Towards high-quality clinical trials and
the implementation of genomic medicine

ATLAS Training Program

Course: Cancer Genomic Medicine: Essential Course

Lecture: Next-generation sequencing principles and analysis files

Speaker: Takashi Kubo

Takashi Kubo, Ph.D.

Department of Laboratory Medicine, National Cancer Center Hospital

EDUCATION

Kyorin University School of Health Sciences, Japan (1990 - 1994)

Kyorin University Graduated School of Health Sciences, Japan (1994-1999)

WORK EXPERIENCE

Research resident fellowship, Radiobiology Division, National Cancer Center Research Institute (1999 – 2002)

Postdoctoral fellow, Division of Pharmacology, National Institute of Health Sciences (2002 – 2005)

Fellowship researcher, Cancer Genomics Project, National Cancer Center Research Institute (2006 – 2009)

Fellowship researcher, Molecular Oncology Division, National Cancer Center Research Institute(2009 – 2010)

Research assistant, Division of Pharmacology, National Institute of Health Sciences (2010 – 2013)

Staff scientist, Division of Translational Genomics, National Cancer Center EPOC (2014 – 2019)

Staff scientist, Department of Clinical Genomics, National Cancer Center Research Institute (2014 – present)

Medical technologist, Division of Clinical Laboratory, National Cancer Center Hospital (2019 – present)

Objectives

- Understand the flow of gene panel testing
- Understand the principles of next-generation sequencing (NGS)
- Learn about data formats for analyses

Overview

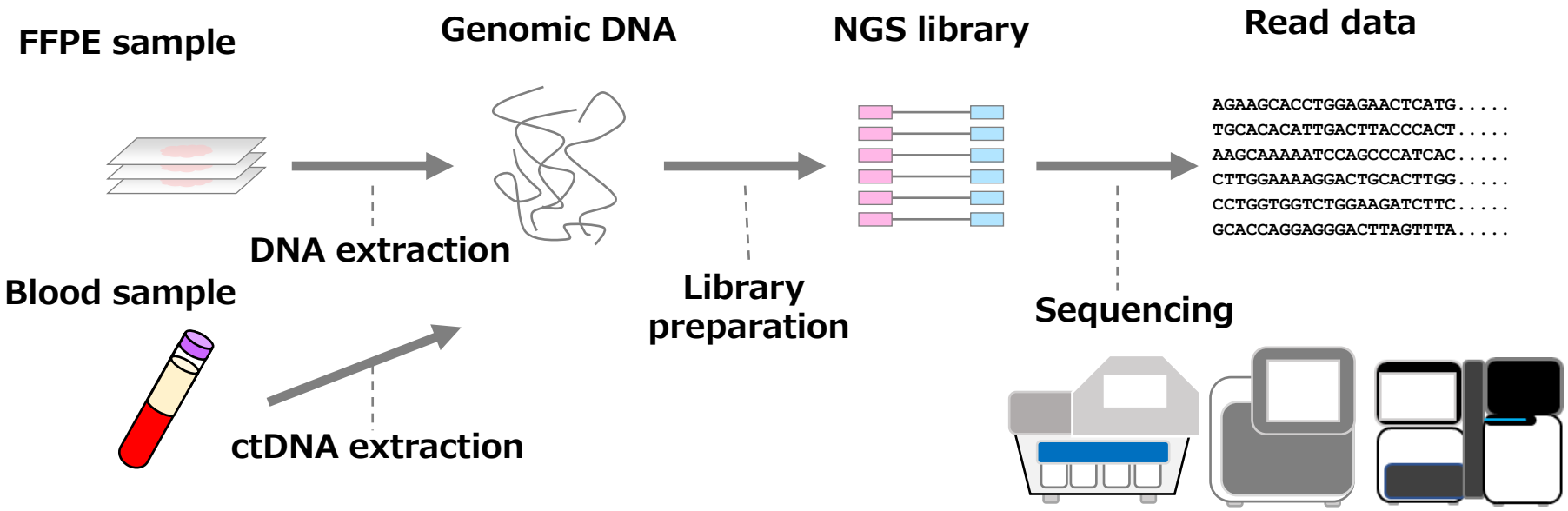
- Gene panel testing workflow (Wet bench process)
- Principles of base sequencing by NGS
- Gene panel test workflow (Dry bench process)
- File format output in NGS analyses

Overview

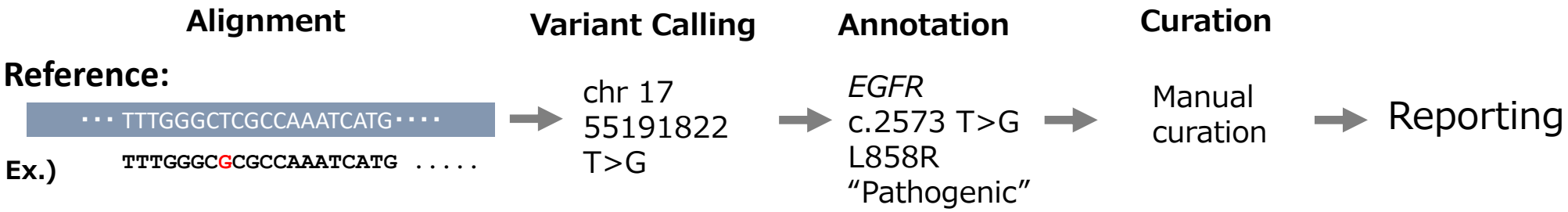
- Gene panel testing workflow (Wet bench process)
- Principles of base sequencing by NGS
- Gene panel test workflow (Dry bench process)
- Data file output by NGS analysis

Overview of the main steps in the NGS workflow

Wet bench process

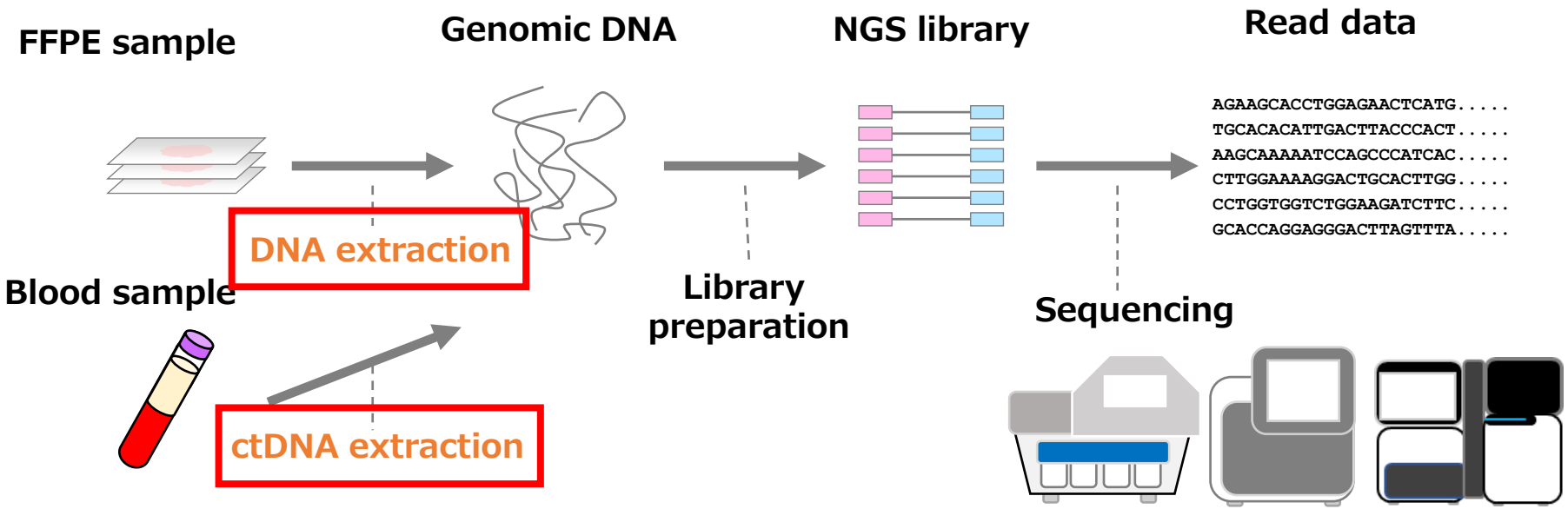


Dry bench process

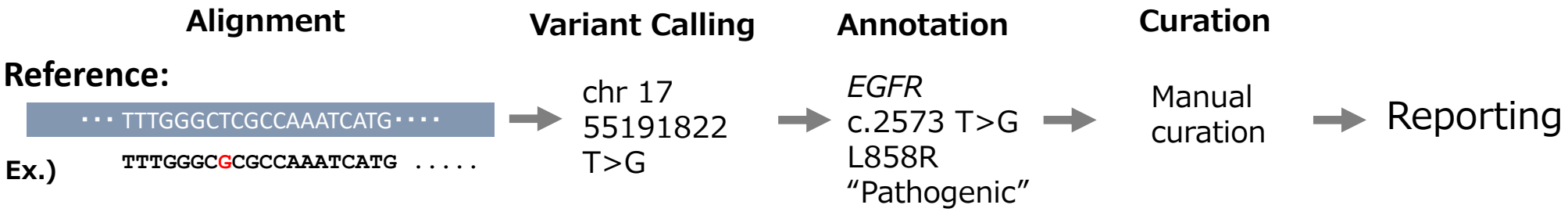


Overview of the main steps in the NGS workflow

Wet bench process

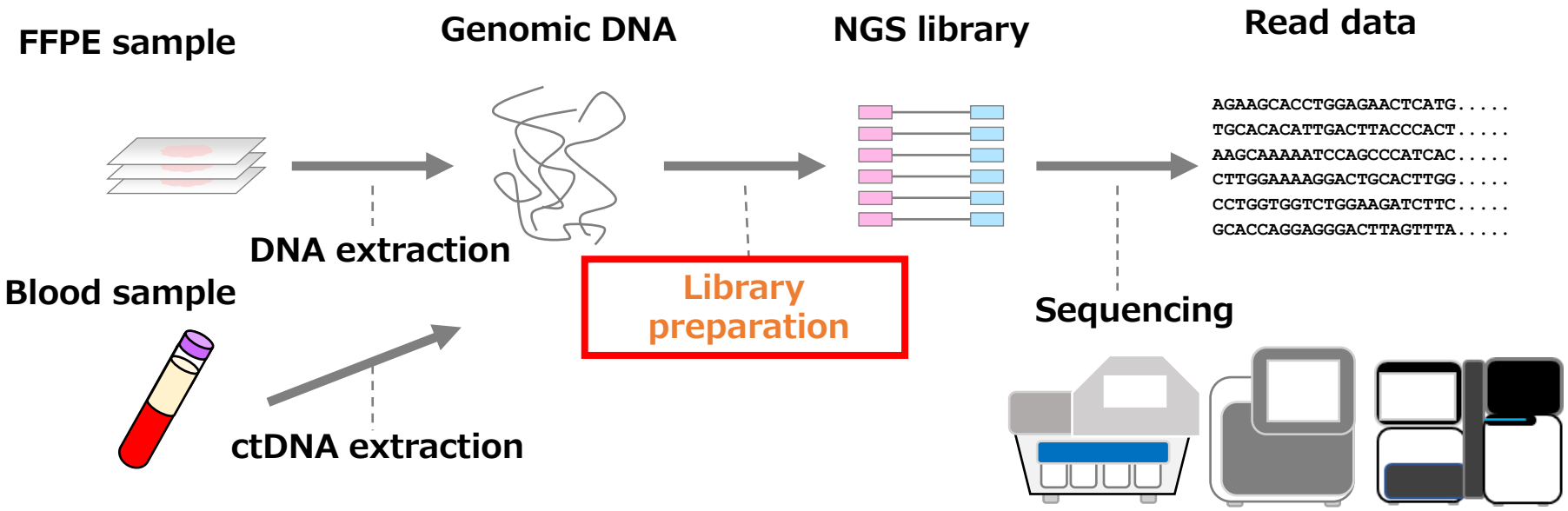


Dry part

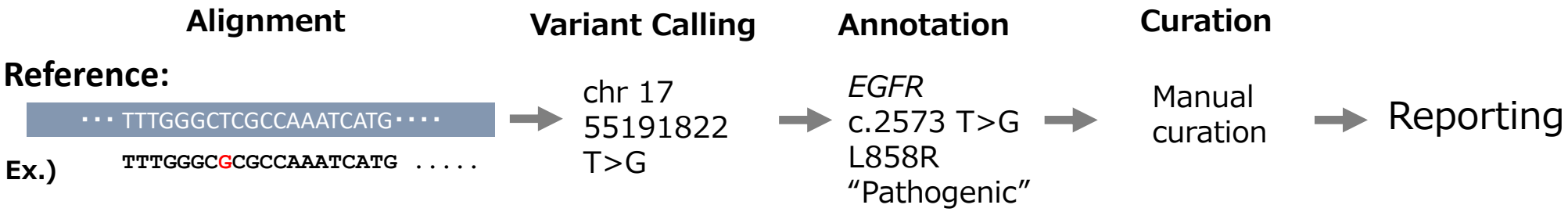


Overview of the main steps in the NGS workflow

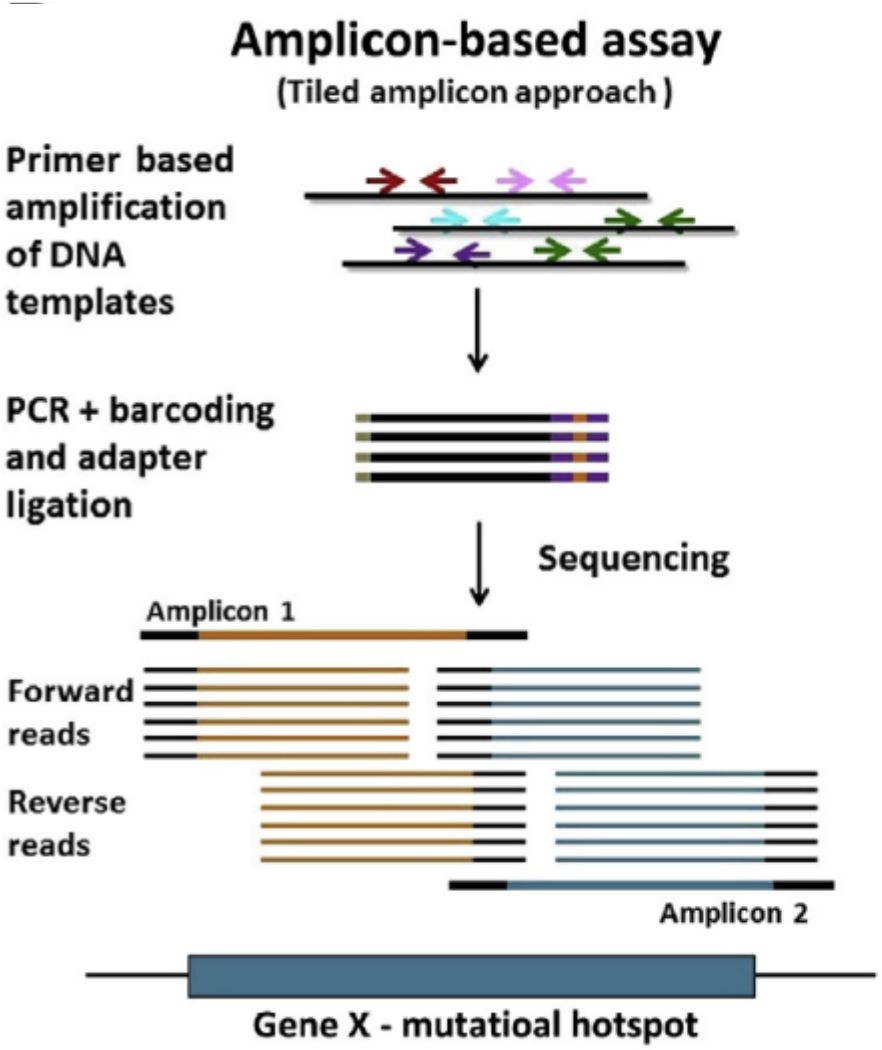
Wet bench process

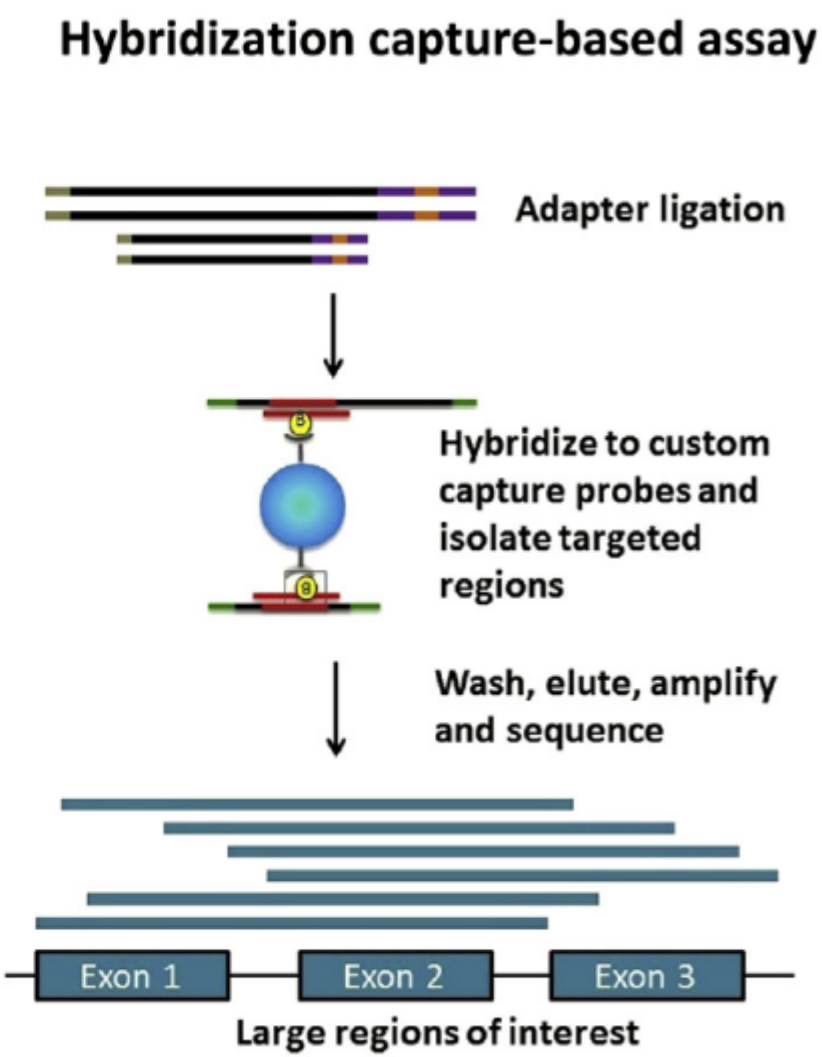


Dry part



Targeted Sequencing Methods





J Mol Diagn. 2017; 19: 341–365.

Characteristics of each library preparation method

	Amplicon-based assay	Hybridization capture-based assay
Necessary amount of DNA	Can be performed with a small amount	Better with a high amount
Work process/required time	Short	Long
Target sequence	Suitable for small regions	Suitable for wide regions
Uniformity of coverage	Low	High
Target sequence stringency	Important	Flexible, to some extent
Test in use	Oncomine Dx TT	Foundation One CDx Foundation One Liquid CDx OncoGuide NCC Oncopanel

➡ For companion diagnostic tests

➡ Suitable for genomic profiling tests

Basic structure of DNA molecules that make up the library (e.g., for Illumina NGS)

Oligonucleotide adapter 1

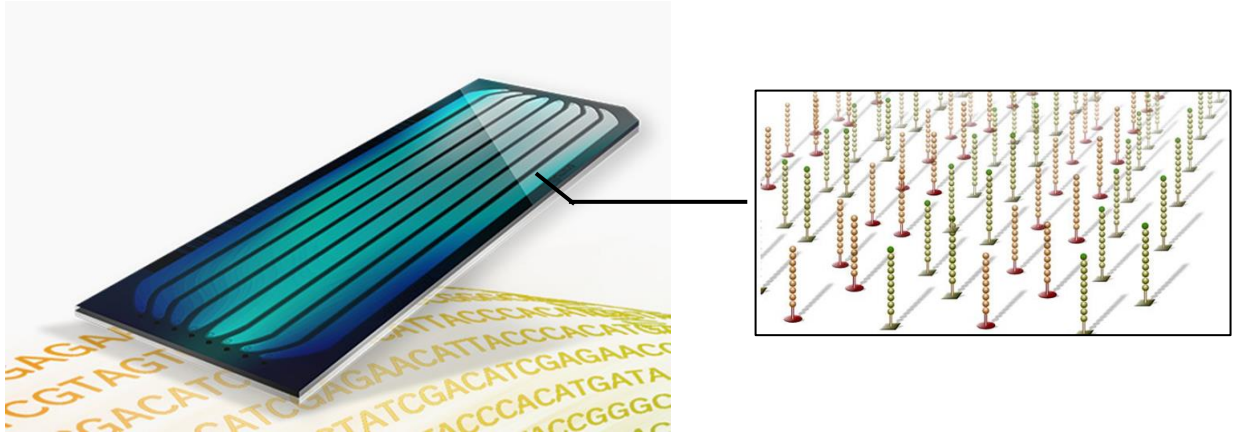
Oligonucleotide adapter 2

DNA Insert: DNA to be analyzed fragmented into hundreds of bp

P5, P7: sequences that bind to flow cells

SP: Sequencing primer binding sites

Index: A sequence for specimen identification



Flow cell (HiSeq 4000)

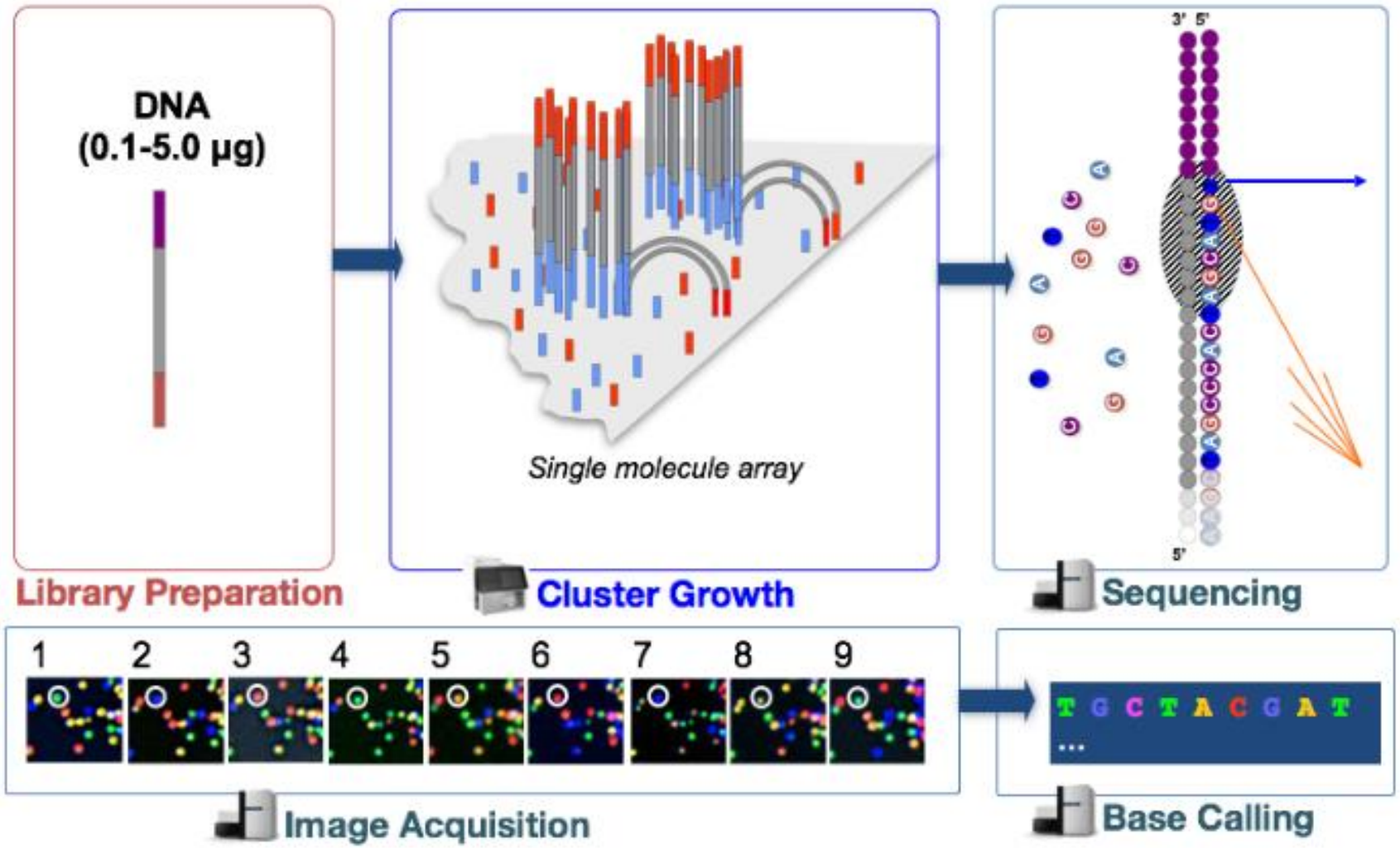
Overview

- Gene panel testing workflow (Wet bench process)
- Principles of base sequencing by NGS
- Gene panel test workflow (Dry bench process)
- Data file output by NGS analysis

Next-generation sequencing (NGS) systems

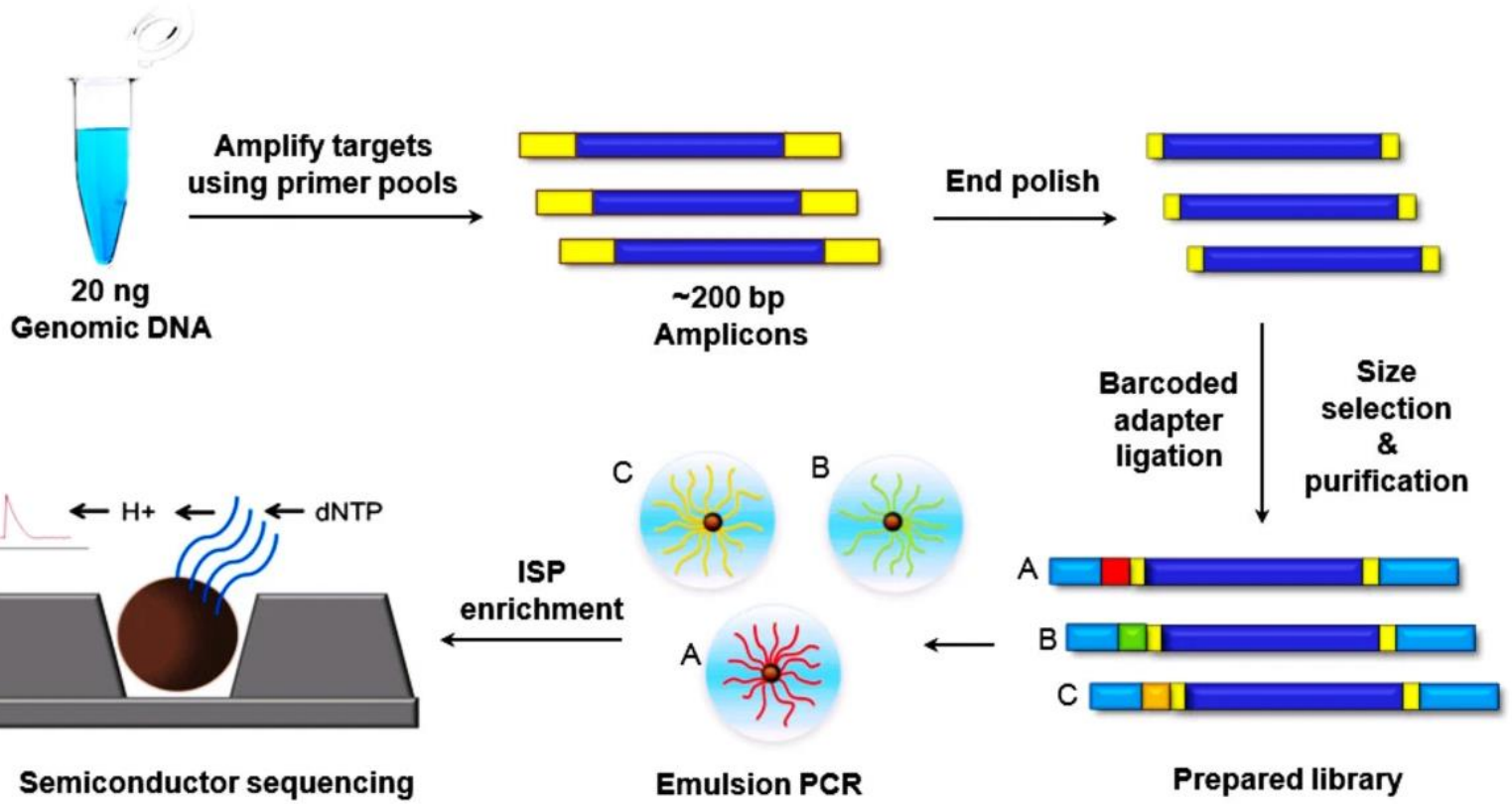
	Illumina Inc.		Ion Torrent (Thermo Fisher Scientific)
NGS machine			
	NextSeq 550Dx	HiSeq 4000	Ion PGM Dx
Detection target/Enhancement method	Bridge PCR on fluorescence/flow cells		pH change/emulsion PCR
Measurement principle	Sequencing By synthesis (SBS)		Proton measurement method
Output data volume/Run	120 Gb	1300 to 1500 Gb	600 Mb to 1 Gb

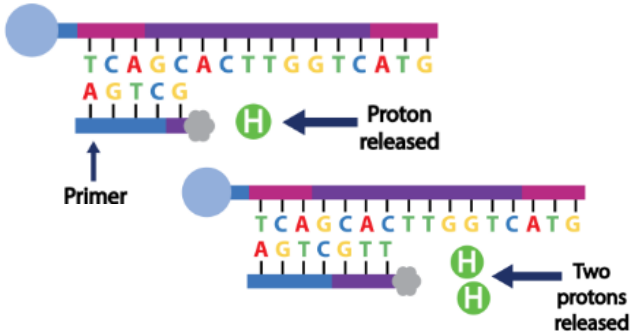
Sequencing By Illumina NGS

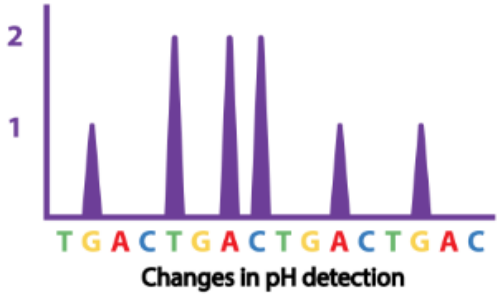


Mary Piper and Radhika Khetani. *Intro-to-ChIPseq in Github, modified*

Sequencing By Ion Torrent NGS







GTTAACCAG

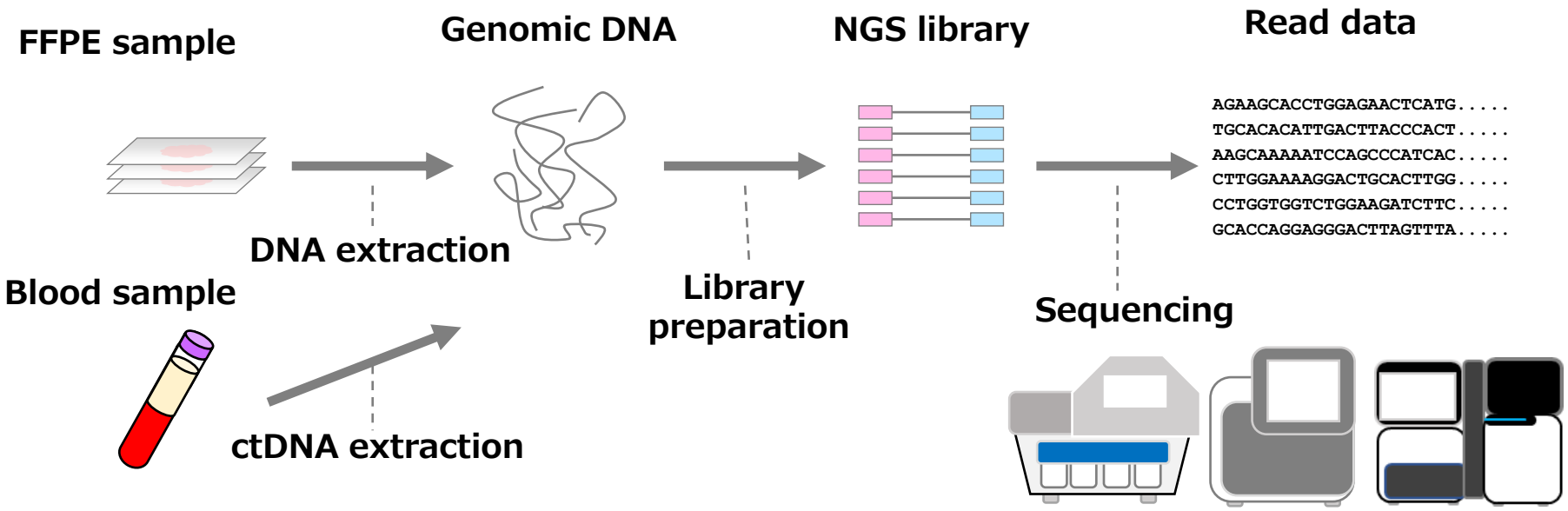
Zongzhe et al. J Transl. Medicine 173, 2014, modified
Julie Nguyen. Apollo Institute web site, modified

Overview

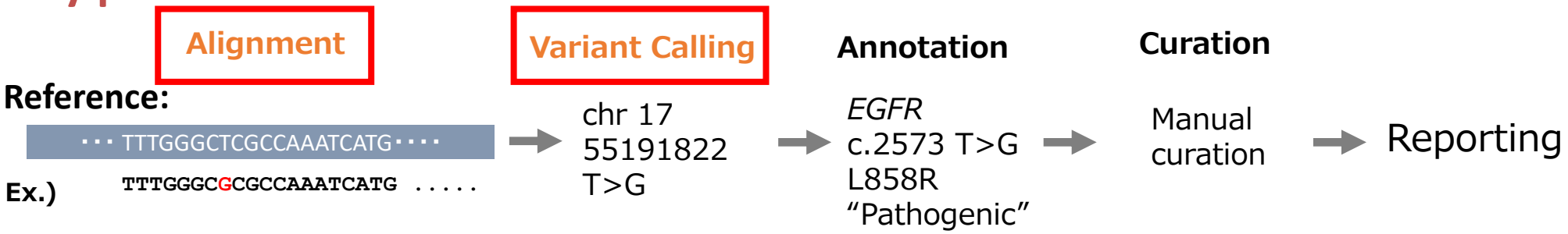
- Gene panel testing workflow (Wet bench process)
- Principles of base sequencing by NGS
- Gene panel test workflow (Dry bench process)
- Data file output by NGS analysis

Overview of the main steps in the NGS workflow

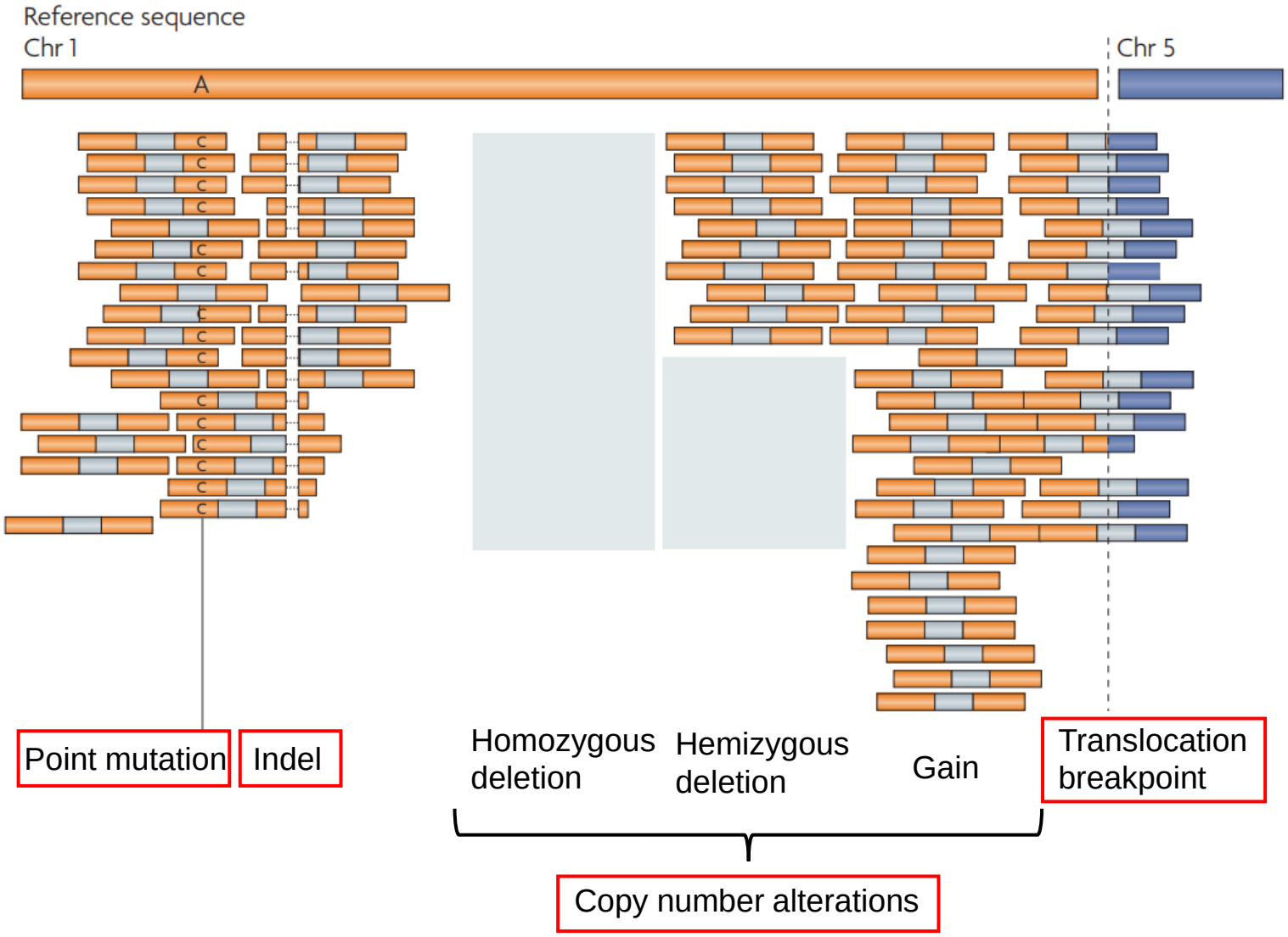
Wet bench process

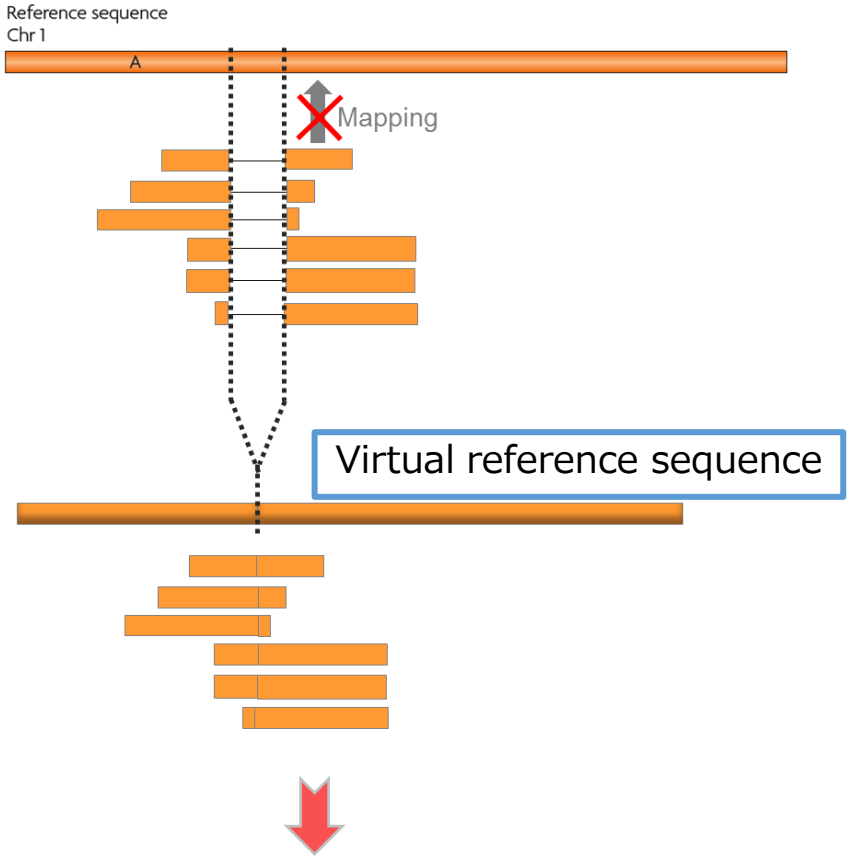


Dry part



Detection of gene mutations, amplifications/deletions, and fusions

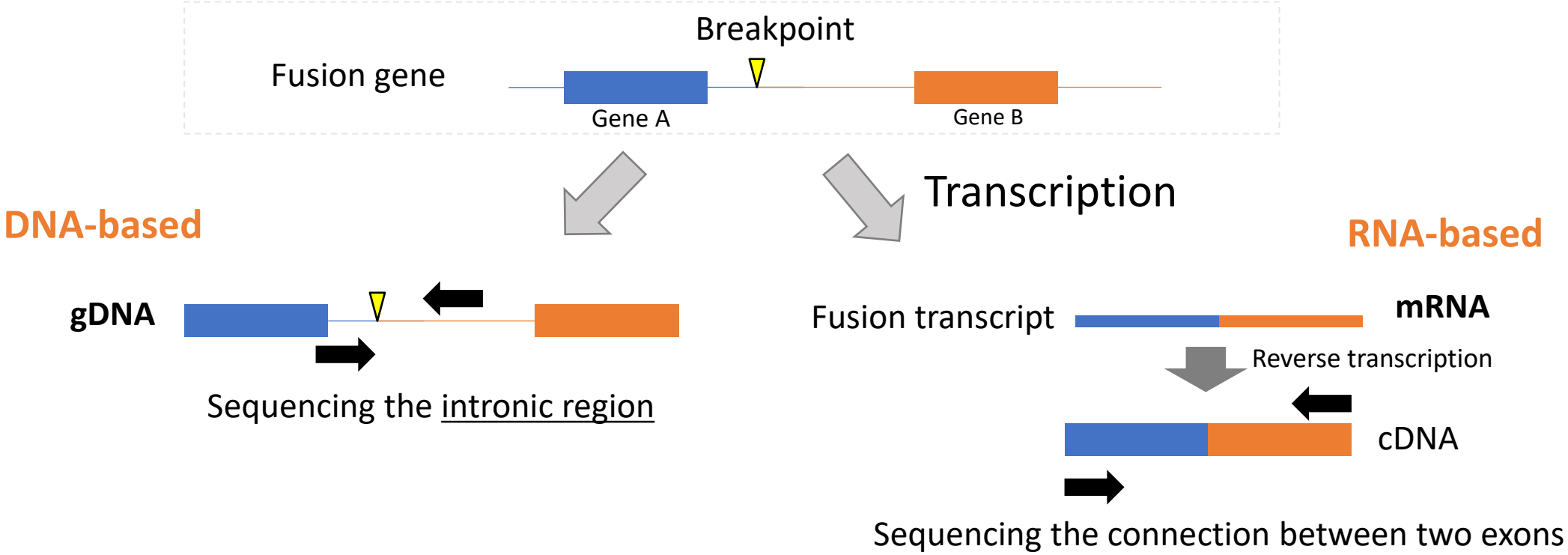




detecting complex mutations (e.g., long deletions and insertions)

Meyerson *et al.*, *Nat Rev Genet* **11**, 685 (2010), modified

DNA- *vs.* RNA-based tests for detecting fusion genes



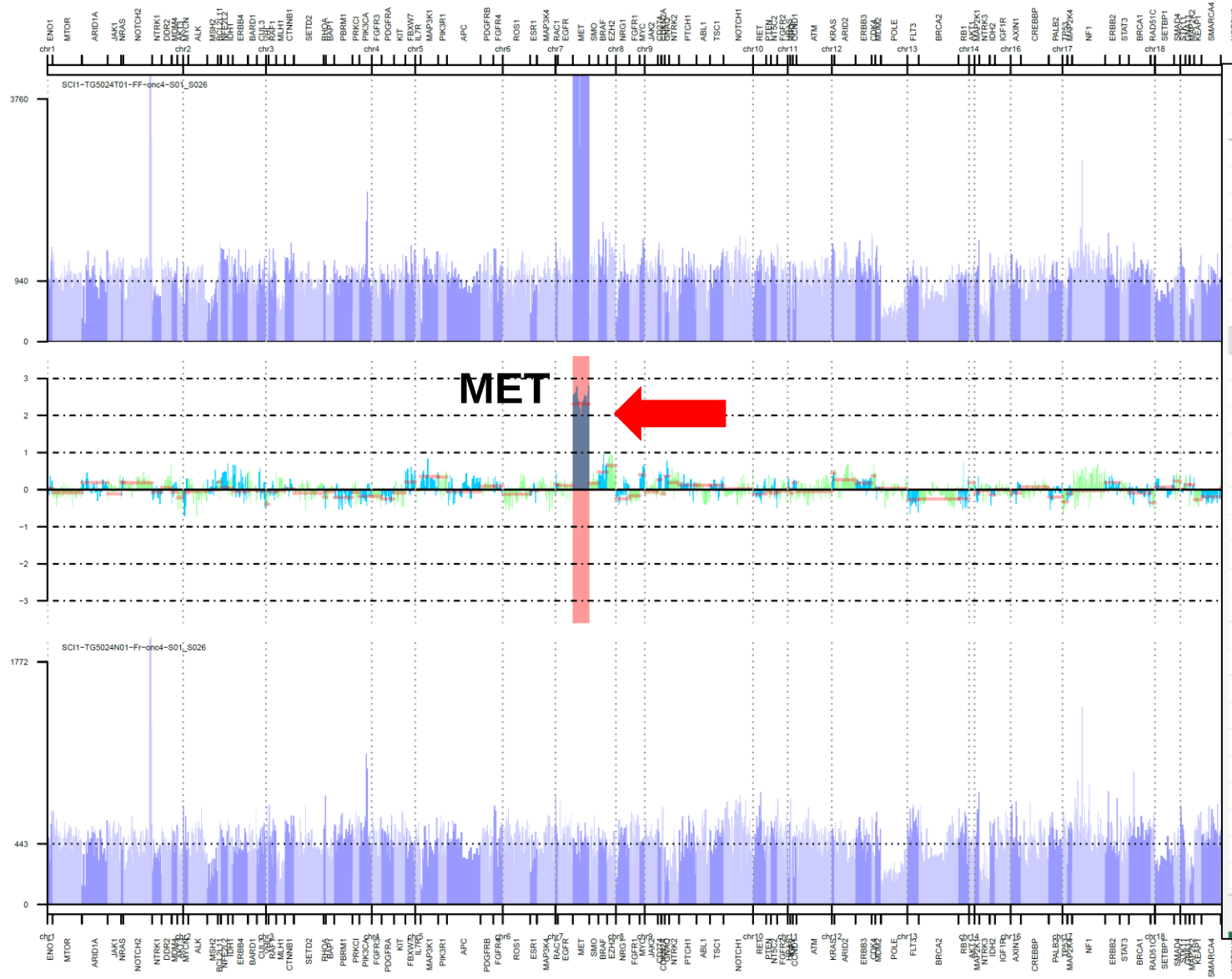
Amplification/deletion analysis

Copy number ratio

Tumor tissue

(tumor) - (normal)

Normal Tissue

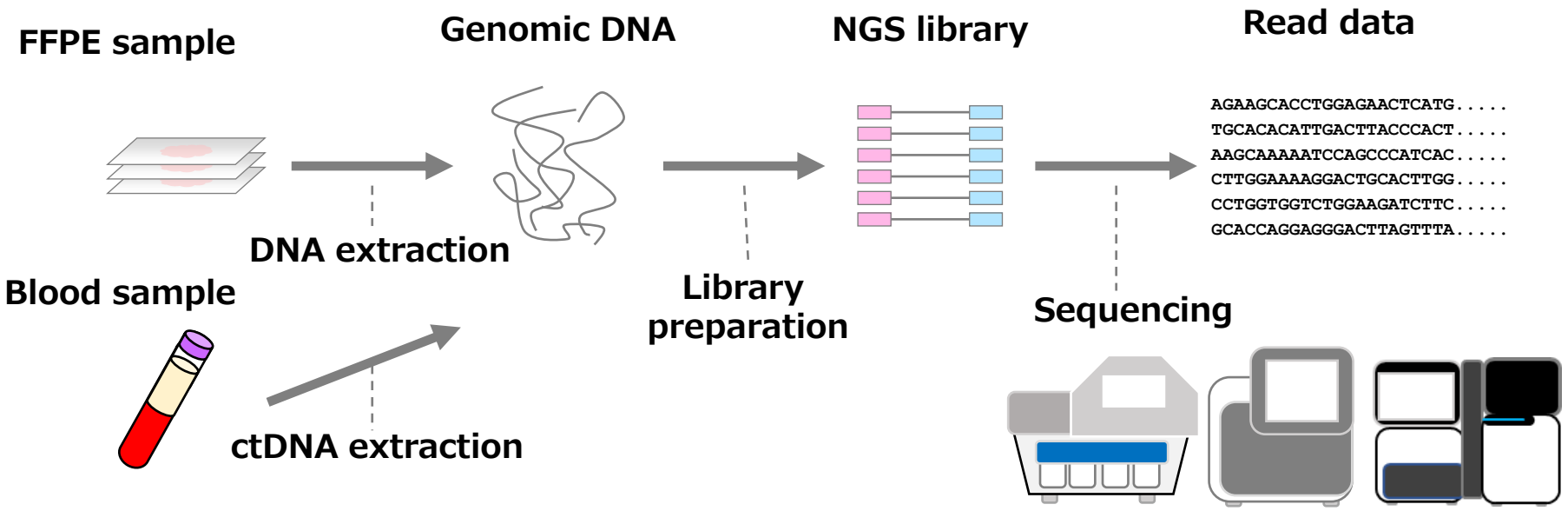


	1	2	3	4
1	Chr	Gene	Matched SCI1-TG	Matched SCI1-TG
39	5	PDGFRB	0.049833023	NotAmplified
40	5	FGFR4	0.06378649	NotAmplified
41	6	ROS1	-0.095340778	NotAmplified
42	6	ESR1	-0.032254974	NotAmplified
43	6	MAP3K4	-0.046316759	NotAmplified
44	7	RAC1	0.107825619	NotAmplified
45	7	EGFR	0.041985642	NotAmplified
46	7	MET	2.403834415	Amplified
47	7	SMO	0.169437069	NotAmplified
48	7	BRAF	0.389560012	NotAmplified
49	7	EZH2	0.660616221	NotAmplified
50	8	NRG1	-0.110659849	NotAmplified
51	8	FGFR1	-0.255713441	NotAmplified
52	8	MYC	0.406503583	NotAmplified
53	9	JAK2	0.02944836	NotAmplified
54	9	CD274	0.411647442	NotAmplified
55	9	CDKN2A	-0.11073413	NotAmplified
56	9	GNAQ	0.366262943	NotAmplified
57	9	NTRK2	0.282641559	NotAmplified
58	9	PTCH1	0.120612122	NotAmplified
59	9	ABL1	-0.019033239	NotAmplified
60	9	TSC1	0.107717038	NotAmplified
61	9	NOTCH1	-0.049057647	NotAmplified
62	10	RET	-0.12507721	NotAmplified
63	10	PTEN	-0.069297297	NotAmplified
64	10	NT5C2	0.050348424	NotAmplified
65	10	FGFR2	-0.027340331	NotAmplified
66	11	HRAS	0.074482023	NotAmplified

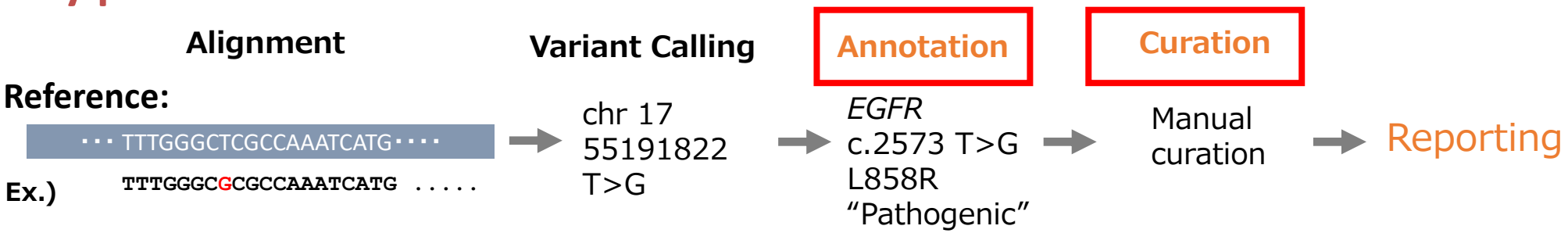
bed.exon.lowess.cnr bed.gene.lowe

Overview of the main steps in the NGS workflow

Wet bench process

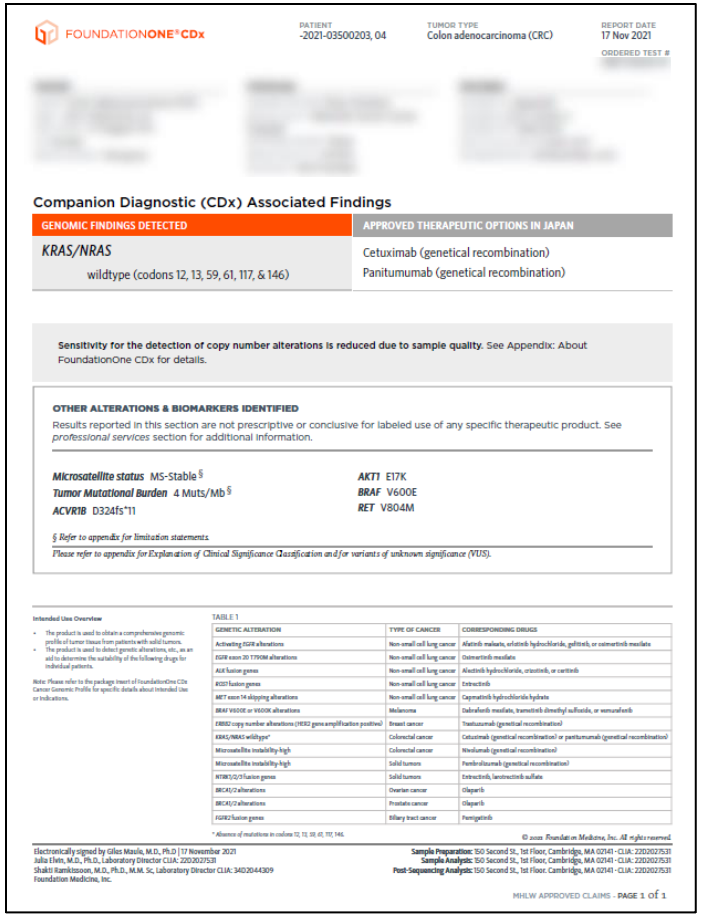


Dry part

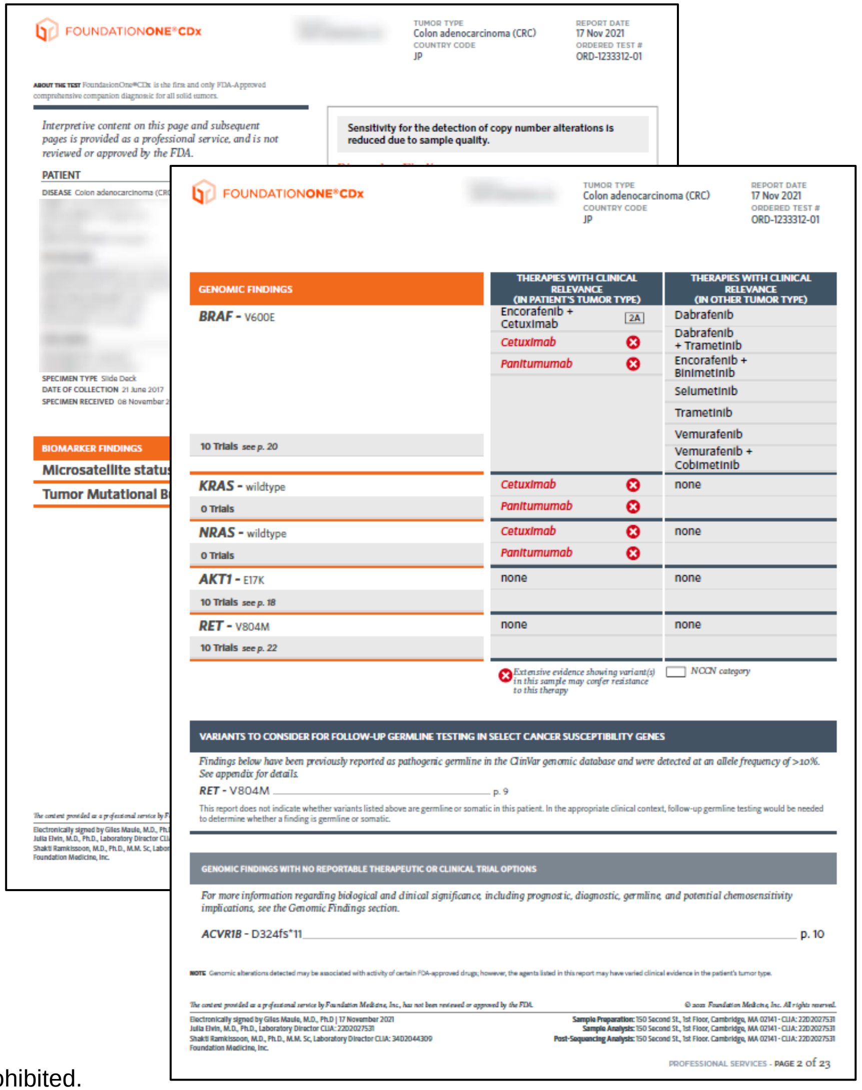


Reporting（FoundationOne CDx）

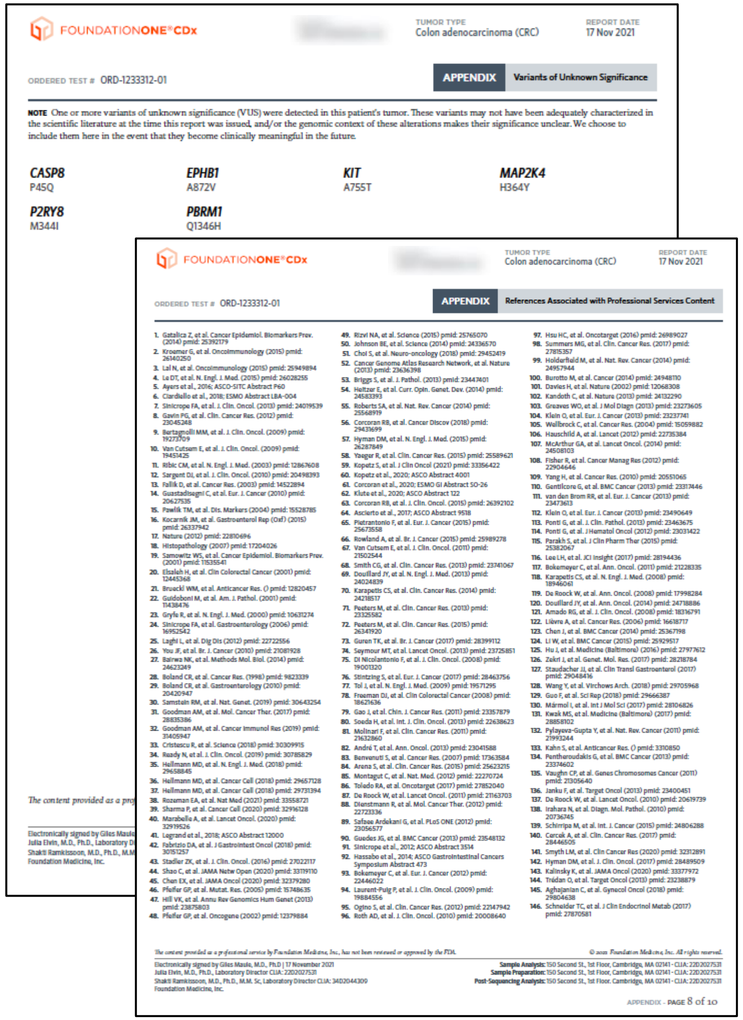
MHLW APPROVED CLAIMS



PROFESSIONAL SERVICES

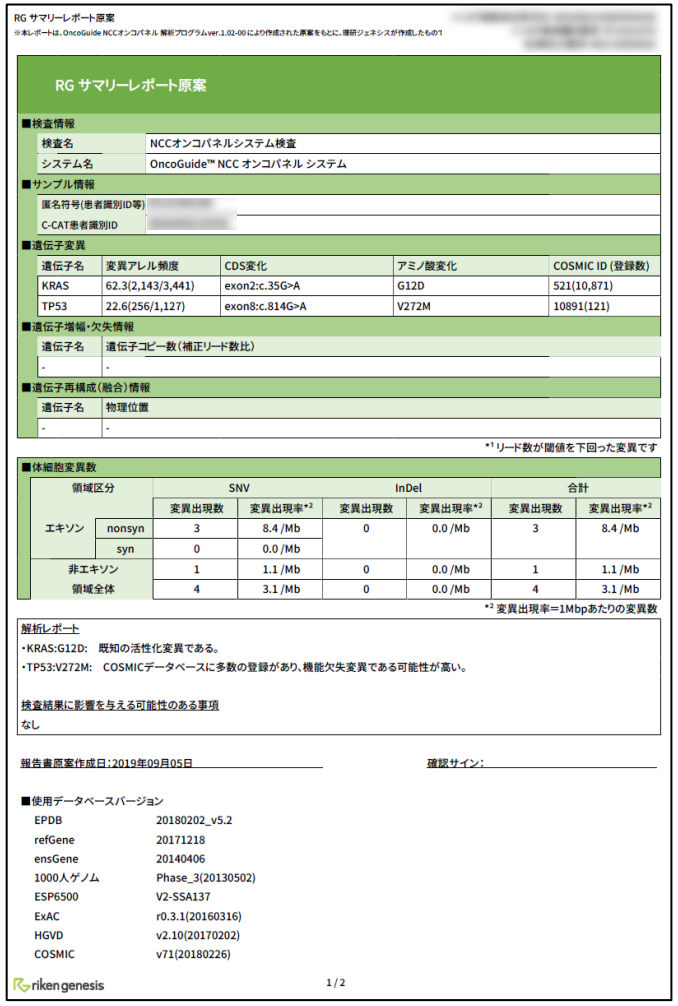


APPENDIX



Reporting（OncoGuide NCC Oncopanel System）

RG Summary Report draft



RG サマリーレポート原案

※本レポートは、OncoGuide NCCオンコパネル 解析プログラムver.1.02-00により作成された原案をもとに、理研ジェネシスが作成したもの

RG サマリーレポート原案

■検査情報

検査名	NCCオンコパネルシステム検査
システム名	OncoGuide™ NCC オンコパネル システム

■サンプル情報

匿名符号(患者識別ID等)	
C-CAT患者識別ID	

■遺伝子変異

遺伝子名	変異アレル頻度	CDS変化	アミノ酸変化	COSMIC ID (登録数)
KRAS	62.3(2,143/3,441)	exon2:c.35G>A	G12D	521(10,871)
TP53	22.6(256/1,127)	exon8:c.814G>A	V272M	10891(121)

■遺伝子増幅・欠失情報

遺伝子名	遺伝子コピー数(補正リード数比)
-	-

■遺伝子再構成(融合)情報

遺伝子名	物理位置
-	-

*1 リード数が閾値を下回った変異です

■体細胞変異数

領域区分		SNV		InDel		合計	
		変異出現数	変異出現率*2	変異出現数	変異出現率*2	変異出現数	変異出現率*2
エキソン	nonsyn	3	8.4 /Mb	0	0.0 /Mb	3	8.4 /Mb
	syn	0	0.0 /Mb				
非エキソン		1	1.1 /Mb	0	0.0 /Mb	1	1.1 /Mb
領域全体		4	3.1 /Mb	0	0.0 /Mb	4	3.1 /Mb

*2 変異出現率=1Mbpあたりの変異数

解析レポート

・KRAS:G12D:　既知の活性化変異である。

・TP53:V272M:　COSMICデータベースに多数の登録があり、機能欠失変異である可能性が高い。

検査結果に影響を与える可能性のある事項

なし

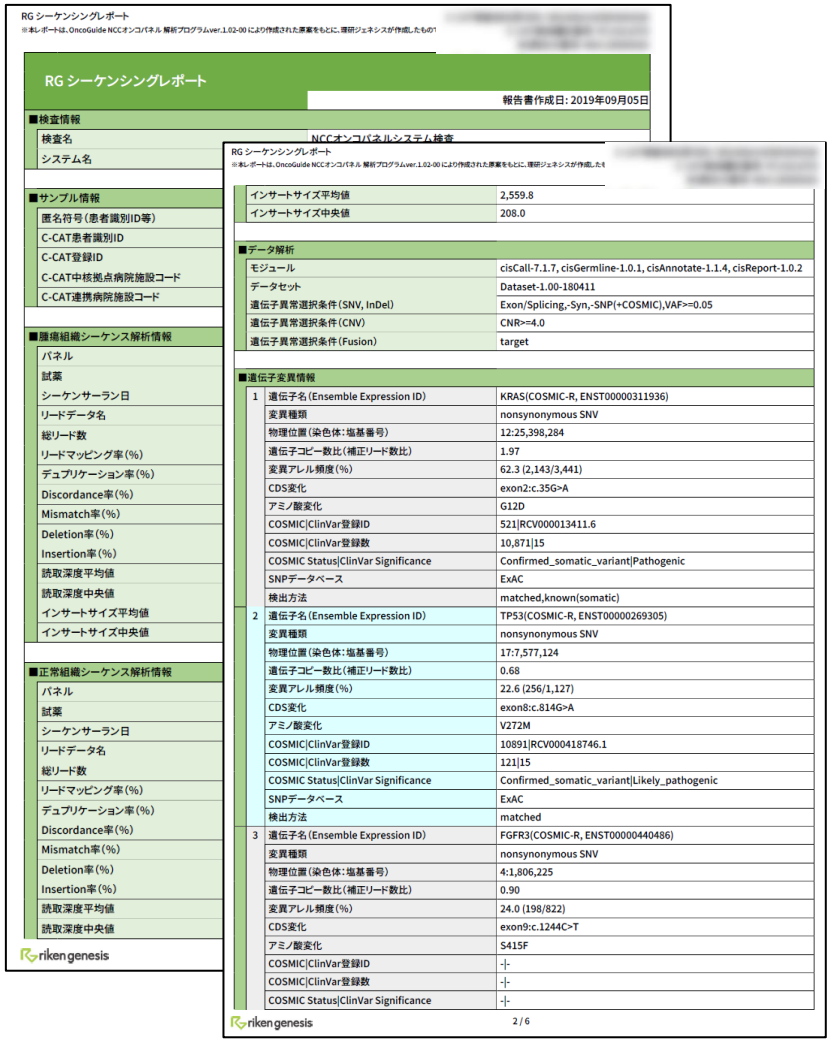
報告書原案作成日：2019年09月05日　　確認サイン：

■使用データベースバージョン

EPDB	20180202_v5.2
refGene	20171218
ensGene	20140406
1000人ゲノム	Phase_3(20130502)
ESP6500	V2-SSA137
ExAC	r0.3.1(20160316)
HGVD	v2.10(20170202)
COSMIC	v71(20180226)

rikengenesis　1 / 2

RG Sequencing Report



RG シーケンシングレポート

※本レポートは、OncoGuide NCCオンコパネル 解析プログラムver.1.02-00により作成された原案をもとに、理研ジェネシスが作成したもの

RG シーケンシングレポート

報告書作成日：2019年09月05日

■検査情報

検査名	NCCオンコパネルシステム検査
システム名	

■サンプル情報

匿名符号（患者識別ID等）	
C-CAT患者識別ID	
C-CAT登録ID	
C-CAT中核拠点病院施設コード	
C-CAT連携病院施設コード	

■腫瘍組織シーケンス解析情報

パネル	
試薬	
シーケンサーラン日	
リードデータ名	
総リード数	
リードマッピング率(%)	
デュプリケーション率(%)	
Discordance率(%)	
Mismatch率(%)	
Deletion率(%)	
Insertion率(%)	
読取深度平均値	
読取深度中央値	
インサートサイズ平均値	
インサートサイズ中央値	

■正常組織シーケンス解析情報

パネル	
試薬	
シーケンサーラン日	
リードデータ名	
総リード数	
リードマッピング率(%)	
デュプリケーション率(%)	
Discordance率(%)	
Mismatch率(%)	
Deletion率(%)	
Insertion率(%)	
読取深度平均値	
読取深度中央値	

rikengenesis

RG シーケンシングレポート

※本レポートは、OncoGuide NCCオンコパネル 解析プログラムver.1.02-00により作成された原案をもとに、理研ジェネシスが作成したも

インサートサイズ平均値	2,559.8
インサートサイズ中央値	208.0

■データ解析

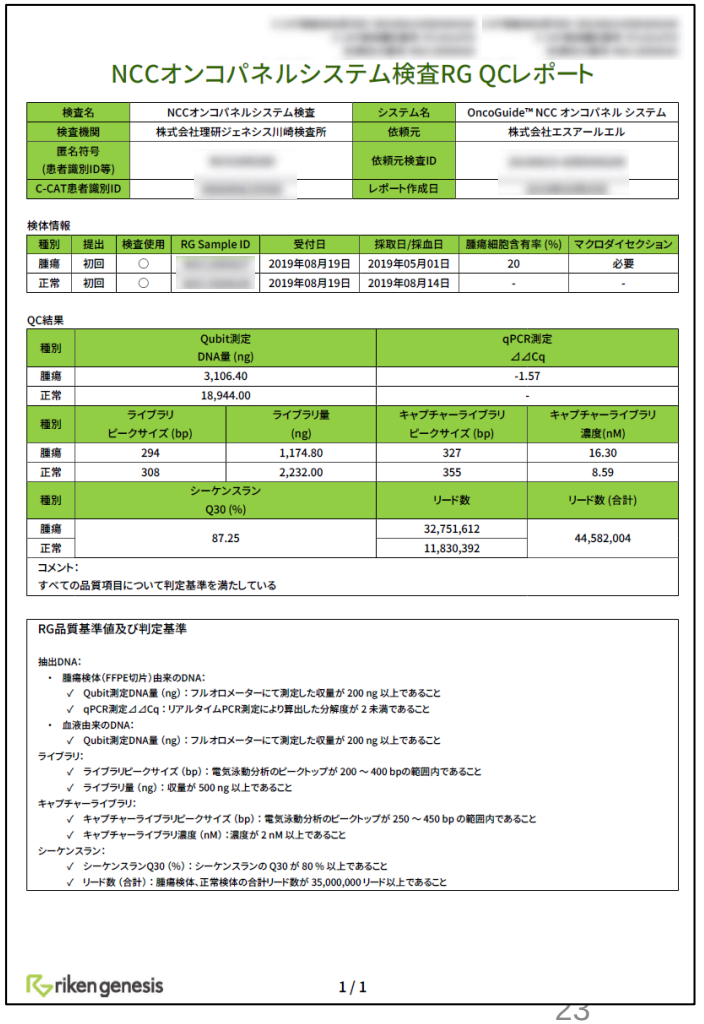
モジュール	cisCall-7.1.7, cisGermline-1.0.1, cisAnnotate-1.1.4, cisReport-1.0.2
データセット	Dataset-1.00-180411
遺伝子異常選択条件(SNV, InDel)	Exon/Splicing,-Syn,-SNP(+COSMIC),VAF>=0.05
遺伝子異常選択条件(CNV)	CNR>=4.0
遺伝子異常選択条件(Fusion)	target

■遺伝子変異情報

1	遺伝子名(Ensemble Expression ID)	KRAS(COSMIC-R, ENST00000311936)
	変異種類	nonsynonymous SNV
	物理位置(染色体:塩基番号)	12:25,398,284
	遺伝子コピー数比(補正リード数比)	1.97
	変異アレル頻度(%)	62.3 (2,143/3,441)
	CDS変化	exon2:c.35G>A
	アミノ酸変化	G12D
	COSMIC|ClinVar登録ID	521|RCV000013411.6
	COSMIC|ClinVar登録数	10,871|15
	COSMIC Status|ClinVar Significance	Confirmed_somatic_variant|Pathogenic
	SNPデータベース	ExAC
	検出方法	matched,known(somatic)
2	遺伝子名(Ensemble Expression ID)	TP53(COSMIC-R, ENST00000269305)
	変異種類	nonsynonymous SNV
	物理位置(染色体:塩基番号)	17:7,577,124
	遺伝子コピー数比(補正リード数比)	0.68
	変異アレル頻度(%)	22.6 (256/1,127)
	CDS変化	exon8:c.814G>A
	アミノ酸変化	V272M
	COSMIC|ClinVar登録ID	10891|RCV000418746.1
	COSMIC|ClinVar登録数	121|15
	COSMIC Status|ClinVar Significance	Confirmed_somatic_variant|Likely_pathogenic
	SNPデータベース	ExAC
	検出方法	matched
3	遺伝子名(Ensemble Expression ID)	FGFR3(COSMIC-R, ENST00000440486)
	変異種類	nonsynonymous SNV
	物理位置(染色体:塩基番号)	4:1,806,225
	遺伝子コピー数比(補正リード数比)	0.90
	変異アレル頻度(%)	24.0 (198/822)
	CDS変化	exon9:c.1244C>T
	アミノ酸変化	S415F
	COSMIC|ClinVar登録ID	-|-
	COSMIC|ClinVar登録数	-|-
	COSMIC Status|ClinVar Significance	-|-

rikengenesis　2 / 6

RG QC report



NCCオンコパネルシステム検査RG QCレポート

検査名	NCCオンコパネルシステム検査	システム名	OncoGuide™ NCC オンコパネル システム
検査機関	株式会社理研ジェネシス川崎検査所	依頼元	株式会社エスアールエル
匿名符号(患者識別ID等)		依頼元検査ID	
C-CAT患者識別ID		レポート作成日	

検体情報

種別	提出	検査使用	RG Sample ID	受付日	採取日/採血日	腫瘍細胞含有率 (%)	マクロダイセクション
腫瘍	初回	○		2019年08月19日	2019年05月01日	20	必要
正常	初回	○		2019年08月19日	2019年08月14日	-	-

QC結果

種別	Qubit測定 DNA量 (ng)	qPCR測定 ⊿⊿Cq
腫瘍	3,106.40	-1.57
正常	18,944.00	-

種別	ライブラリ ピークサイズ (bp)	ライブラリ量 (ng)	キャプチャーライブラリ ピークサイズ (bp)	キャプチャーライブラリ 濃度(nM)
腫瘍	294	1,174.80	327	16.30
正常	308	2,232.00	355	8.59

種別	シーケンスラン Q30 (%)	リード数	リード数 (合計)
腫瘍	87.25	32,751,612	44,582,004
正常		11,830,392	

コメント：

すべての品質項目について判定基準を満たしている

RG品質基準値及び判定基準

抽出DNA:

- 腫瘍検体(FFPE切片)由来のDNA:
  - ✓ Qubit測定DNA量 (ng)：フルオロメーターにて測定した収量が 200 ng 以上であること
  - ✓ qPCR測定⊿⊿Cq：リアルタイムPCR測定により算出した分解度が 2 未満であること
- 血液由来のDNA:
  - ✓ Qubit測定DNA量 (ng)：フルオロメーターにて測定した収量が 200 ng 以上であること

ライブラリ:

- ✓ ライブラリピークサイズ (bp)：電気泳動分析のピークトップが 200 ～ 400 bpの範囲内であること
- ✓ ライブラリ量 (ng)：収量が 500 ng 以上であること

キャプチャーライブラリ:

- ✓ キャプチャーライブラリピークサイズ (bp)：電気泳動分析のピークトップが 250 ～ 450 bp の範囲内であること
- ✓ キャプチャーライブラリ濃度 (nM)：濃度が 2 nM 以上であること

シーケンスラン:

- ✓ シーケンスランQ30 (%)：シーケンスランの Q30 が 80 % 以上であること
- ✓ リード数 (合計)：腫瘍検体、正常検体の合計リード数が 35,000,000 リード以上であること

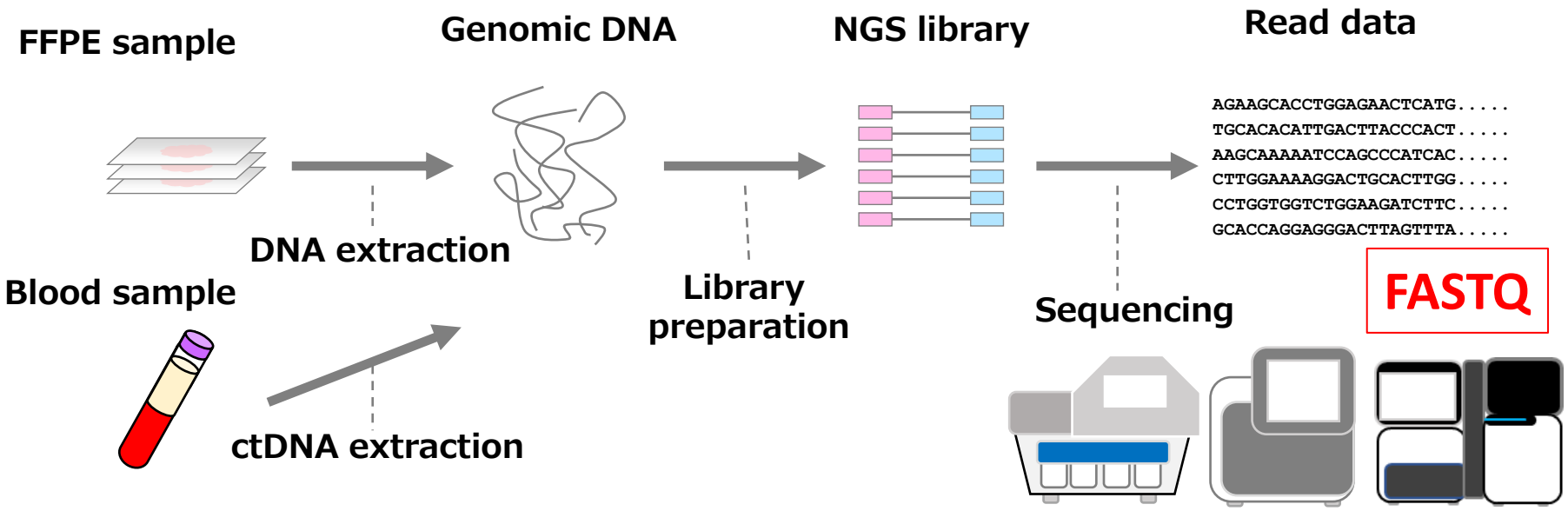
rikengenesis　1 / 1

Overview

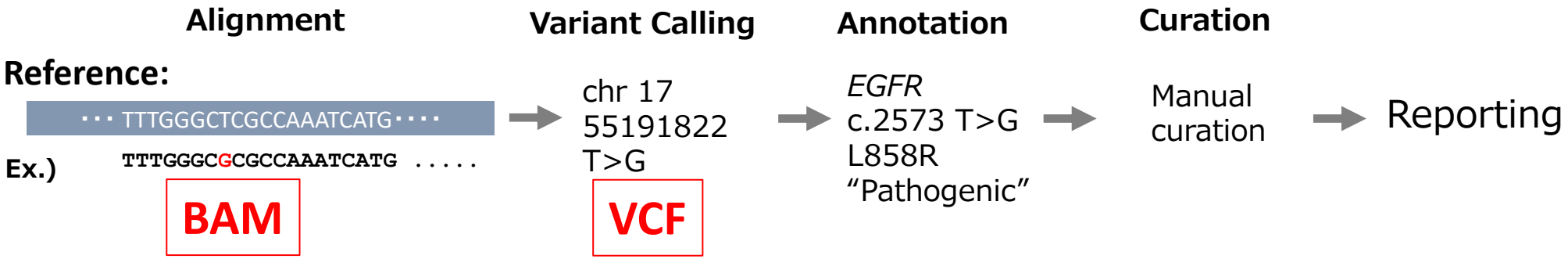
- Gene panel testing workflow (Wet bench process)
- Principles of base sequencing by NGS
- Gene panel test workflow (Dry bench process)
- Data file output by NGS analysis

Overview of the main steps in the NGS workflow

Wet bench process



Dry part



FASTQ File

- Sequence information read by the sequencer (read information)
- Expresses the information for one read in four lines

```
@NB501521:131:HFN75AFXY:1:11101:18629:1018 1:N:0:CTGGCATA
CCTGGNCTTGAGCCCCCGTTCTTAATCACCATGCCCTACTTCCCCCAGGGGTTCCCCCTGTGGTGTTGCTGTGACATTTTTATGCCATAGTTCGGACTCCTGCTGCCTATTTCAGCAGTCATGATGGTCCCCTGCCTGTTGTGGTGCAGGA
+
AAAAA#EEEE/EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAEAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEAEEEEEEEEAEEEEEEEAEEEEEEEEEEEEA<<AEEAEEEEEEEEEAEEE<
```

1st line : In the Read ID Paired-end, the end of Read-1 is "1:" and the end of Read-2 is "2:" to maintain associativity.

2nd line : Base sequence of reads

3rd line : Comment box

4th line : Estimated accuracy (Q-score) of each base in ASCII code

BAM File

- Alignment information: read mapping to the reference genome

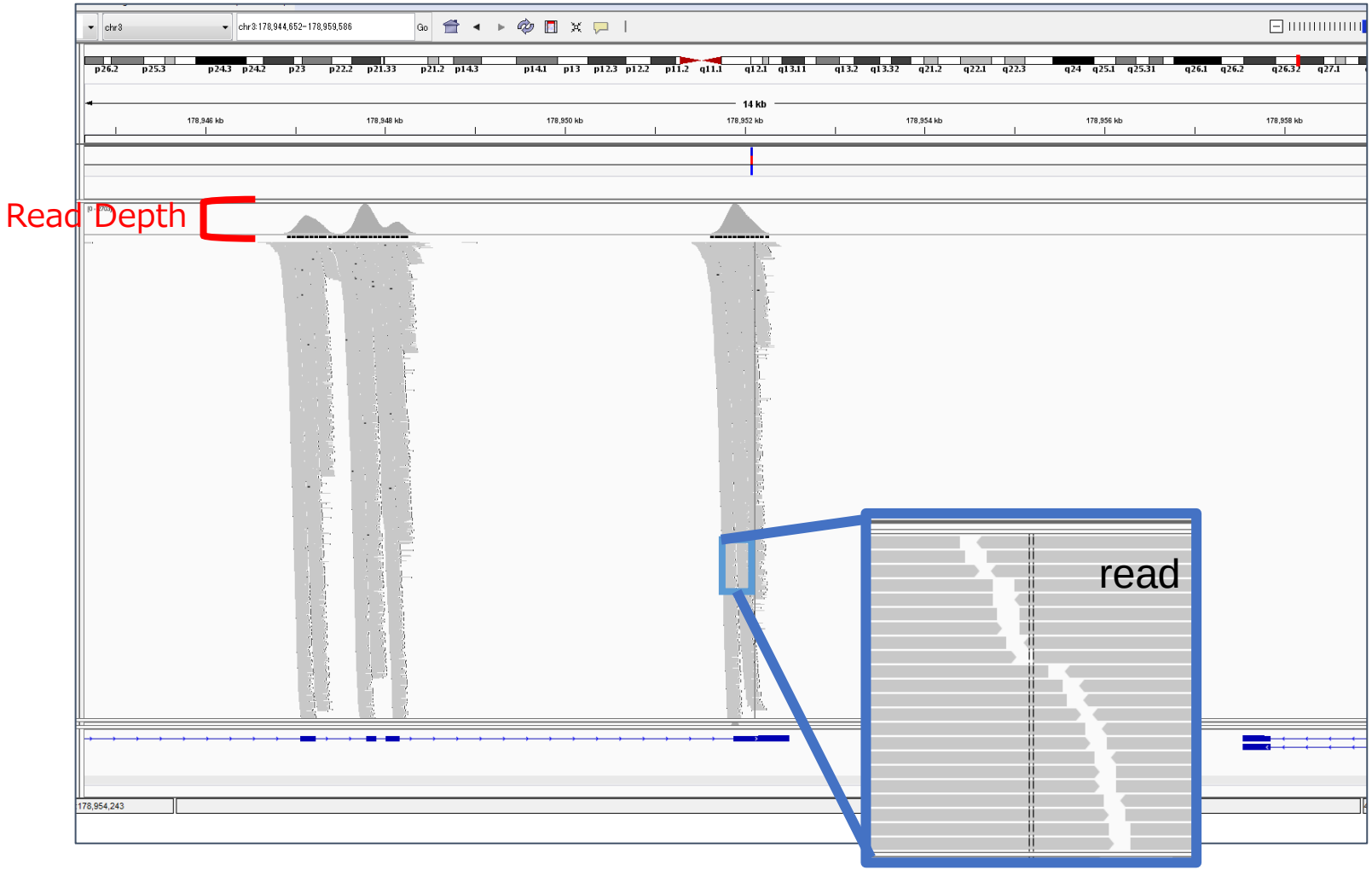
- A single line shows data from one sequence read, containing information about sequences that differ from the reference sequence.

```
NB501521:147:HHVG3AFXY:2:11212:17736:6634    19    chr1    13133    0    151M    chr9    13153    0
TCTCCTACCTGAGGCTGAGGAAGGAGAAGGGGATGCACTGTTGGGAAGGCAGCTGTAACTCAAAGCCTTAGCCTCTGTGCCCACGAAGGCAGGGCCATCAGGCACCAAATGGATTCTGCCAGCATAGTGCTCCTGGACCAGTGATACACCC
/AE/A/A//EEEE//EEEE/E6/6///EA//EA/E/A/AA/A/<A/AA<AE/E/6EA////E//E/AEA<//EA//AE/EEEEE//</EE<//<EA///AE/EEE///////E/EA////A/E/E/////E/E/AE//AE//EE////AA
XT:A:R  NM:i:3  XO:i:2  X1:i:2  XM:i:3  XO:i:0  XG:i:0  MD:Z:45G32T30G41    XA:Z:chr15:1:102531392,+102517887,151M,3;chr2:1:243199373,+114357733,151M,4;
chr16:1:90354753,-62814,151M,4;
```

- Can be visualized with Integrative Genomics Viewer (IGV), etc.

IGV: Integrative Genomics Viewer

- A genome browser (freeware) created by the Broad Institute (United States)



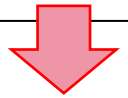
Read information can be visualized by opening the BAM file with IGV

VCF File

- All information for a single mutation is listed on one line, separated by tabs.
- Includes detected variants, sequencing data, and various annotation information

1 variant information

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NOAxxxxxxxxxx
chr17	7577552	_	CAT	C	100	PASS	GENE=T	GT:DP:CN	1/1:1805:0.948004



GENE=TP53;AF=0.5224;DP=1805;VFR=inf;DPP=7577553;DPB=1904;FG=836|10|16|0|943|0;BG=1904|0|0|0|0|0;HL=1;SMAF=0.5224;Func=exonic;ExomeFunc=frame_shift;FullAA=TP53:NM_001126117:exon3:c.331_332delAT:p.M111fs*20:7:4;FullAAEns=ENSG00000141510:ENST00000413465:exon6:c.727_728delAT:p.M243fs*43:7:4;Transcript=NM_001126112;NucleotideChange=c.727_728delAT;AA=p.M243fs*20;1000G_ALL=.;1000G_AFR=.;1000G_AMR=.;1000G_EAR=.;1000G_EUR=.;1000G_SAS=.;ExAC_Freq=.;ExAC_AFR=.;ExAC_AMR=.;ExAC_EAS=.;ExAC_FIN=.;ExAC_NFE=.;ExAC_OTH=.;ExAC_SAS=.;ESP6500si_ALL=.;ESP6500si_AA=.;HGVD_Qual=.;HGVD_ALL=.;COSMIC_ID=.;COSMIC_STATUS=.;COSMIC_CNT=.;COSMIC_DIS=.;ClinVar_SIG=.;ClinVar_DIS=.;ClinVar_STATUS=.;ClinVar_ID=.;ClinVar_DB=.;1000G_VER=Phase_3(20130502);ExAC_VER=r0.3.1(20160316);HGVD_VER=v2.10(20170202);COSMIC_VER=v71(20180226);ClinVar_VER=20170905;RefGene_VER=20171218;RMCNT=0;CURATED;;ClinVar_DBID=.;ESP6500si_EA=.;ESP6500si_VER=V2-SSA137;ExonNo=7;GSD=.;RMASK=.;SMPLRPT=.

Ex) NCC Oncopanel System: vcf by cisCall contains annotation information, such as:

- Changes in bases and amino acids
- Registration status in the polymorphism database
- Registration status in COSMIC and ClinVar

Overview

- Gene panel testing workflow (Wet bench process)
- Principles of base sequencing by NGS
- Gene panel test workflow (Dry part)
- Data file output by NGS analysis