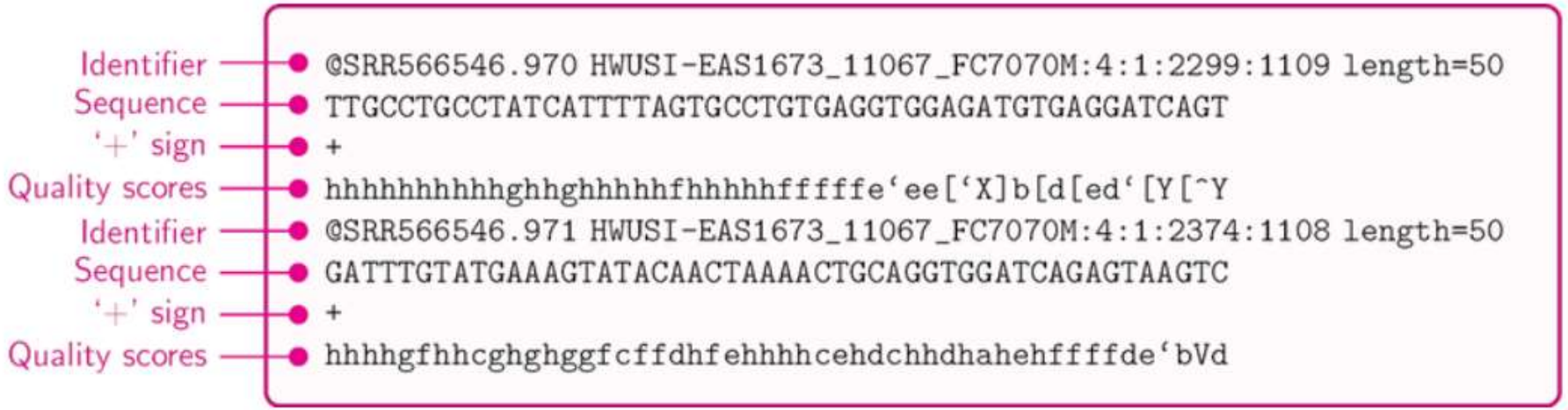
Coverage uniformity: A measurement of the evenness of the coverage depth of target (another metric: proportion of reads **covered by more than 20 reads**)

Mapped read: The total number of bases sequenced and aligned (Mapped %: rate)

Mapping quality: quality score for mapping

FASTQ

```
@A00125:23:H3CYNDMXX:1:1101:2049:1000 1:N:0:CGGCTATG+AGGCGAAG
CTCAGACCGCGTTCTCTCCCTCTCACTCCCCAATACGGAGAGAAGAACGATCATCAATGGCTGACGGCAGTTGCGGCGAAGCAACGCCAAAAAGCCAGCTT
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00125:23:H3CYNDMXX:1:1101:2700:1000 1:N:0:CGGCTATG+AGGCGAAG
CCTCGTCCGACTCCTCCGACTCTTCCTTGGCTTCAACCTTAGCTGGGGCTGCAGCAGCAGCAGGAGCAGCTGTGGTGGCAGCAGCCACAGGGGCAGCAGCC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00125:23:H3CYNDMXX:1:1101:10908:1000 1:N:0:CGGCTATG+AGGCGAAG
GTCCACTCTCGACTGCCGGCGACGGCCGGGTATGGGCCCGACGCTCCAGCGCCATCCATTTTCAGGGCTAGTTGATTCGGCAGGTGAGTTGTTACACACTC
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00125:23:H3CYNDMXX:1:1101:11342:1000 1:N:0:CGGCTATG+AGGCGAAG
GGGCTGGCAGGAGTCAGGCTGTTCAAGACAACTGGAAGGAGTTGAATAACATCTATCCAGTGAGTCCTGCAAGACTTCAGGCTCTACTACCTCCAGCAGCT
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
```



FASTQ

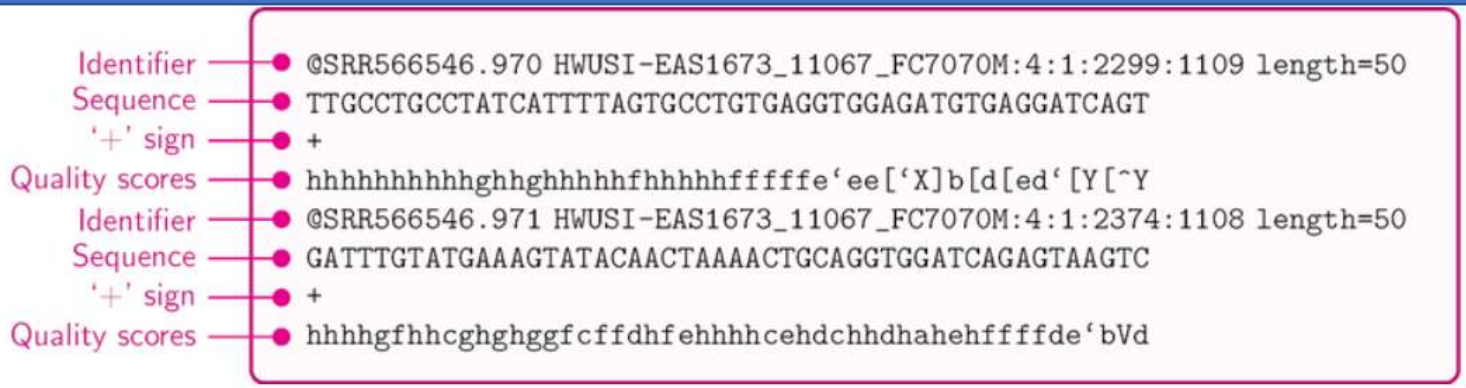


Table 1 ASCII Characters Encoding Q-scores 0–40

Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score
!	33	0	/	47	14	=	61	28
"	34	1	0	48	15	>	62	29
#	35	2	1	49	16	?	63	30
$	36	3	2	50	17	@	64	31
%	37	4	3	51	18	A	65	32
&	38	5	4	52	19	B	66	33
'	39	6	5	53	20	C	67	34
(	40	7	6	54	21	D	68	35
)	41	8	7	55	22	E	69	36
*	42	9	8	56	23	F	70	37
+	43	10	9	57	24	G	71	38
,	44	11	:	58	25	H	72	39
-	45	12	;	59	26	I	73	40
.	46	13	<	60	27			

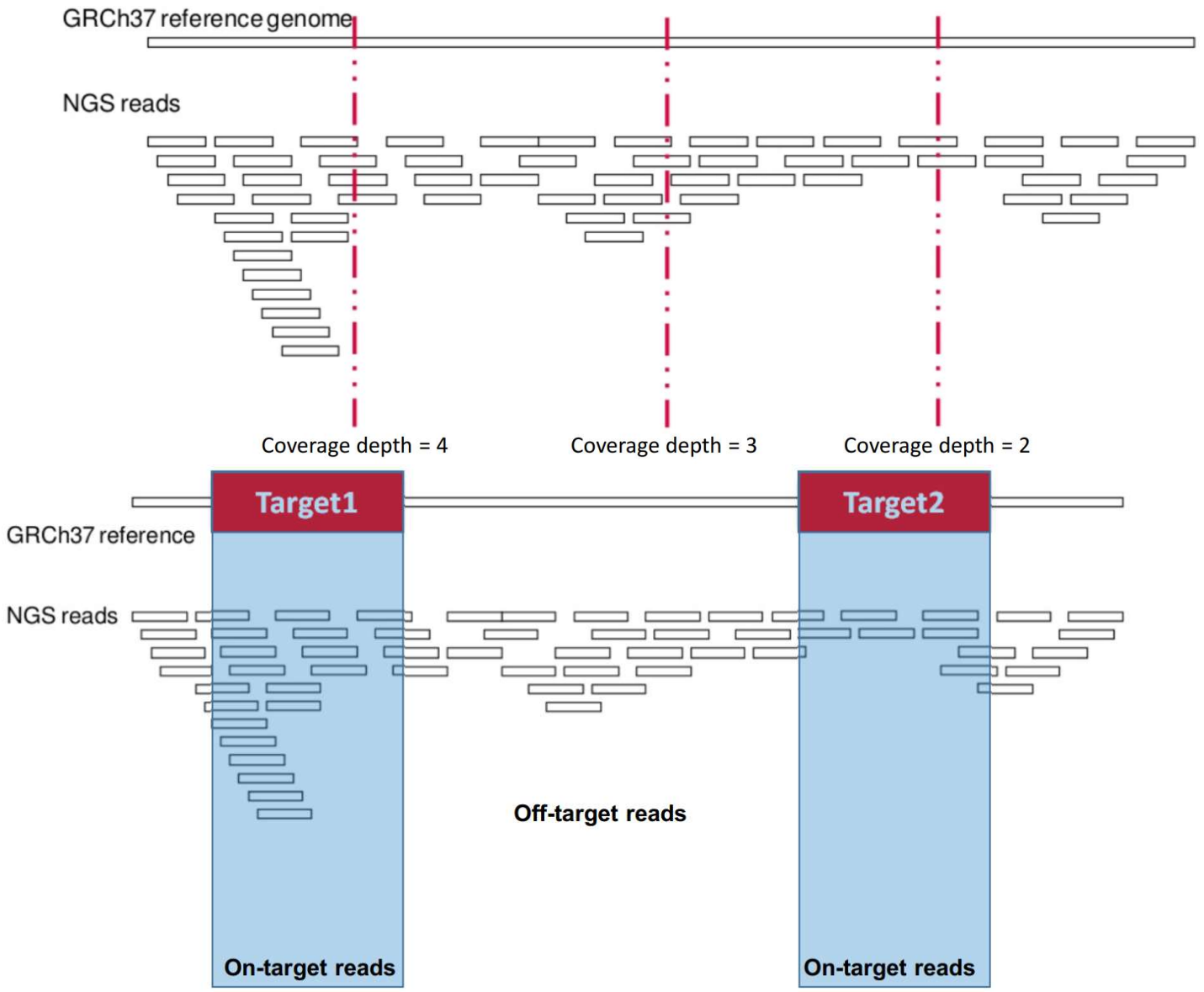
Q10 = 90%

Q20 = 99%

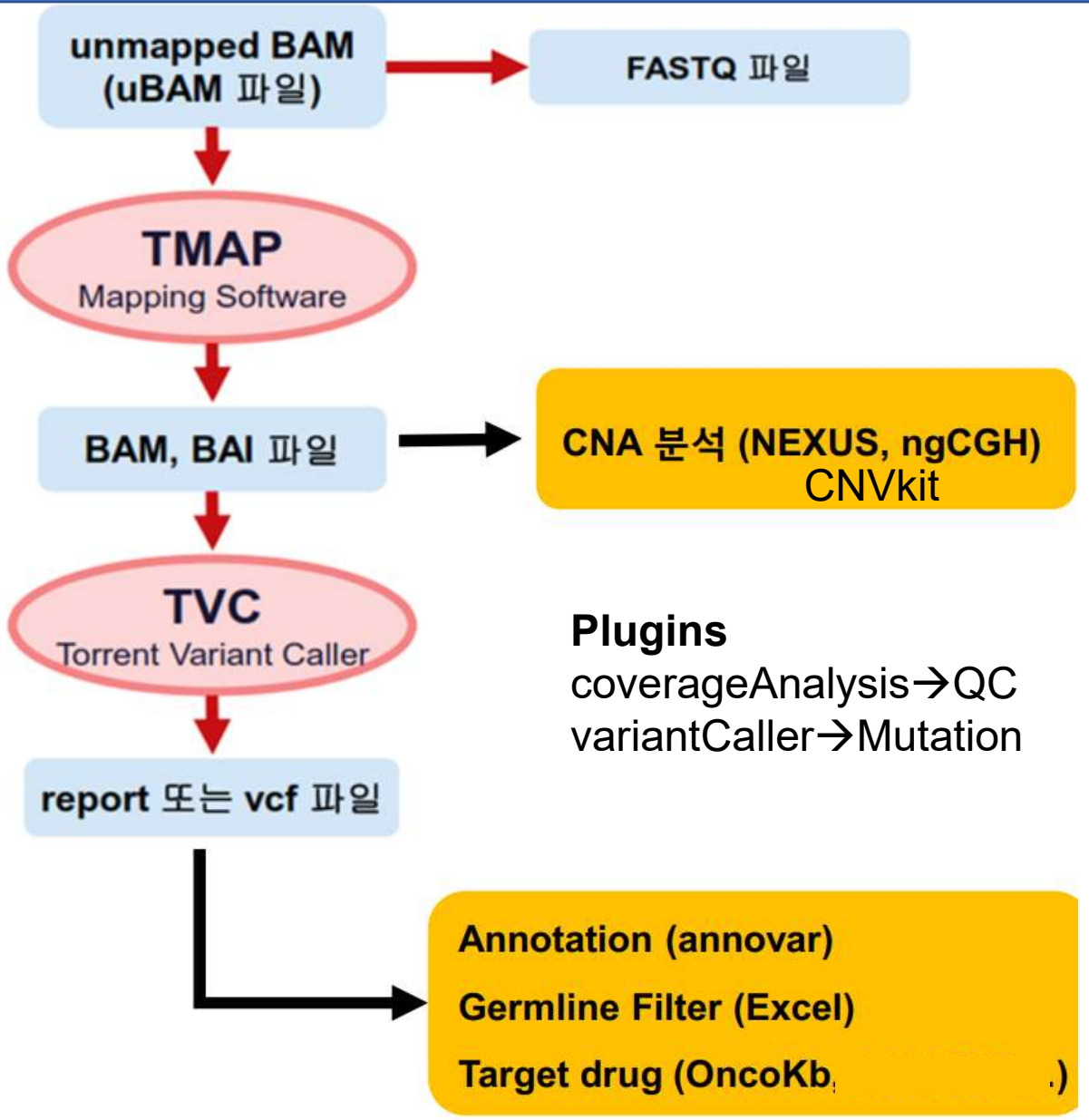
Q30 = 99.9%

Q40 = 99.99%

Read depth, on-target rate



Ion torrent system



Chip Type: 530

Library Type: AmpliSeq

Variant Frequency: Somatic

AmpliSeq Panel: Unspecified — Add panel...

Reference Genome: hg19 - Homo sapiens

Targeted Regions: WG_IAD119462.20170404.designed — Add targets...

Hotspot Regions: CCP.20131001.hotspots — Add hotspots...

Parameter Settings:

- (●) Generic - S5/S5XL (510/520/530) - Somatic - Low Stringency
  somatic_low_stringency_520_530, TS version: 5.6
- ( ) Custom
  custom, TS version: 5.6

Load external parameter file

Hide Advanced Settings ▲

Parameter	INDEL	SNP	MNP	Hotspot
Minimum allele frequency min_allele_freq	0.02	0.0185	0.02	0.02
Minimum quality min_variant_score	6	6	15	6
Minimum coverage min_coverage	100	100	100	100
Minimum coverage on either strand min_cov_each_strand	200	100	20	4
Maximum strand bias strand_bias	0.69	0.95	0.95	0.95
Minimum relative read quality data_quality_stringency	10			
Maximum common signal shift filter_unusual_predictions	0.12			

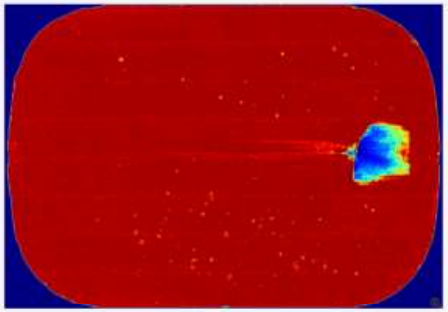
Summary

Run Name user S5-00284-45-CMC melanoma 02	Run Date Dec 20 2018	Projects
Report Name Auto_user_S5-00284-45-CMC_melanoma_02_204	Report Date Dec 21 2018	Notes

Unaligned Reads

2.22 G — Total Bases
138 — Key Signal
93 % ISP Loading

ISP Density

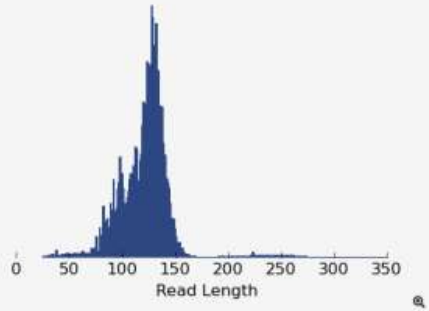


18,316,578 — Total Reads
53 % Usable Reads

ISP Summary

121 bp — Mean
124 bp — Median
127 bp — Mode

Read Length

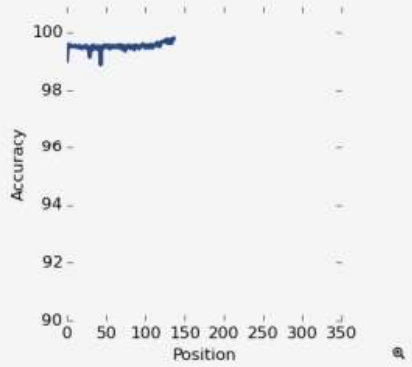


Aligned Reads hg19(Homo sapiens)

2.06 G — Total Aligned Bases
0.7X — Reference Coverage

	Count	%
Total Reads	17,551,449	—
Aligned Reads	17,517,055	99.8%
Unaligned Reads	34,394	0.2%

99.5% — Mean Raw Accuracy 1x



1.98 G — AQ17 Total Bases

Alignment Quality

	AQ17	AQ20	Perfect
Total Number of Bases [bp]	1.98 G	1.89 G	1.69 G
Mean Length [bp]	115	113	104
Longest Alignment [bp]	333	330	326
Mean Coverage Depth [x]	0.6	0.6	0.5

Barcode Name	Sample	Bases	>=Q20 Bases	Reads	Mean Read Length	Read Length Histogram	Files
No barcode	None	114,954,488	106,652,847	868,107	132 bp		UBAM BAM BAI
IonXpress_049	MIS 2N	183,081,985	170,865,582	1,501,716	121 bp		UBAM BAM BAI
IonXpress_050	MIS 2T	236,630,008	221,112,089	1,944,960	121 bp		UBAM BAM BAI
IonXpress_051	MIS 8N	210,229,008	196,316,760	1,775,009	118 bp		UBAM BAM BAI
IonXpress_052	MIS 8T	177,032,606	164,363,676	1,500,380	117 bp		UBAM BAM BAI
IonXpress_053	none	196,771,749	183,322,645	1,678,487	117 bp		UBAM BAM BAI
IonXpress_054	MM 1T	239,666,633	222,636,477	2,000,637	119 bp		UBAM BAM BAI
IonXpress_055	MM 2N	201,283,875	188,107,462	1,704,706	118 bp		UBAM BAM BAI
IonXpress_056	MM 2T	209,749,403	196,572,727	1,770,737	118 bp		UBAM BAM BAI
IonXpress_057	MM 3N	139,145,455	129,611,282	1,182,958	117 bp		UBAM BAM BAI

1 2 10 items per page 1 - 10 of 14 items

Barcode Name	Sample	Bases	>=Q20 Bases	Reads	Mean Read Length	Read Length Histogram	Files
IonXpress_058	MM 3T	163,207,303	152,640,087	1,392,422	117 bp		UBAM BAM BAI
IonXpress_062	1_calu-3	59,314,281	55,331,819	495,473	119 bp		UBAM BAM BAI
IonXpress_063	2_FTC-133	51,391,001	47,861,432	421,815	121 bp		UBAM BAM BAI
IonXpress_064	3_ML-1	63,349,689	59,239,939	515,198	122 bp		UBAM BAM BAI

1 2 10 items per page 11 - 14 of 14 items

coverageAnalysis v5.6.0.1 (262)

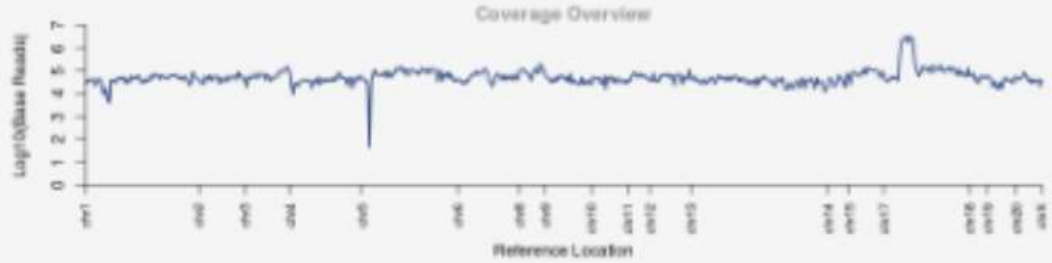
Completed 42.2 MB

•coverageAnalysis.html

Library type: AmpliSeq DNA
Target regions: WG_IAD119462.20170404.designed

Barcode Name	Sample	Mapped Reads	On Target	Mean Depth	Uniformity
IonXpress_049	MIS 2N	1,498,454	98.31%	1,215	96.34%
IonXpress_050	MIS 2T	1,941,425	98.47%	1,586	95.93%
IonXpress_051	MIS 8N	1,771,277	97.57%	1,421	96.31%
IonXpress_052	MIS 8T	1,496,229	97.00%	1,195	96.28%
IonXpress_053	None	1,674,744	97.27%	1,332	96.62%
IonXpress_054	MM 1T	1,996,273	97.98%	1,614	95.66%
IonXpress_055	MM 2N	1,701,453	97.30%	1,354	96.42%
IonXpress_056	MM 2T	1,765,942	95.92%	1,397	96.49%
IonXpress_057	MM 3N	1,179,391	96.53%	929.6	96.64%
IonXpress_058	MM 3T	1,389,029	97.46%	1,106	96.57%
IonXpress_062	1_calu-3	494,352	98.63%	402.4	90.97%
IonXpress_063	2_FTC-133	421,012	94.55%	337	96.88%
IonXpress_064	3_ML-1	512,519	98.09%	428.7	96.92%

Library type: AmpliSeq DNA
Reference: hg19 (DNA)
Target regions: WG_IAD119462.20170404.designed



Number of mapped reads	494,352
Percent reads on target	98.63%
Average base coverage depth	402.4
Uniformity of base coverage	90.97%

Amplicon Read Coverage

Number of amplicons	1,366
Percent assigned amplicon reads	98.63%
Average reads per amplicon	356.9
Uniformity of amplicon coverage	92.48%
Amplicons with at least 1 read	99.85%
Amplicons with at least 20 reads	98.24%
Amplicons with at least 100 reads	85.58%
Amplicons with at least 500 reads	4.39%
Amplicons with no strand bias	96.19%
Amplicons reading end-to-end	96.71%
Amplicon base composition bias	0.353

Target Base Coverage

Bases in target regions	136,918
Percent base reads on target	94.52%
Average base coverage depth	402.4
Uniformity of base coverage	90.97%
Target base coverage at 1x	99.86%
Target base coverage at 20x	98.48%
Target base coverage at 100x	87.22%
Target base coverage at 500x	9.27%
Target bases with no strand bias	95.91%
Percent end-to-end reads	93.17%

variantCaller v5.6.0.4 (263)

Completed 8.90 GB

•variantCaller.html

Library type:	AmpliSeq
Reference genome:	hg19
Targeted regions:	WG_IAD119462.20170404.designed
Hotspot regions:	none
Configuration:	Generic - S5/S5XL (510/520/530) - Somatic - Low Stringency
Output directory:	variantCaller_out.263
Download all barcodes:	VCF.ZIP XLS.ZIP XLS COV
Please note:	Variant calling was carried out for all barcodes with reference genome as specified above

Barcode Name	Sample	Variants	Downloads
IonXpress_049	MIS 2N	99	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_050	MIS 2T	99	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_051	MIS 8N	101	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_052	MIS 8T	104	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_053	none	110	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_054	MM 1T	107	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_055	MM 2N	105	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_056	MM 2T	104	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_057	MM 3N	116	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_058	MM 3T	114	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_062	1_calu-3	36	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_063	2_FTC-133	74	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS
IonXpress_064	3_ML-1	67	VCF.GZ VCF.GZ.TBI gVCF.GZ gVCF.GZ.TBI XLS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Chr	Start	End	Ref	Variant	Frequency	Quality	Original Coverage	Coverage	Coverage+	Coverage-	Allele Cov	Allele Cov+	Allele Cov-	Strand Bias	Sample Name
2	chr3	37053568	37053568	A	G	35	1055.56	317	317	182	135	111	66	45	0.521	1_calu-3
3	chr5	112162854	112162854	T	C	100	5168.6	303	303	154	149	303	154	149	0.5	1_calu-3
4	chr5	112175770	112175770	G	A	99.6	8866.01	520	521	240	281	519	239	280	0.5002	1_calu-3
5	chr5	112176325	112176325	G	A	99	8051.83	478	478	196	282	473	196	277	0.5045	1_calu-3
6	chr5	112176559	112176559	T	G	99.8	7266.3	432	432	200	232	431	200	231	0.5011	1_calu-3
7	chr5	112176756	112176756	T	A	100	6946.4	406	406	253	153	406	253	153	0.5	1_calu-3
8	chr5	112177171	112177171	G	A	100	12061.9	705	707	375	332	707	375	332	0.5	1_calu-3
9	chr5	149457678	149457678	G	A	20.6	707.711	485	485	126	359	100	20	80	0.584	1_calu-3
10	chr5	176517797	176517797	C	T	99.5	7272.58	434	435	213	222	433	211	222	0.5024	1_calu-3
11	chr5	176520243	176520243	G	A	19	433.98	346	347	178	169	66	32	34	0.5281	1_calu-3
12	chr7	6441891	6441891	T	C	13	231.323	347	347	148	199	45	24	21	0.6057	1_calu-3
13	chr7	55249063	55249063	G	A	99.4	8292.53	493	496	236	260	493	234	259	0.5012	1_calu-3
14	chr7	128846328	128846328	G	C	99.6	4629.59	279	279	172	107	278	171	107	0.5015	1_calu-3
15	chr8	38285914	38285916	TCA	-	2.2	6.01272	417	417	169	248	9	4	5	0.5399	1_calu-3
16	chr8	128748744	128748744	C	T	80.9	4652.23	397	398	184	214	322	154	168	0.516	1_calu-3
17	chr9	5557672	5557672	T	C	99.6	4175.23	247	248	133	115	247	133	114	0.5022	1_calu-3
18	chr11	62609160	62609160	A	T	2.6	10.7706	540	540	262	278	14	7	7	0.5148	1_calu-3
19	chr11	62609171	62609174	TATT	CTGC	58.7	2557.57	367	368	158	210	216	92	124	0.5035	1_calu-3
20	chr11	69462910	69462910	G	A	67.1	4032.7	459	459	248	211	308	167	141	0.5019	1_calu-3
21	chr13	32913055	32913055	A	G	99.7	5542.08	331	332	163	169	331	162	169	0.5015	1_calu-3
22	chr14	20811577	20811577	A	G	99.5	6416.22	383	385	229	156	383	228	155	0.5005	1_calu-3
23	chr14	20811588	20811588	T	C	70.1	3523.4	379	381	227	154	267	169	98	0.5392	1_calu-3
24	chr14	105253009	105253009	A	G	100	4813.66	286	286	144	142	286	144	142	0.5	1_calu-3
25	chr16	68857289	68857289	T	C	27.1	492.711	229	229	106	123	62	26	36	0.544	1_calu-3
26	chr16	68857441	68857441	T	C	100	10100.6	585	585	337	248	585	337	248	0.5	1_calu-3
27	chr17	7579472	7579472	G	C	98.3	4648.48	281	287	120	167	282	119	163	0.504	1_calu-3
28	chr17	41223094	41223094	T	C	19.5	349.428	261	262	126	136	51	25	26	0.5093	1_calu-3
29	chr17	41234470	41234470	A	G	18.7	1027.15	843	846	399	447	158	75	83	0.5031	1_calu-3
30	chr17	41242939	41242940	CA	-	3.1	16.151	445	451	267	184	14	7	7	0.5919	1_calu-3
31	chr17	41244000	41244000	T	C	21	931.237	633	629	317	312	132	69	63	0.5188	1_calu-3
32	chr17	41244435	41244435	T	C	17.5	546.726	504	504	217	287	88	37	51	0.5103	1_calu-3

=B2+LEN(D2)-1

Basic Information

http://wannovar.wglab.org/

Email hyun@catholic.ac.kr

Sample Identifier Sample Identifier

Input File + Input File

or Paste Variant Calls

```
1    11308243     11308243     A  G
5    1295228  1295228  G  A
10   115511590    115511590    G  A
11   62609254     62609254
```

Submit | Reset

I agree to the Terms of Use . Plea
license.

Disease/Phenotype

Enter Disease or Phenotype Terms please enter your focused disea

Please use semicolon or enter as
Try to use multiple terms instead
OMIM IDs are also accepted, like
Better Combined with wANNOVAI

Parameter Settings

Result duration	15 days
Reference Genome	hg19
Input Fomat	ANNOVAR
Gene Definition	RefSeq Gene
Individual analysis	Individual analysis
Disease Model	none

Variant filter

1) Germline filter
2) Non-exonic variant filter (Func.refgene)
3) Synonymous variant filter (ExonicFunc.refgene)
4) Germline-like variant filter: population AF > 0.1 % (ExAC, gnomAD)
5) Variant coverage <100 on normal or tumor
6) Variant allele frequency < 2~5%

Variant prioritization among non-silent

1) Hotspot/Druggable are the most important (such as BRAFV600E)
2) Truncating (Nonsense, Frameshift indel, Splicing) > Missense
3) Polyphen_HDIV "probably damaging" (0.957-1) > "possibly damaging" (0.453-0.956) > "benign" (0-0.452)
4) ClinVar (Pathogenic > likely pathogenic > others)

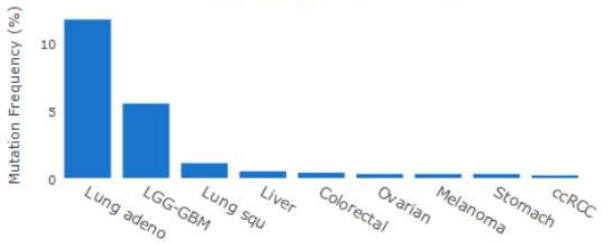
EGFR

Oncogene
Highest level of evidence: Level 1
Also known as PIG61, ERBB1, mENA, ERBB, HER1, NISBD2
Isoform: ENST00000275493 RefSeq: NM_005228.3

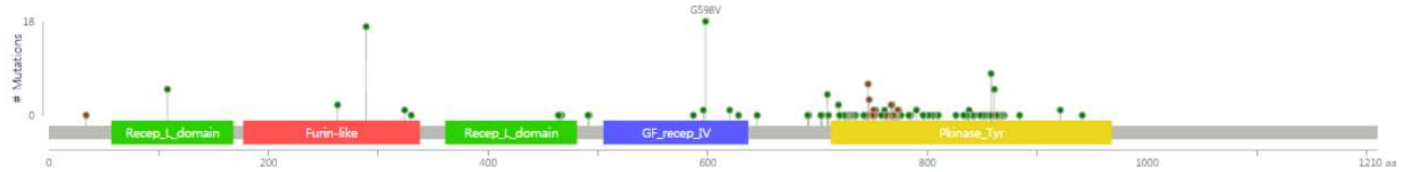
EGFR, a receptor tyrosine kinase, is altered by amplification, mutation and/or overexpression in various cancers, most frequently in lung and brain cancers.

See EGFR background

Tumor Types with EGFR Mutations



Annotated Mutation Distribution across 20 Disease Specific Studies

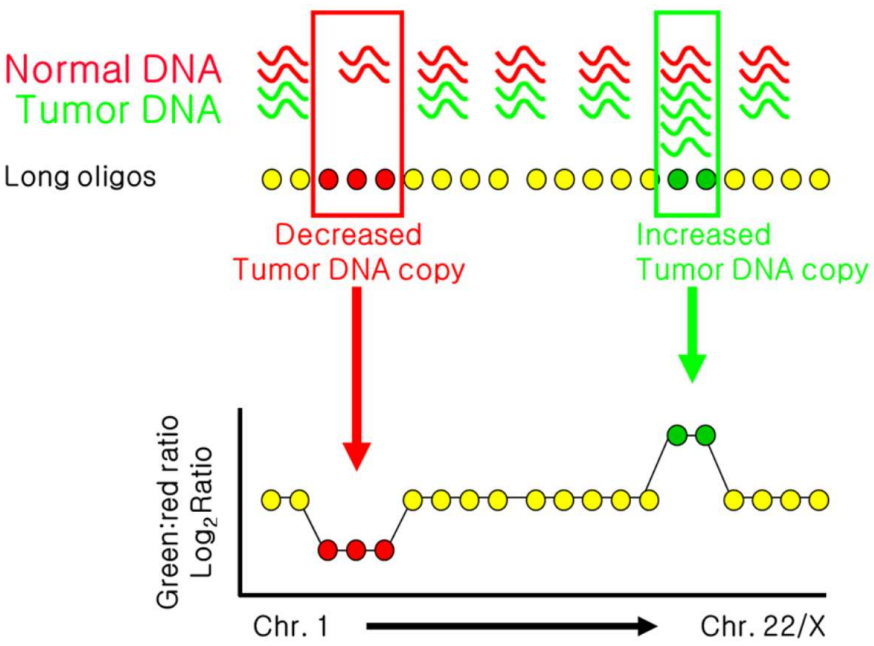


Clinically Relevant Variants (20) | All Annotated Variants (157)

If you notice any mistakes or missing variants / citations, please send an email to feedback@oncokb.org.

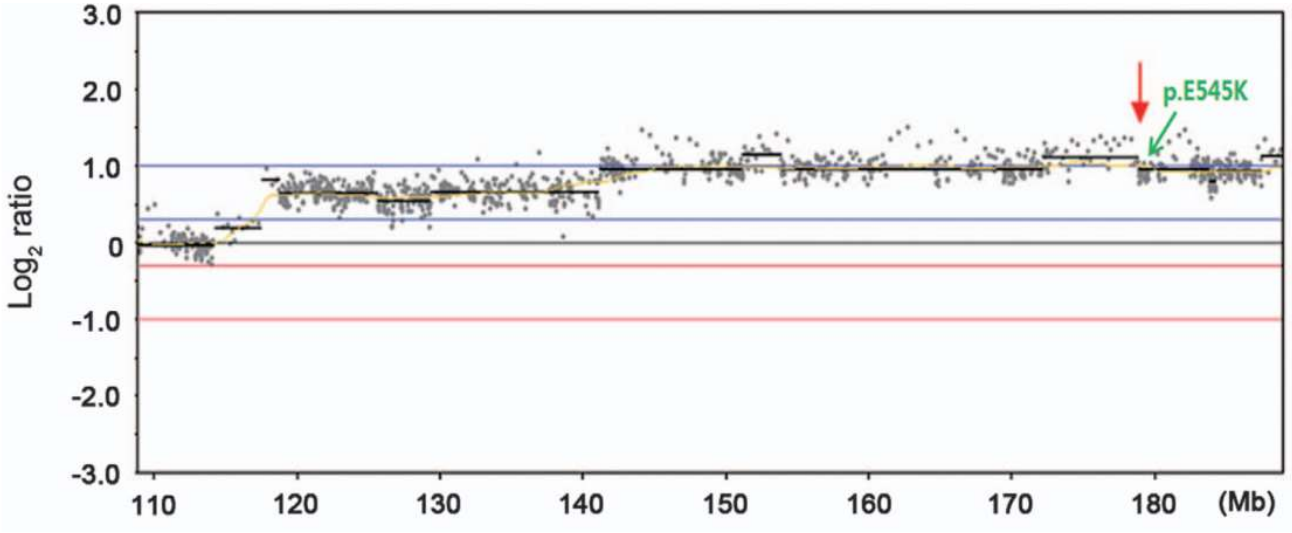
Search:

Variant	Cancer Type	Drug(s)	Level	Citations
Exon 19 deletion/insertion	Non-Small Cell Lung Cancer	Erlotinib Afatinib Gefitinib	1	12 references
Exon 19 deletion	Non-Small Cell Lung Cancer	Erlotinib Afatinib Gefitinib	1	12 references
Exon 19 insertion	Non-Small Cell Lung Cancer	Afatinib Erlotinib Gefitinib	1	1 reference
E709K	Non-Small Cell Lung Cancer	Afatinib Erlotinib Gefitinib	1	6 references
G719A	Non-Small Cell Lung Cancer	Afatinib Erlotinib Gefitinib	1	11 references



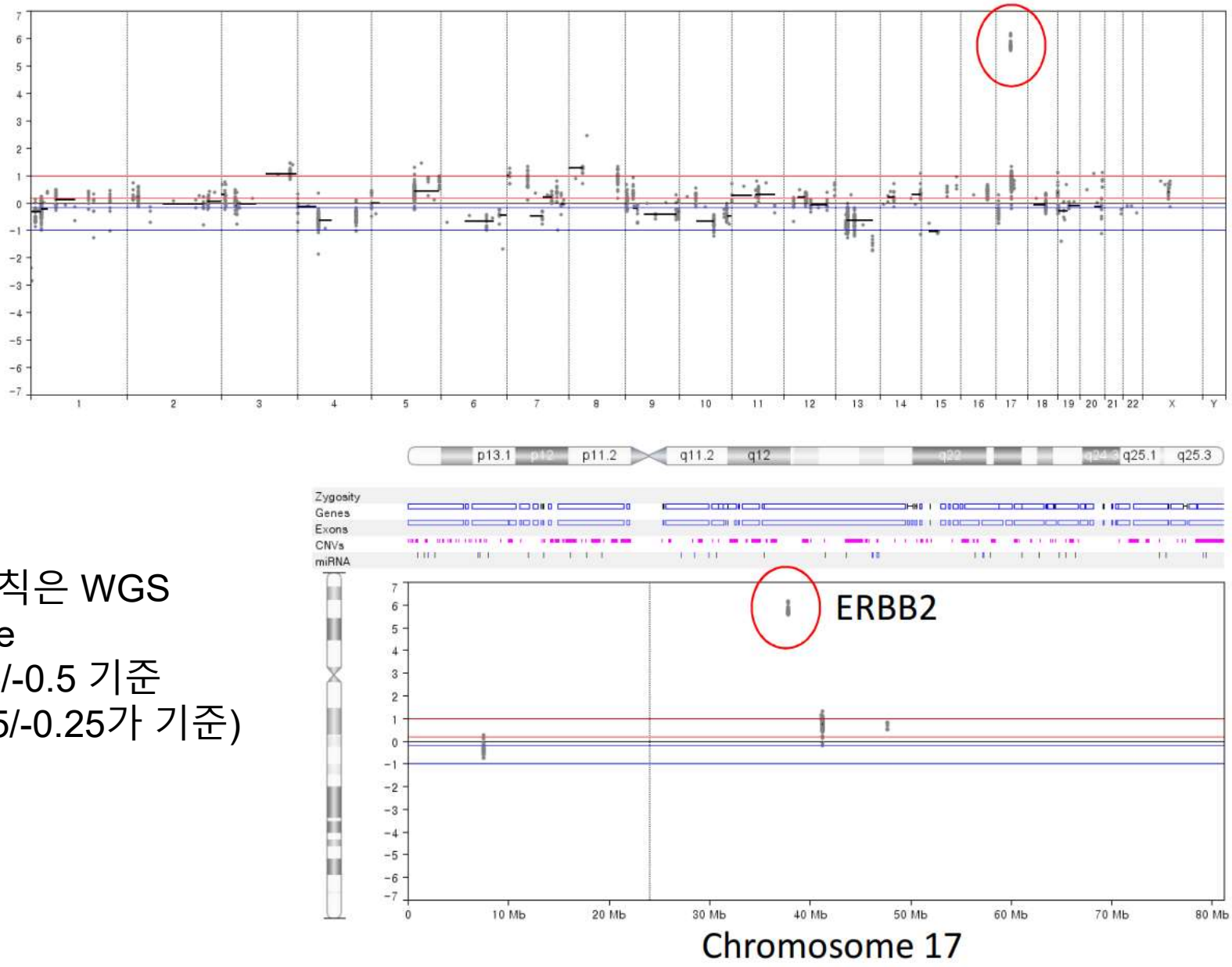
CNA analysis
Control needed
Various analysis tool
(GISTIC, NEXUS, CNVkit)
Difficult to detect in Ampliseq

Probe: Average calculated read depth within region after analysis (Point)
Segment: Final copy region, Median probe signal of CBS segmentation (Line)
Log2 Ratio: gain(3) & loss(1), 0.25 or 0.2; Amplification (>3), 0.6, Deletion(0), -1



Chung YJ, EMM
Chromosome 3

Calu3
(lung ca)

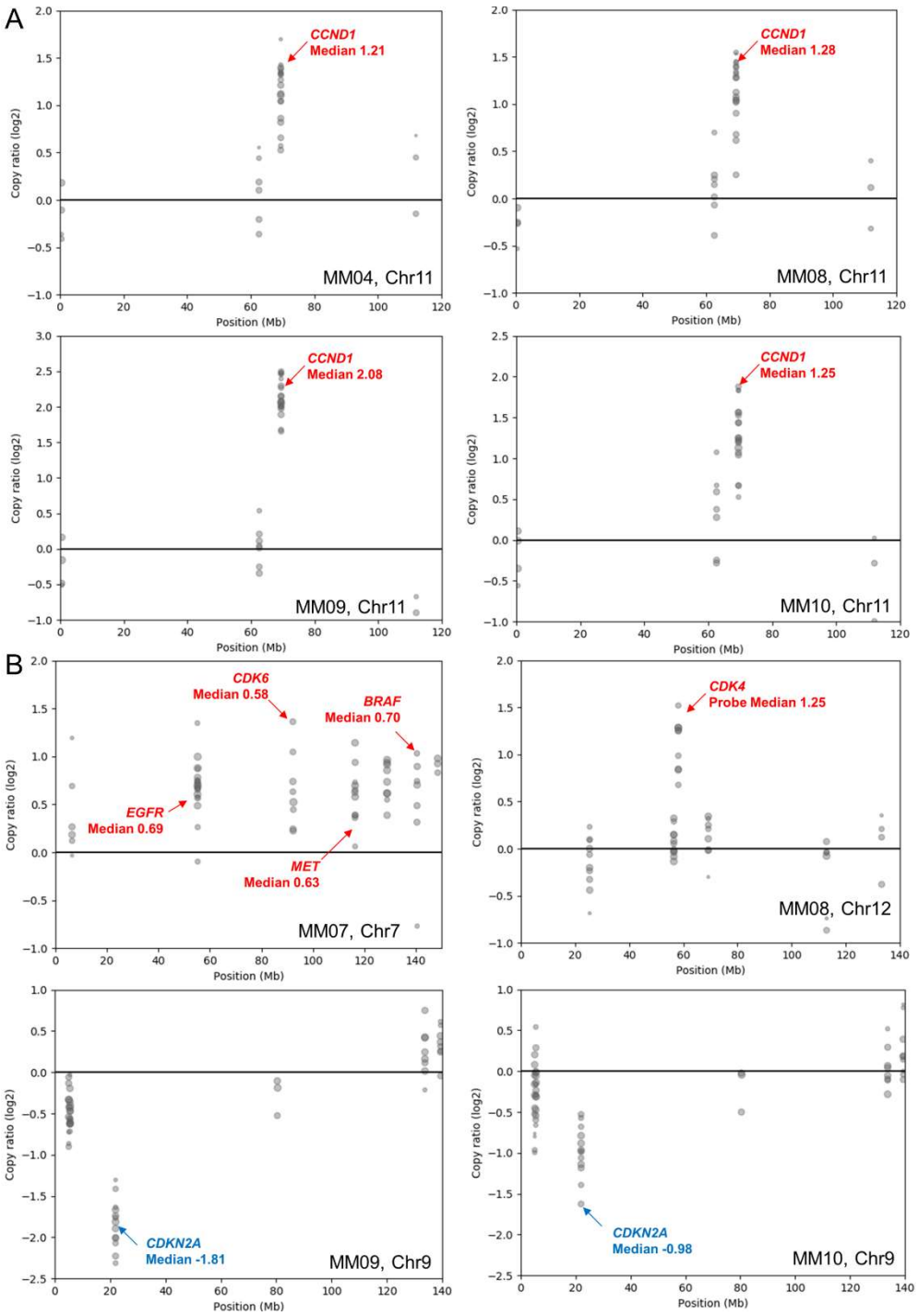


CNAs: 원칙은 WGS
High noise
logFC 0.5/-0.5 기준
(보통 0.25/-0.25가 기준)

Sample_ID	Sequencing reads	Mapped reads (%)	On target (%)	Coverage (mean)	Bases (%) (>=20 reads)
MM01N	1,501,716	1,498,454 (99.78)	98.10	1,216	99.73
MM01T	1,944,960	1,941,425 (99.82)	98.29	1,588	99.77
MM02N	1,775,009	1,771,277 (99.79)	97.36	1,423	99.74
MM02T	1,500,380	1,496,229 (99.72)	96.73	1,196	99.64
MM03N	1,704,706	1,701,453 (99.81)	97.11	1,355	99.66
MM03T	1,770,737	1,765,942 (99.73)	95.66	1,398	99.56
MM04N	1,182,958	1,179,391 (99.70)	93.23	930	99.48
MM04T	1,392,422	1,389,029 (99.76)	97.22	1,107	99.48
MM05N	1,520,480	1,517,787 (99.82)	97.24	1,213	99.64
MM05T	1,709,870	1,707,458 (99.86)	97.88	1,385	99.76
MM06N	1,427,399	1,425,178 (99.84)	97.19	1,144	99.76
MM06T	2,029,408	2,027,002 (99.88)	98.11	1,647	99.90
MM07N	1,678,487	1,674,744 (99.78)	97.06	1,332	99.75
MM07T	2,000,637	1,996,273 (99.78)	97.77	1,615	99.74
MM08N	1,671,415	1,668,112 (99.80)	98.17	1,366	99.65
MM08T	1,816,654	1,812,118 (99.75)	97.83	1,482	99.68
MM09N	1,563,892	1,560,412 (99.78)	98.14	1,276	99.76
MM09T	1,625,262	1,622,103 (99.81)	98.28	1,332	99.80
MM10N	2,023,058	2,018,771 (99.79)	98.05	1,645	99.74
MM10T	1,695,306	1,691,310 (99.76)	97.77	1,382	99.75
Average	1,676,738	1,673,223 (99.79)	97.36	1,352	99.70

Table 1. Non-silent somatic mutations identified across 10 patients with localized acral melanoma

Patient	Gene	Chr	Position	Ref	Alt	Amino acid	Exonic function	COSMIC	VAF
MM03	*NRAS*	1	115258747	C	T	p.G12D	Missense	O	32.3
MM04	*KIT*	4	55593661	T	C	p.L576P	Missense	O	7.5
MM05	*KRAS*	12	25380276	T	A	p.Q61L	Missense	O	2.5
MM07	*KIT*	4	55599320	G	C	p.D816H	Missense	O	43.6
MM07	*TP53*	17	7577574	T	C	p.Y104C	Missense	O	40.8
MM08	*KIT*	4	55594233	T	G	p.Y646D	Missense		22.7
MM08	*APC*	5	112179207	A	T	p.E2621V	Missense		16.3
MM08	*TP53*	17	7577120	C	T	p.R141H	Missense	O	30.4



CNAs: 원칙은 WGS
High noise
logFC 0.5/-0.5 기준
(보통 0.25/-0.25가 기준)