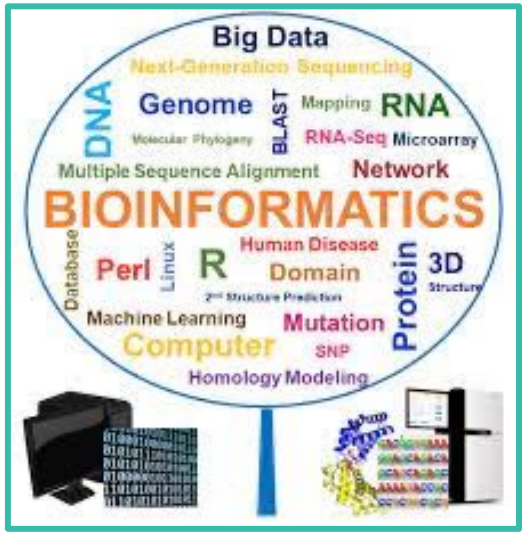
Next Generation Sequencing

GUNADI

Pediatric Surgery Division, Department of Surgery
Genetics Working Group/Translational Research Unit
FK-KMK UGM/RSUP Dr. Sardjito

Yogyakarta, 7 December 2022



Outlines

1. Teori Dasar – Sequencing (NGS)

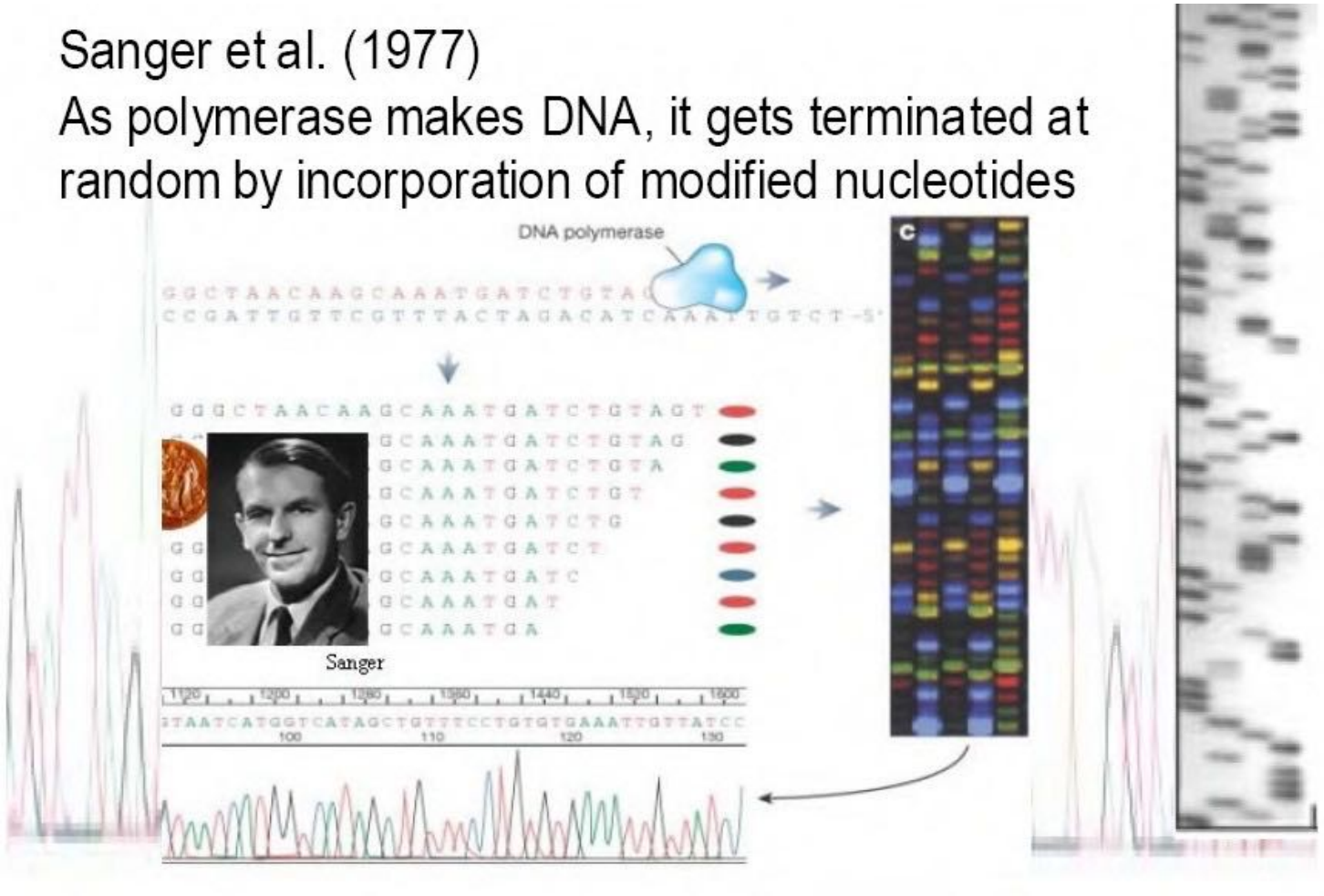
2. Analisis NGS

3. Aplikasi NGS

TEORI DASAR SEQUENCING (NGS)

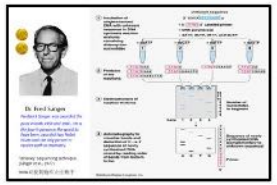
HOW DOES SEQUENCING WORK?

Refers to determining the order of nucleotide (G, A, T, and C) in a stretch of DNA



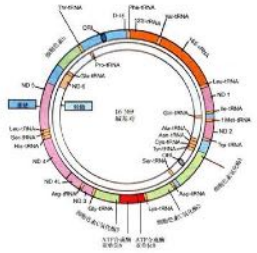
(Axygen, 2008)

HISTORY OF SEQUENCING TECHNOLOGY



1977 Sanger method

1981 Human mitochondrial genome sequence



1990 Human Genome Project

1995 Complete cell genome

1996 Complete eukaryotic genome

2001 Complete the Human Genome Project

2005 Second generation sequencer：454 GS20

2007 Second generation sequencer：Genetic Analyzer 2

2008 Research Human Microbiome Project

2011 Third generation sequencer：PacBio RS

2014 Nanospace sequencing

2019 The third stage Human microbiome project

CLASSICAL CHAIN-TERMINATION SEQUENCING

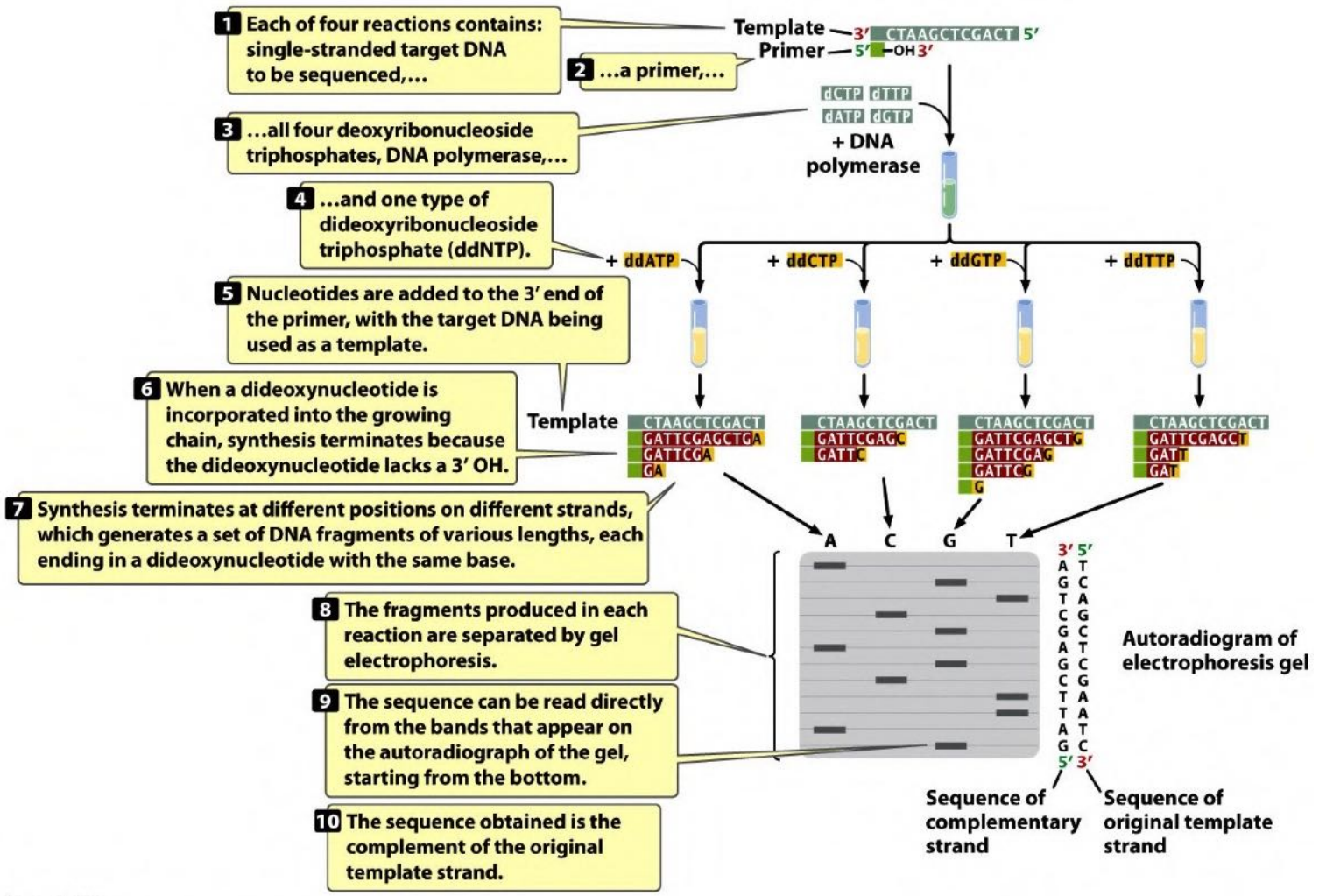
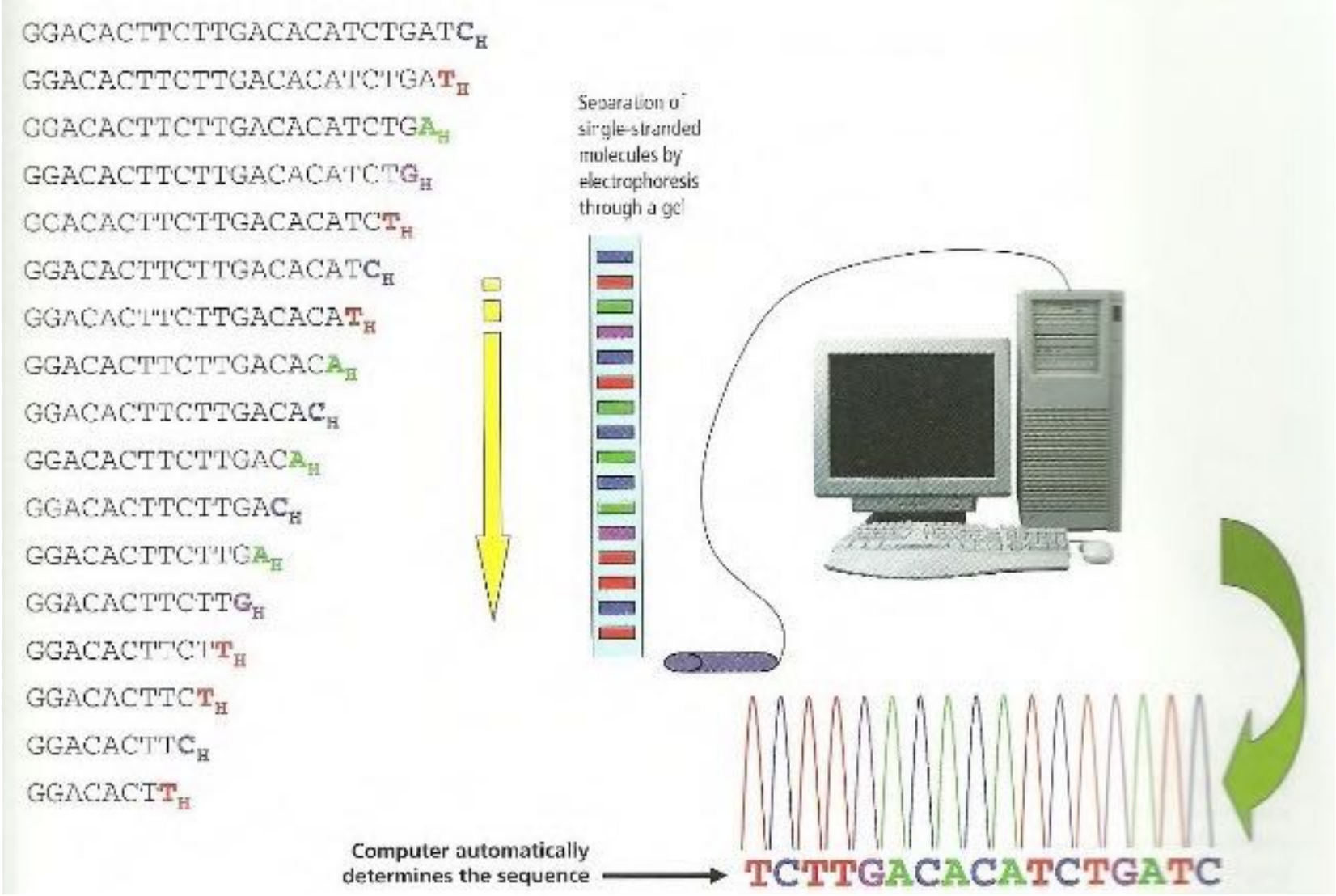


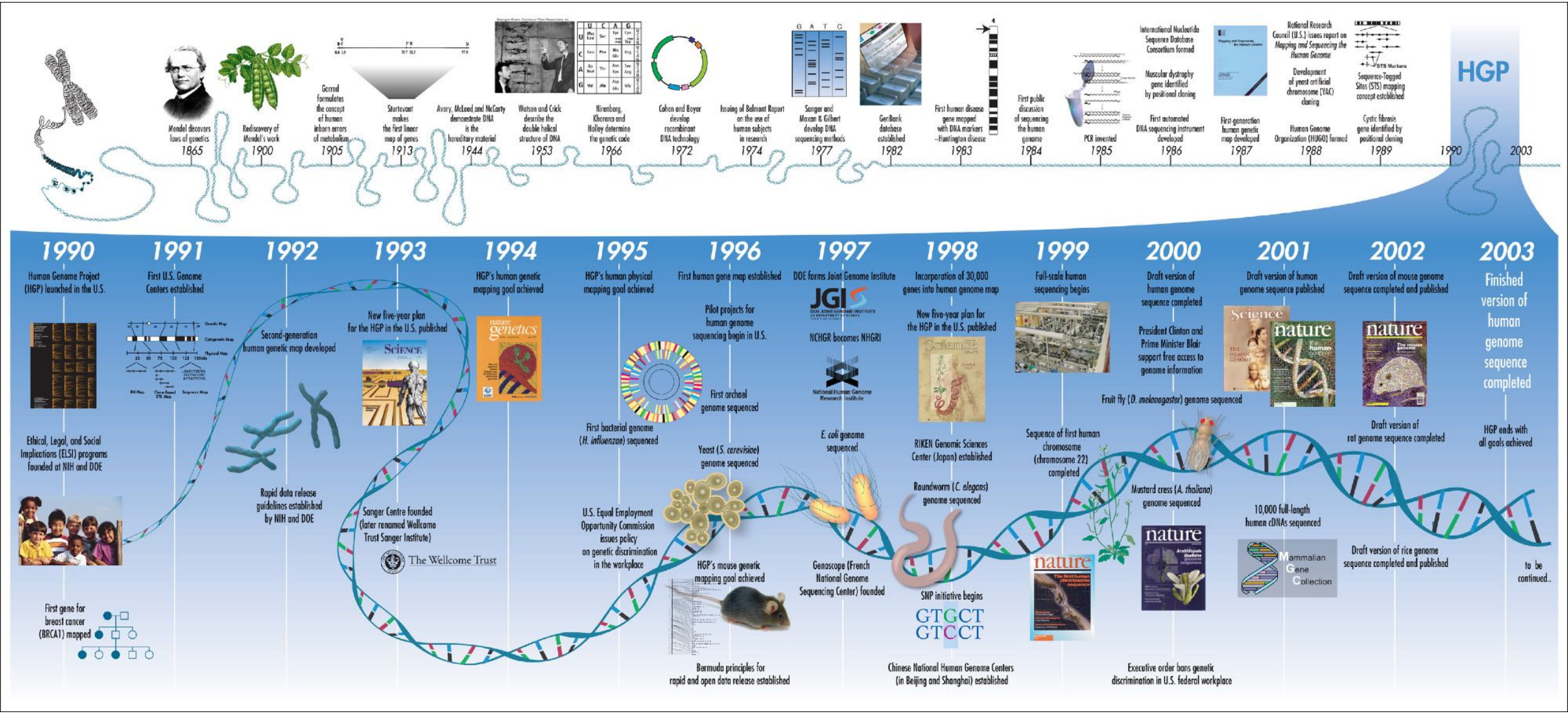
Figure 19-26
Genetics: A Conceptual Approach, Third Edition

(Freeman, 2009)

DYE-TERMINATOR SEQUENCING

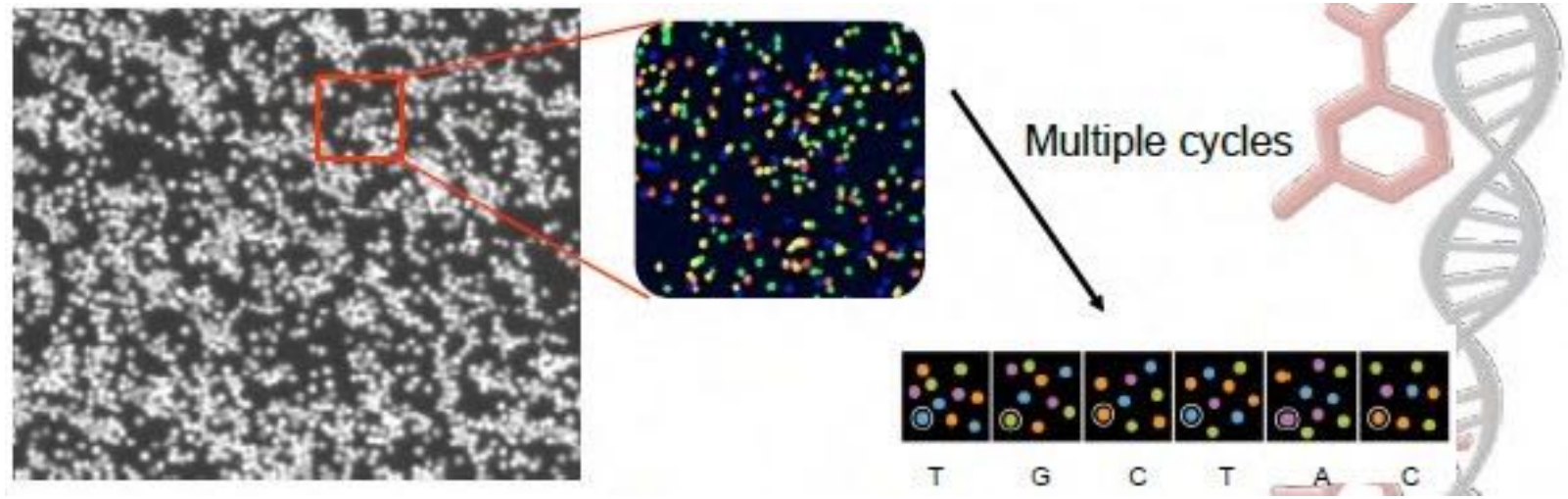


Human Genome Project



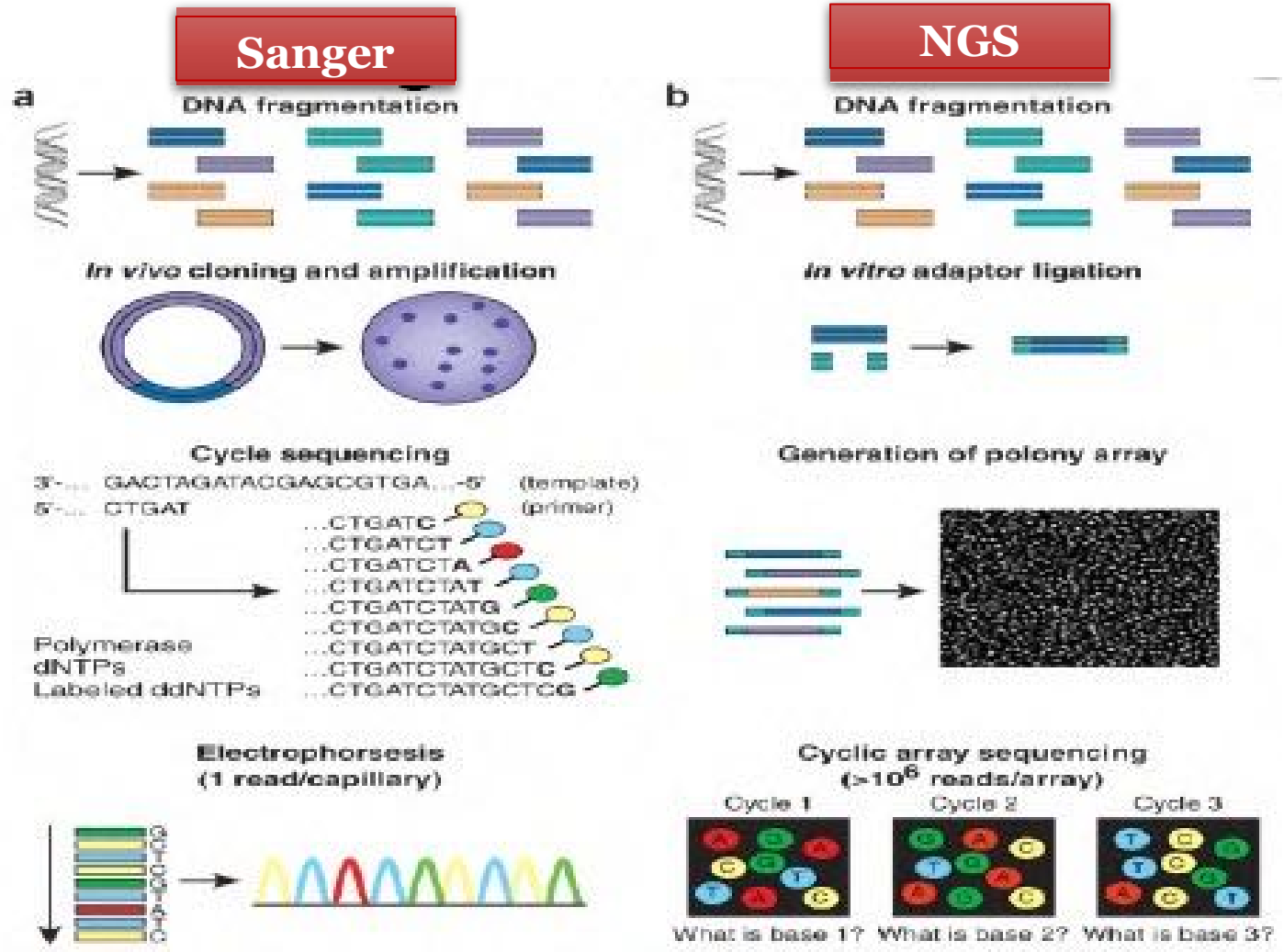
NEXT-GENERATION SEQUENCING

- Employs micro- & nanotechnologies to reduce the size of sample components, reducing reagent costs.
- Highly multiplexed, allowing simultaneous sequencing and analysis of millions of samples.



(Axygen, 2008)

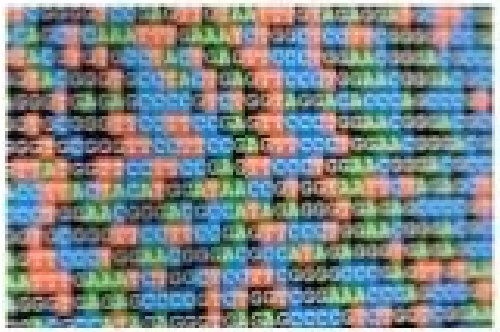
SANGER *vs*. NGS

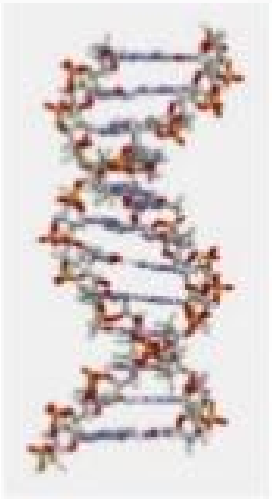


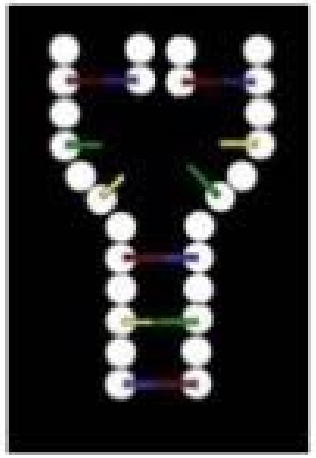
(Axygen, 2008)

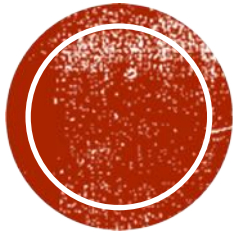
SANGER *vs.* NGS

Features	Sanger	NGS
Sequencing Samples	Clones, PCR	DNA Libraries
Preparation Steps	Few, Sequencing reactions clean up	Many, Complex procedures
Data Collection	Samples in plates : 96, 384	Samples on slides 1-16+
Data	1 Read/ Sample	Thousands & Millions of Reads/ Samples.



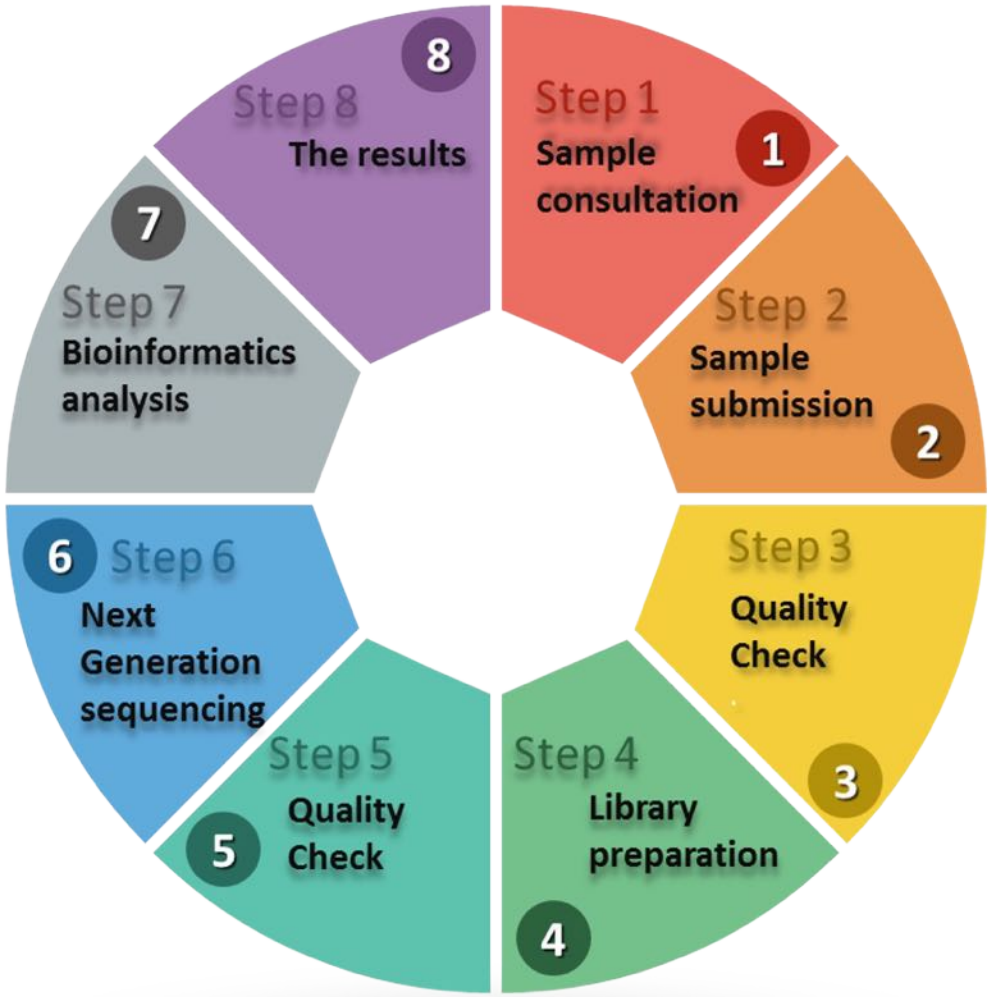






ANALISIS NGS

HOW END-TO-END SOLUTION *NGS* WORKS?



1. Complimentary project consultation and experimental design
3. *Qubit quantification
   *Bioanalyser analysis
5. *Qubit quantification
   *Bioanalyser analysis
8. The results will be reported to the customers with high confidentiality and security.

http://www.ccb.usm.my/index.php/services/next-generation-sequencing

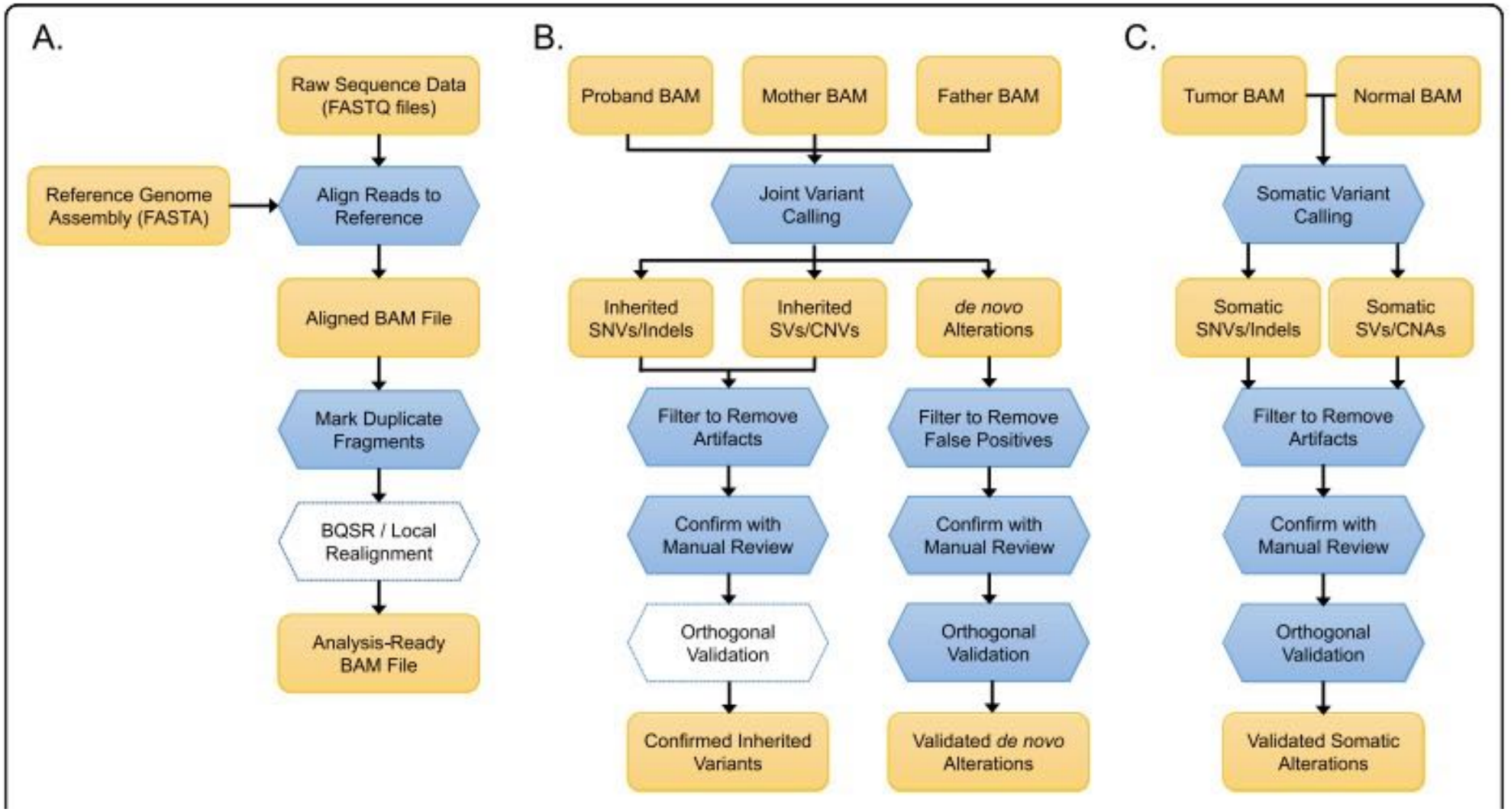
KEY COMPONENTS OF NGS ANALYSIS & A LIST OF EXEMPLAR TOOLS

Strategy	Variant callers
Alignment and pre-processing	
Read alignment	BWA-MEM [25], Bowtie 2 [26], minimap2 [27], Novoalign
Marking duplicates	Picard tools [28], Sambamba [29], SAMBLASTER [30]
BAM file creation	Samtools [31], GATK [19]
Sequencing metrics	BEDTools [32], Picard tools [28], QualiMap 2 [33]
Sample quality control	KING [34], VerifyBamID [35]
Variant calling	
Inherited SNVs/indels	FreeBayes [36], GATK HaplotypeCaller [19], Platypus [20], Samtools/BCFtools [37]
Somatic mutations	deepSNV [38], MuSE [39], MuTect2 [40], SomaticSniper [41], Strelka2 [42], VarDict [43], VarScan2 [44]
Copy number variants	cn.MOPS [45], CONTRA [46], CoNVEX [47], ExomeCNV [48], ExomeDepth [49], XHMM [50]
Structural variants	DELLY [51], Lumpy [52], Manta [53], Pindel [54], SVMerge [55]
Gene fusions (RNA-seq)	fusionCatcher [56], fusionMap [57], mapSplice [58], SOAPfuse [59], STAR-Fusion [60], TopHat-Fusion [61]
Variant review/storage	
Visualization and review	Artemis [62], Integrative Genomics Viewer [63]
VCF/BCF file manipulation	BCFtools [37]

BAM binary alignment/map, *SNV* single nucleotide variant, *VCF* variant call format, *BCF* binary variant call format

(Koboldt1, 2020)

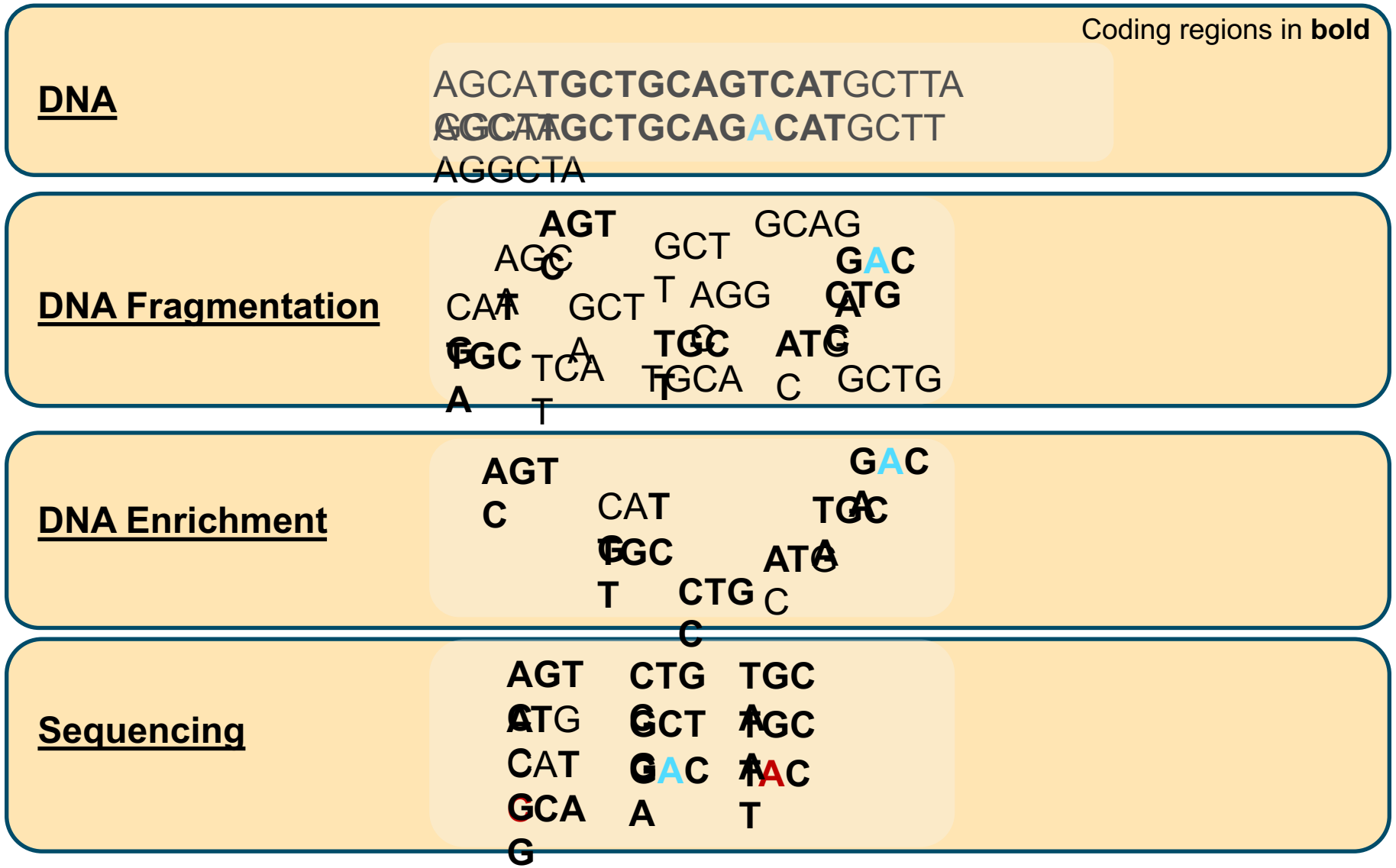
STANDARD PIPELINES FOR NGS ANALYSIS



INDIVIDUAL TRIOS SOMATIC

(Koboldt1, 2020)

HOW DOES IT WORK?



(Gilissen 2022)

HOW DOES IT WORK? ANALYZE

Sequence reads

FASTQ file

Reference

AGCATGCTGCAGTCATGCTTAGGCTA

Alignment / Mapping

Mapped reads

BAM file

Variant calling

Chr5, position 9, G → C
Reads = 5, Variants reads = 1, (20%), score = 10

Chr5, position 13, T → A
Reads = 2, Variants reads = 1, (50%), score = 50

VCF file

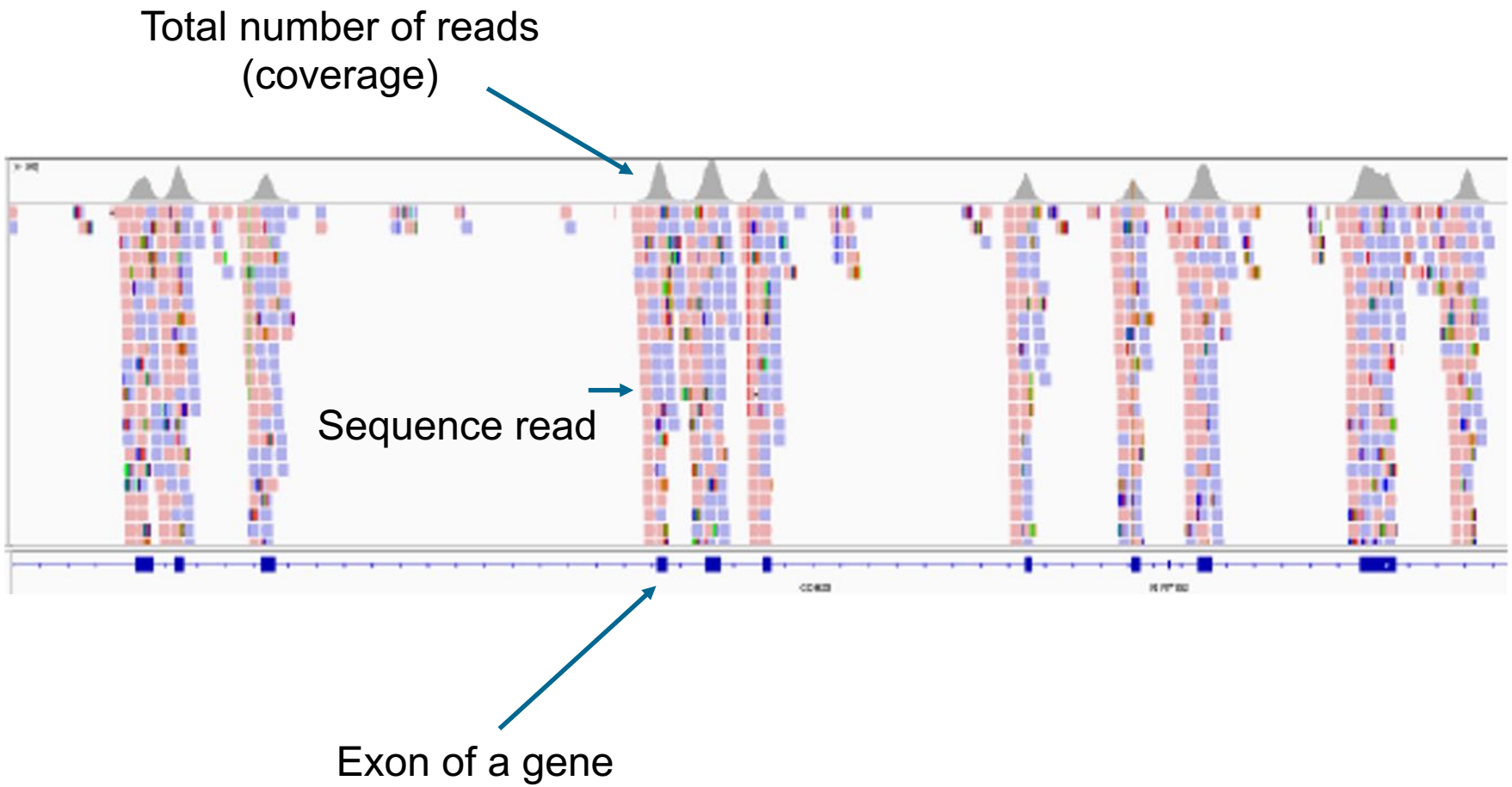
Variant Annotation

Annotated VCF file

Variant Interpretation

(Gilissen 2022)

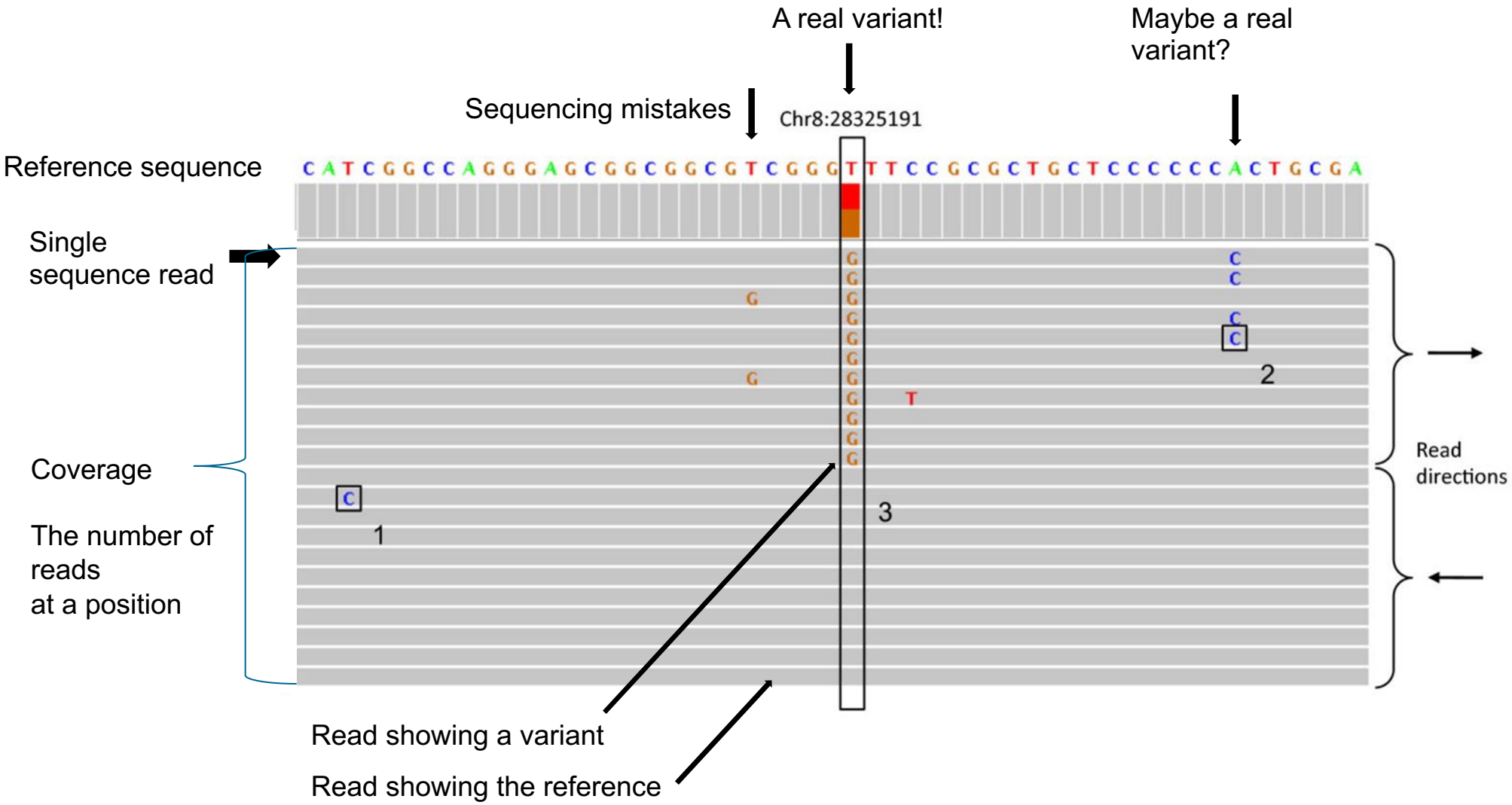
WHAT DO WE GET?



IGV screenshot: https://software.broadinstitute.org/software/igv/

(Gilissen 2022)

SEQUENCING DATA & VARIANTS



(Gilissen 2022)

IGV screenshot: https://software.broadinstitute.org/software/igv/

VARIANTS & ANNOTATION

1	Chromosome	Start position	End position	reference	mutation	reads	variation reads	% variation	Abberation	SNP id	Gene name	Gene component	AminoAcid changes	mRNA changes
2	chr1	888639	888639	T	C	6	6	100	substitution	rs3748596	*NOC2L*	EXON_REGION	E306E	918T>C
3	chr1	897325	897325	G	C	10	10	100	substitution	rs4970441	*KLHL17*	EXON_REGION	A203A	609G>C
4	chr1	909768	909768	A	G	18	16	89	substitution	rs2340593	*PLEKHN1*	INTRON_REGION		
5	chr1	915350	915350	G	A	28	4	14	substitution		*C1orf170*	UTR		
6	chr1	915495	915495	G	A	48	8	17	substitution		*C1orf170*	UTR		
7	chr1	977330	977330	T	C	11	11	100	substitution	rs2799066	*AGRN*	SA_SITE		
8	chr1	982994	982994	T	C	7	7	100	substitution	rs10267	*AGRN*	EXON_REGION	F1186F	3558T>C
9	chr1	1246004	1246004	A	G	84	19	23	substitution	rs2296474	*PUSL1*	SA_SITE		
10	chr1	1334053	1334057	TAGAG		5	5	100	deletion	rs3831366	*CCNL2*	SA_SITE_CANONICAL		
11	chr1	1479333	1479333	A	G	10	3	30	substitution	rs7533	*SSU72*	EXON_REGION	P133P	399A>G
12	chr1	1581096	1581096	C	T	72	27	38	substitution	rs67488456	*CDK11B*	UTR		
13	chr1	1582202	1582202	C	T	63	29	46	substitution	rs71511303	*CDK11B*	UTR		
14	chr1	1586752	1586752	T	C	35	35	100	substitution	rs11486023	*CDK11B*	UTR		
15	chr1	1647745	1647745	G	A	71	15	21	substitution	rs72634830	*CDK11B*	UTR		
16	chr1	1647753	1647753	C	T	88	27	31	substitution	rs74045984	*CDK11B*	UTR		
17	chr1	1647814	1647814	T	C	236	120	51	substitution	rs72901775	*CDK11A*	EXON_REGION	E142E	324T>C
18	chr1	1647871	1647871	T	C	134	76	57	substitution	rs72909014	*CDK11A*	EXON_REGION	R123R	267T>C
19	chr1	1650787	1650787	T	C	107	51	48	substitution	rs1137003	*CDK11A*	EXON_REGION	H112R	335T>C
20	chr1	1650797	1650797	A	G	115	57	50	substitution	rs1059830	*CDK11A*	EXON_REGION	C109R	325A>G
21	chr1	1650807	1650807	T	C	202	60	30	substitution	rs1137005	*CDK11A*	EXON_REGION	R105R	213T>C
22	chr1	1650832	1650832	A	G	380	236	62	substitution	rs72909030	*CDK11A*	EXON_REGION	V97A	188A>G
23	chr1	1650845	1650845	G	A	381	126	33	substitution	rs1059831	*CDK11A*	EXON_REGION	R93W	175G>A
24	chr1	1653028	1653028	C	T	41	40	98	substitution	rs16825265	*CDK11B*	UTR		
25	chr1	1887019	1887019	A	G	24	23	96	substitution	rs28548017	*KIAA1751*	EXON_REGION	*763Q	2287A>G
26	chr1	1957037	1957037	T	C	8	8	100	substitution	rs2229110	*GABRD*	EXON_REGION	G110G	330T>C

(Gilissen 2022)

SEQUENCE COVERAGE

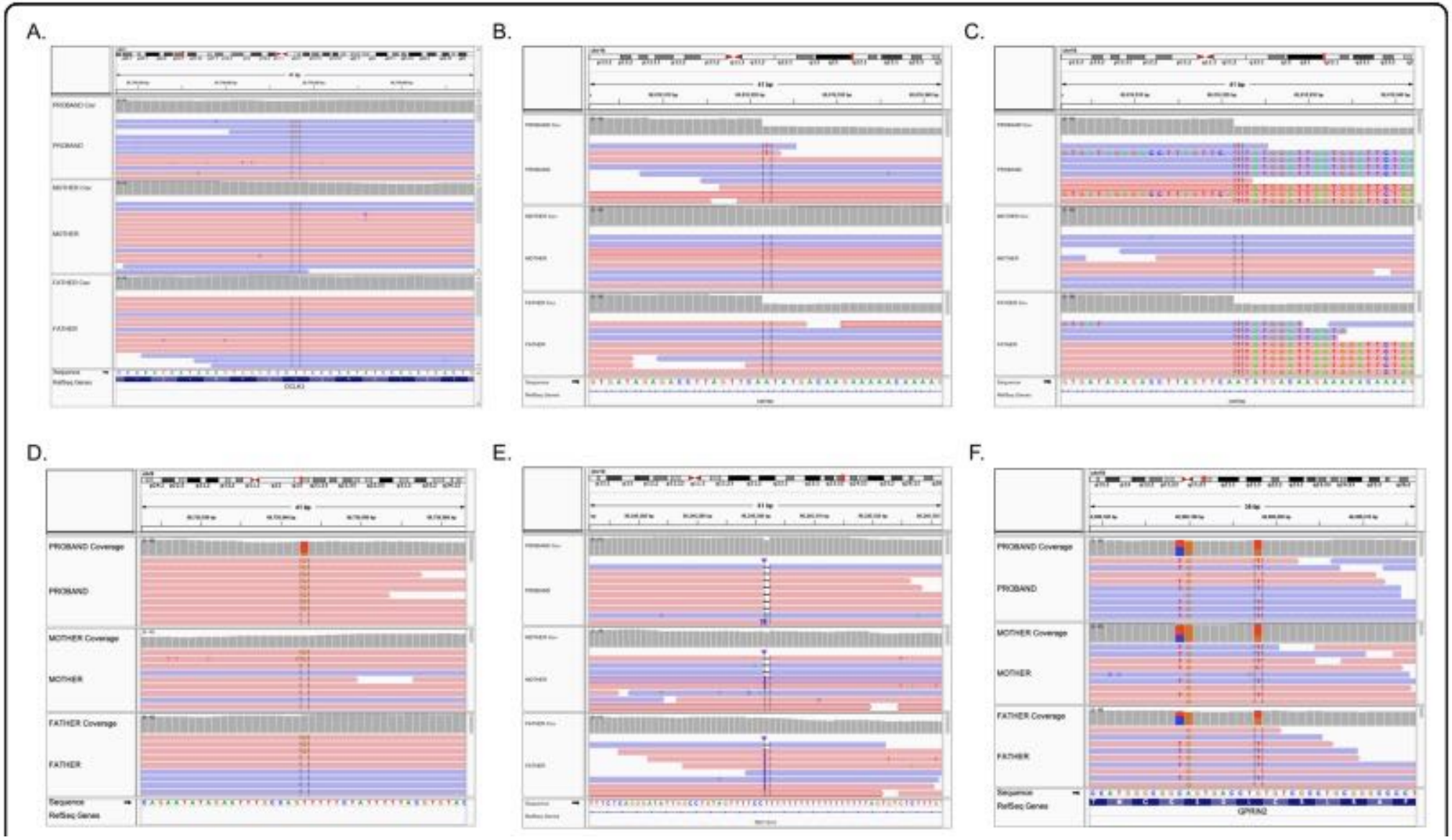
Coverage: the number of times a single position on the DNA has been interrogated by a sequence reads

Why do we need coverage?

1. There are two alleles that we sequence at random: we need to be sure that we see each allele at least once.
2. We want to be able to distinguish variants from sequencing errors.
3. Some regions don't enrich very well, if we sequence more, we will have a higher chance of sequencing these regions as well.eor

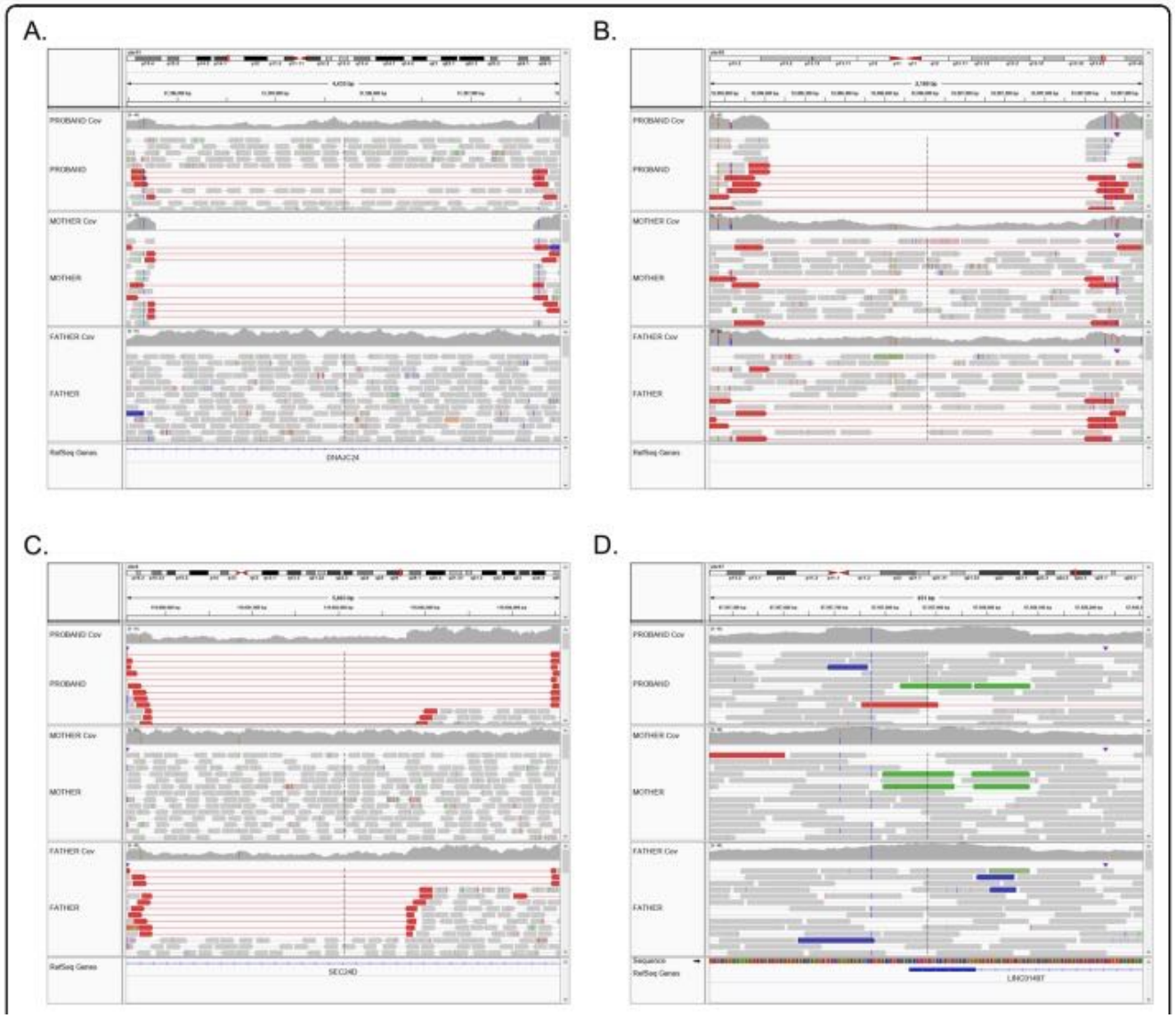
(Gilissen 2022)

COMMON ARTIFACTS IN *NGS* ALIGNMENT



(Koboldt1, 2020)

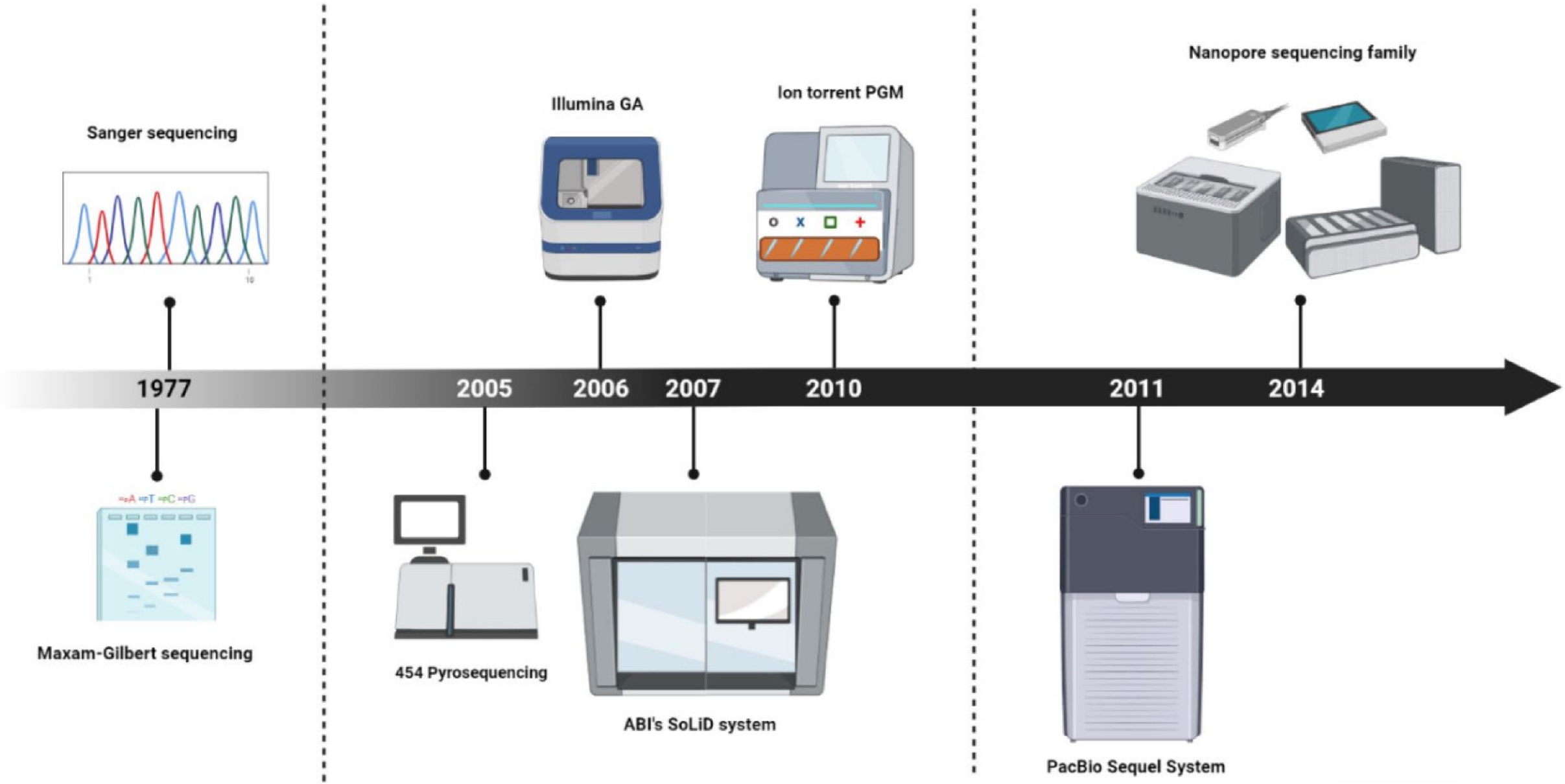
COPY NUMBER (CNV) & STRUCTURAL VARIANTS (SV)



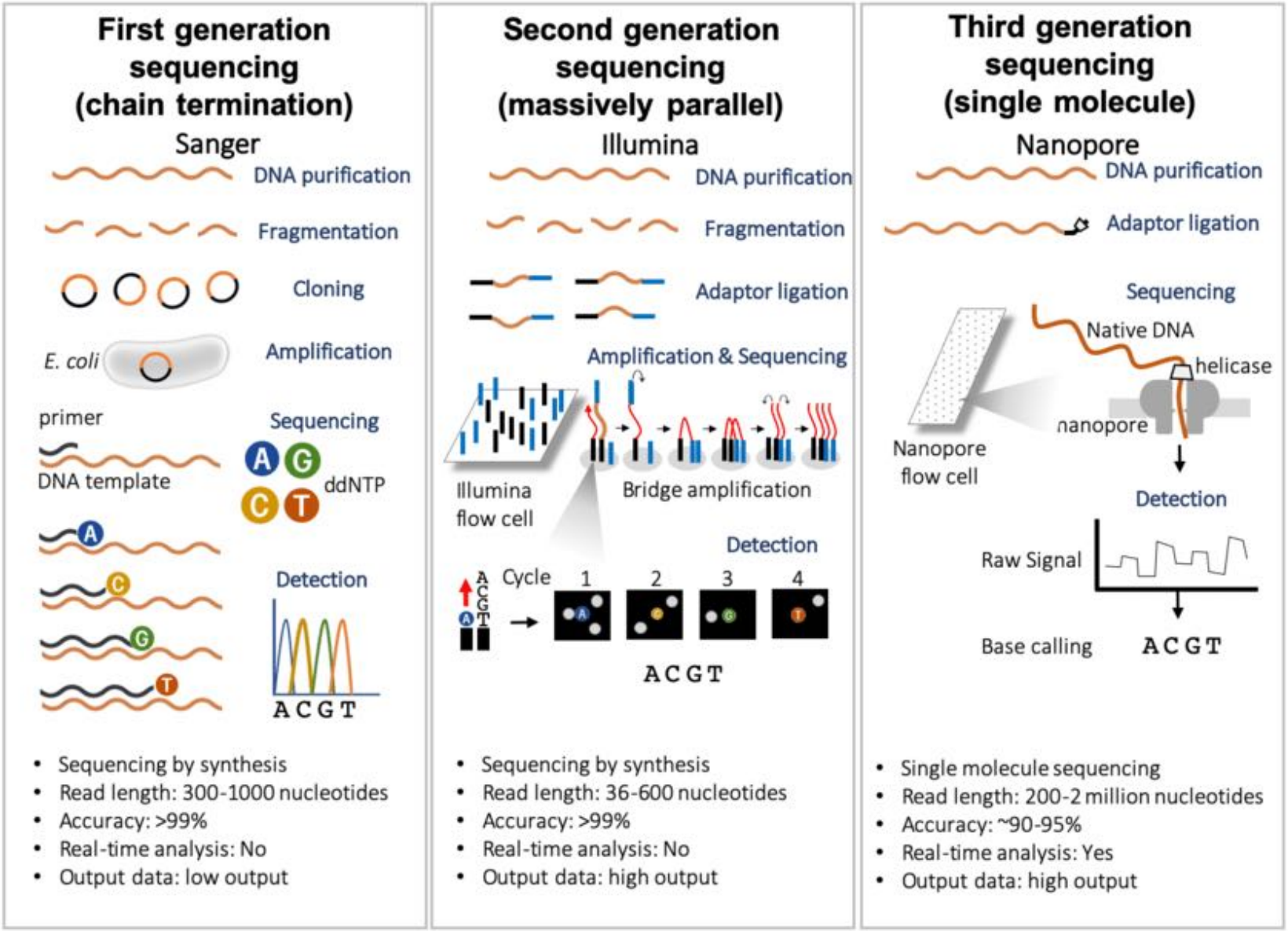
(Koboldt1, 2020)

APLIKASI NGS

Sequencing Advancement



Sequencing Advancement



Sequencing Advancement

1 Sanger Sequencing

Slow but good base and accurate structure

Reads each single base one at a time

2 Amplified Molecule Sequencing

Fast reads but only good base accuracy

Reads 300 base long fragments

Processes millions of these fragments at same time

illumina

ThermoFisher SCIENTIFIC

life technologies

Roche

3 Single Molecule Sequencing

Fast reads but only good structural accuracy

Reads 5,000 base long fragments

Processes hundreds of fragments at same time

10x GENOMICS

PACBIO

Oxford NANOPORE technologies

4 New Technologies

Fast reads with good base and structural accuracy

Reads 100,000 base long fragments

New Technologies in Development

Sequencing Advancement

First generation

Sanger sequencing
Maxam and Gilbert
Sanger chain termination

Infer nucleotide identity using dNTPs, then visualize with electrophoresis

500–1,000 bp fragments

Second generation
(next generation sequencing)

454, Solexa,
Ion Torrent,
Illumina

High throughput from the parallelization of sequencing reactions

~50–500 bp fragments

Third generation

PacBio
Oxford Nanopore

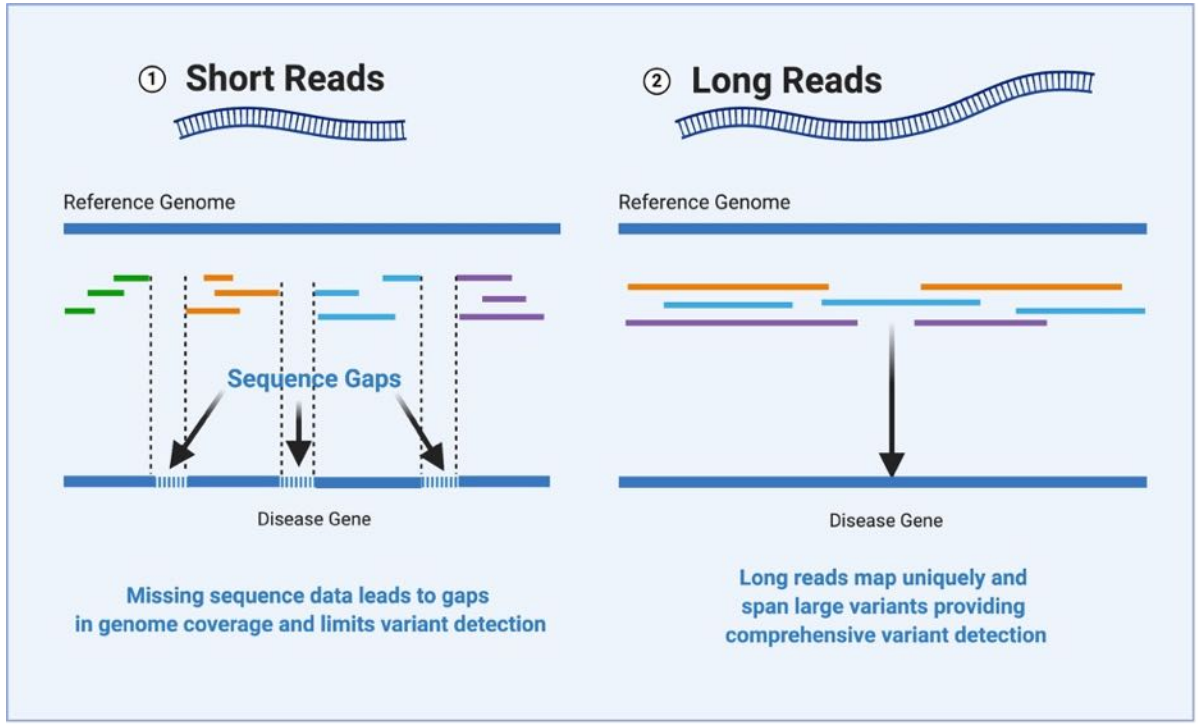
Sequence native DNA in real time with single-molecule resolution

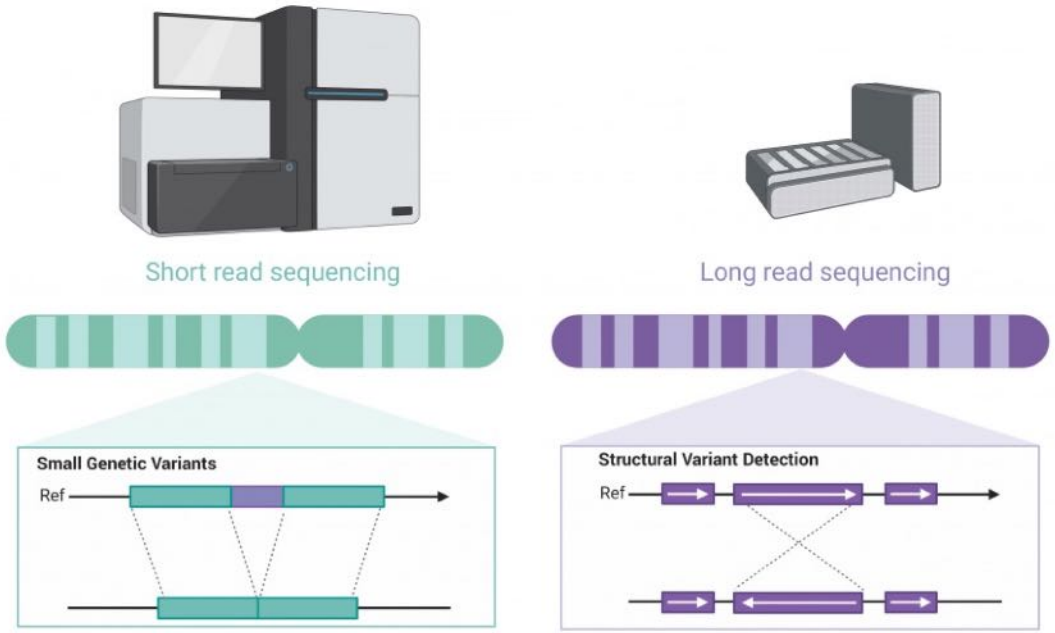
Tens of kb fragments, on average

Short-read sequencing

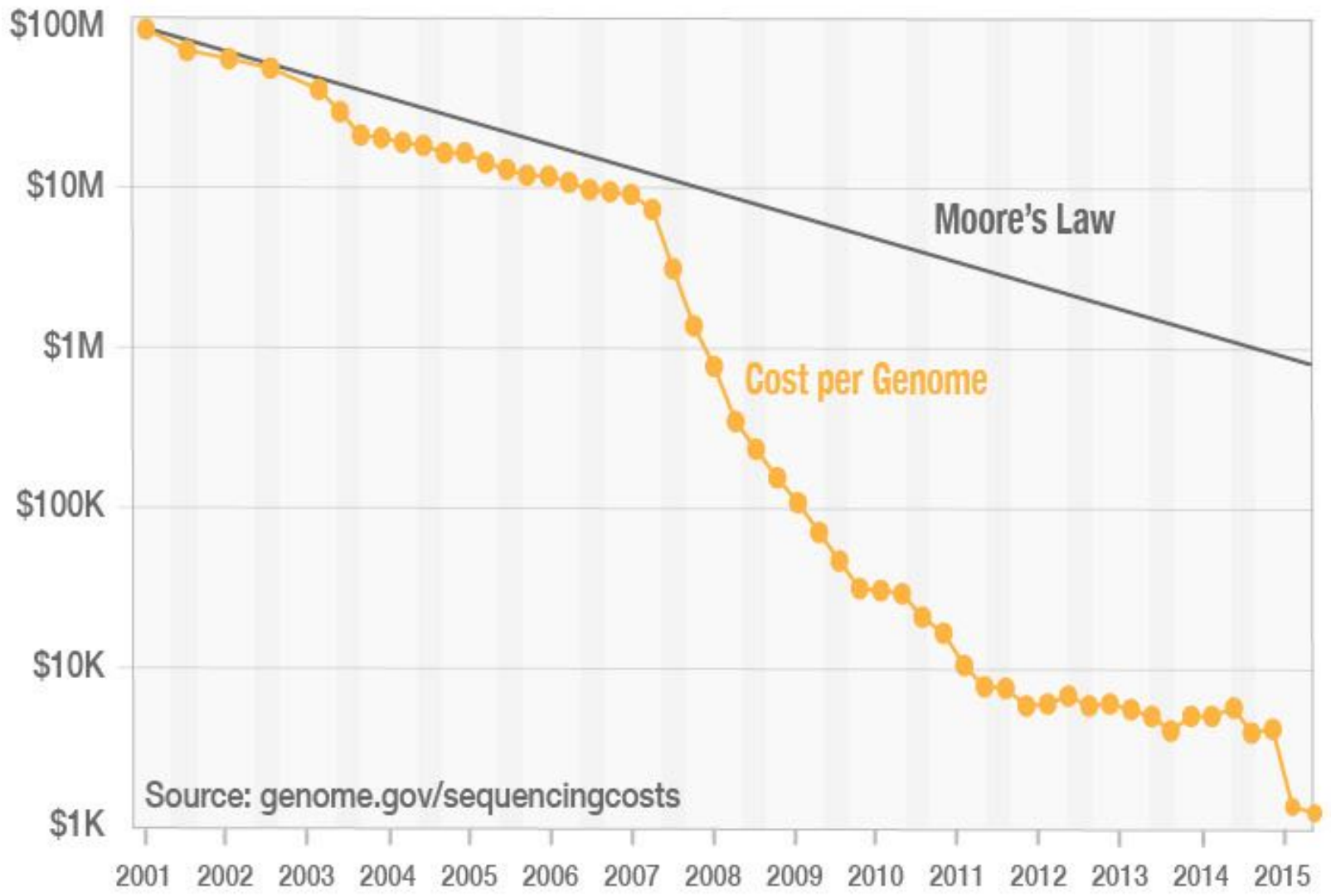
Long-read sequencing

SHORT *vs.* LONG READS SEQ

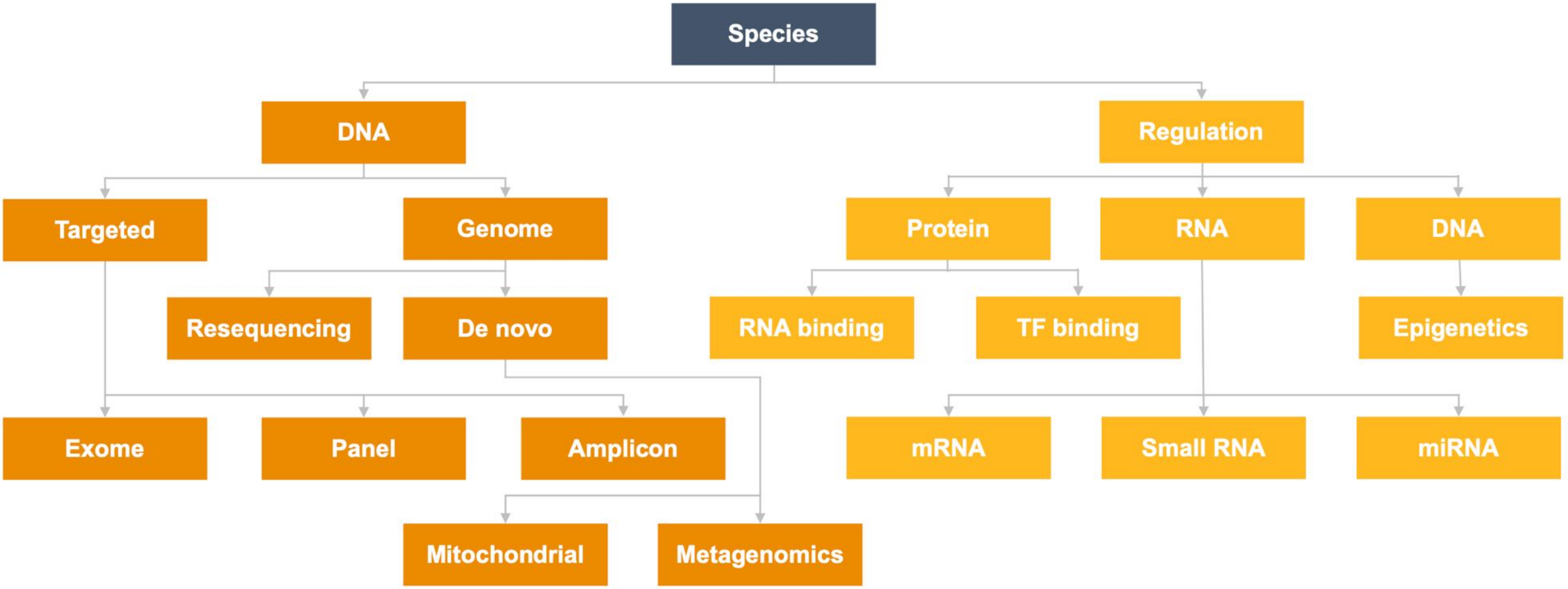




Cost Genome Seq Declining & More Affordable



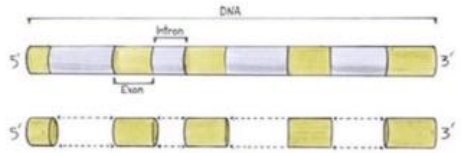
WHAT ARE *NGS* APPLICATIONS?

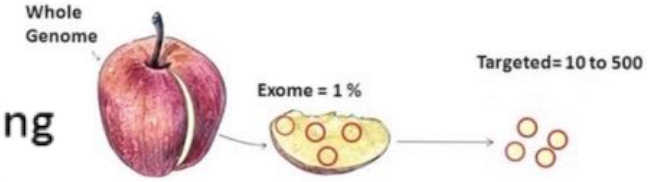


TF = Transcription Factor

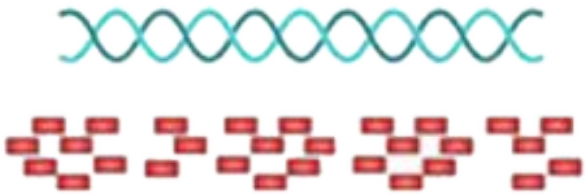
1

DNASeq



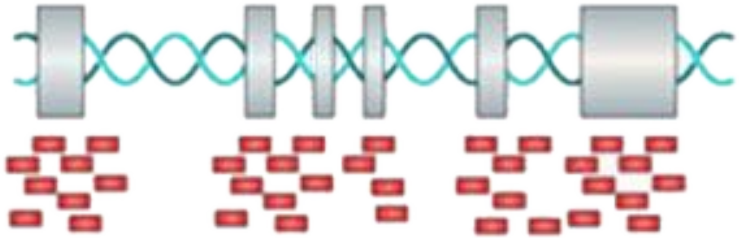


Whole genome sequencing



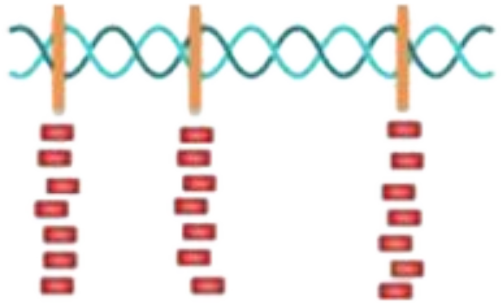
- Sequencing region : whole genome
- Sequencing Depth: >30X
- Covers everything – can identify all kinds of variants including SNPs, INDELs and SV.

Whole exome sequencing



- Sequencing region: whole exome
- Sequencing Depth : >50X ~ 100X
- Identify all kinds of variants including SNPs, INDELs and SV in coding region.
- Cost effective

Targeted sequencing



- Sequencing region: specific regions (could be customized)
- Sequencing Depth : >500X
- Identify all kinds of variants including SNPs, INDELs in specific regions
- Most Cost effective

http://www.genomesop.com/somatic-mutations/

DNASeq

Whole-exome or whole-genome sequencing → becoming popular since they can capture the gene- or genome-level genomic alterations.

Target-sequencing covers several hundreds genes whose alterations are well known for their roles in cancer pathology & targetable by chemical/immunologic agents.

In cancer patients, target or panel sequencing has started to produce clinically relevant findings such as actionable mutations for patients

SEQUENCING STRATEGIES FOR NGS & EMPIRICAL VARIANT DETECTION SENSITIVITY

Strategy	Panel	Exome	Genome
Size of target space (Mbp)	~ 0.5	~ 50	~ 3200
Average read depth	500–100x	100–150x	~ 30–60x
Relative cost	$	$$	$$$
SNV/indel detection	++	++	++
CNV detection	+	+	++
SV detection	–	–	+
Low VAF	++	+	+

(Koboldt1, 2020)

WGS OFFERS THE BROADEST SCOPE OF DETECTABLE VARIANTS TO PROVIDE COMPREHENSIVE VARIANT ANALYSIS

Current testing options	SNVs and indels	CNVs	Repeat expansions	Structural variants	Mitochondrial variants
Sanger	●				●
Targeted NGS	●	LIMITED			●
PCR	●	●	●		
FISH		●		●	
Karyotype				●	
CMA		●			
WES	●	LIMITED		LIMITED	●
WGS	●[1]	●[1]	●[2]	●[3]	●[1]

CMA=chromosomal microarray; CNV=copy number variant; FISH=fluorescence in situ hybridization; Indel=small insertion/deletion; NGS=next-generation sequencing; PCR=polymerase chain reaction; SNV=single nucleotide variant; WES=whole-exome sequencing; WGS=whole-genome sequencing.

References: 1. Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med*. 2018 Apr;20(4):435-443. doi: 10.1038/gim.2017.119. Epub 2017 Aug 3. **2.** Dolzhenko E, van Vugt JJFA, Shaw RJ, et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res*. 2017;27(11):1895-1903. doi: 10.1101/gr.225672.117. **3.** Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 2016;32(8):1220–1222. http://doi.org/10.1093/bioinformatics/btv710.

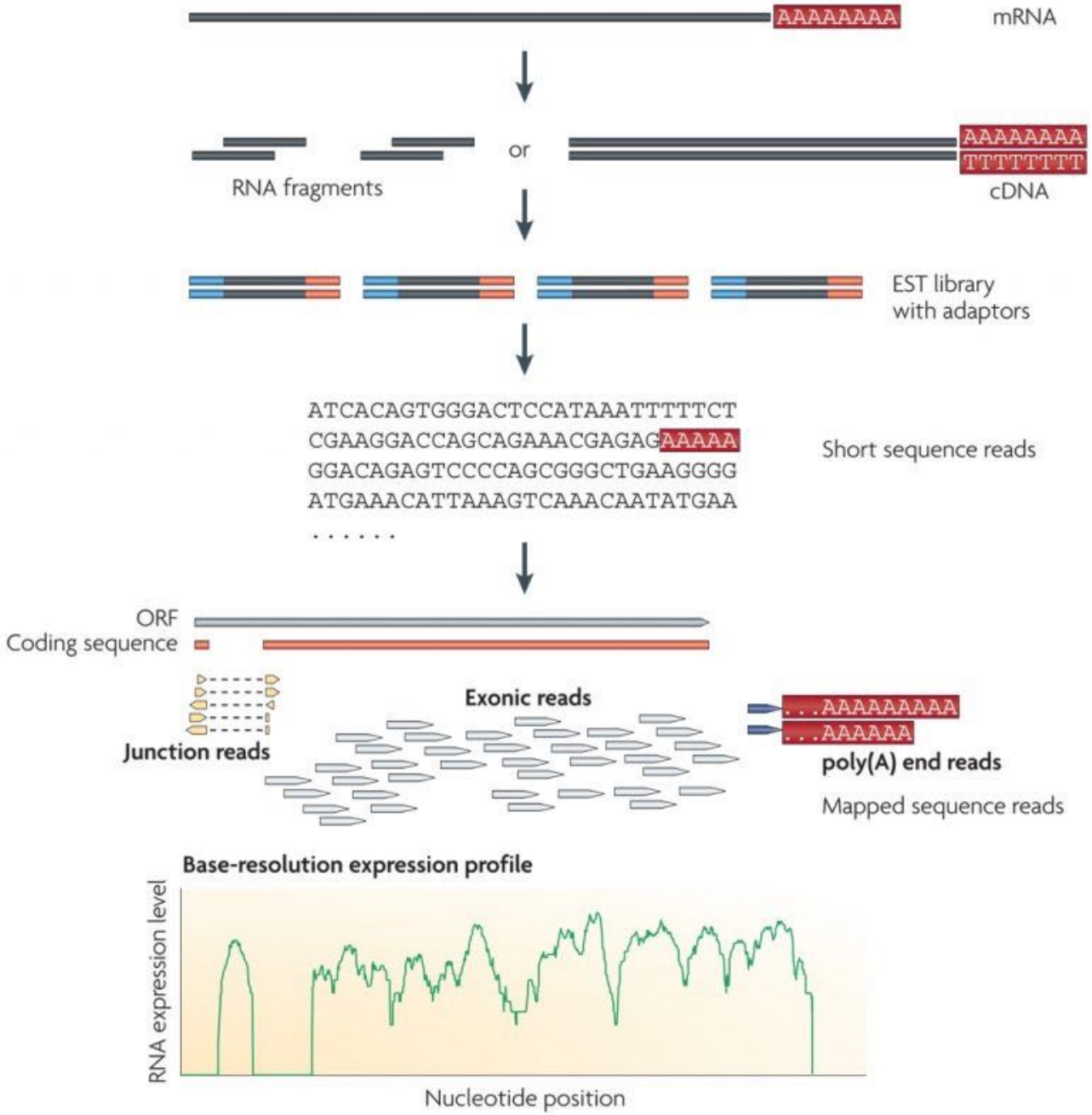
2 RNASeq

RNA-sequencing (Whole-transcriptome sequencing, WTS, transcriptomic) can quantify level of individual genes & identify structural changes in the transcriptome.

These include the gene fusions and alternative splicing that also play roles in the development of cancers.

http://www.genomesop.com/somatic-mutations/

RNASeq

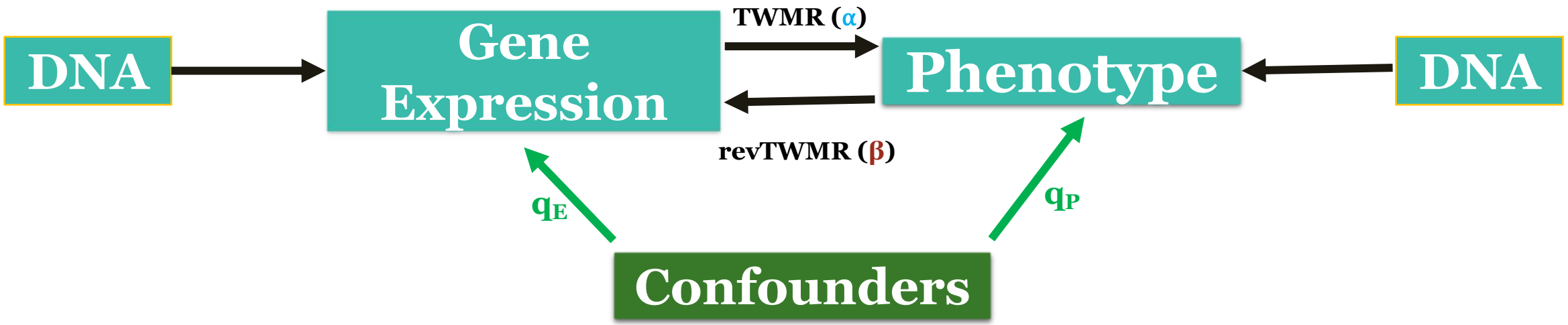


http://www.genomesop.com/somatic-mutations/

3 NGS-BASED ASSAYS TO MAP EPIGENETIC MARKS

Broad epigenetic features	Types of marks (if applicable)	Assays
DNA methylation	5-mC: methyl cytosine Variants 5-hmC: hydroxyl methyl cytosine 5-fC: formyl cytosine 5-caC: carboxyl cytosine	Restriction based: MRE-seq [15] Affinity based: MeDIP-seq [16] and MBD-seq [17] Chemical based: RRBS [18] and WGBS/methylC-seq [19] oxBS-seq (to distinguish between 5-mC and 5-hmC) [20]
Histone modifications	H3K27me3: associated with repressed regions H3K4me1: associated with enhancers H3K4me3: associated with promoters H3K27ac: associated with active enhancers H3K9ac: associated with active promoters H3K36me3: associated with gene bodies H3K9me3: associated with heterochromatin	ChIP-seq [21] ChIP-exo [22]
Nucleosome positioning and occupancy	-	MNase-seq [23] MNase-independent mutated hisotones based mapping [24]
Chromatin accessibility	-	DNase-seq [25] DGF [26] FAIRE-seq [27]
3D chromatin structure	-	3C [28] 4C [29] 5C [30] Hi-C [31] ChIA-PET [32]
Non-coding RNA localization	lncRNA smRNA (siRNA, piRNA, miRNA)	Deep sequencing of transcriptomes by mRNA-seq [33] smRNA-seq [34]

EPIGENETIC → COMPOSITION OF CORRELATION



r = corr (Phenotype, Gene Expression) = α + β + q_E*q_P

ESHG 2021; transcriptome-wide summary statistics-based Mendelian Randomization approach (TWMR); https://www.nature.com/articles/s41467-019-10936-0

MILESTONE CoE Human Genomics & Genetics

2007

Sanger Sequencing
- Complex Genetic Disorder (HSCR) & Mendelian (SMA)

2017

Next-Generation Sequencing (NGS)
Whole-exome seq (WES)
Transcriptomic (HSCR)
Bioinformatics (Mendelian)

2020

COVID-19 Pandemic
Whole-genome seq (WGS)
SARS-CoV-2

2021

Keputusan Menteri Kesehatan
No. HK 01.07/MENKES/4842/2021
Jejaring Lab Surveilans Genom
SARS-CoV-2

Illumina Next-Seq 550

2022 ~

Keputusan Menteri Kesehatan
No. HK 01.07/MENKES/1141/2022
BGSi for Precision Medicine
RSUP Dr Sardjito → *Genetic Disorder Hub*

Precision Medicine

1 SARS-CoV-2 Genomic Surveillance

PeerJ

Full-length genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java, Indonesia

Gunadi[1], Hendra Wibawa[2], Marcellus[1], Mohamad Saifudin Hakim[3], Edwin Widyanto Daniwijaya[4], Ludhang Pradipta Rizki[3], Endah Supriyati[5], Dwi Aris Agung Nugrahaningsih[6], Afiahayati[7], Siswanto[8], Kristy Iskandar[9], Nungki Anggorowati[10], Alvin Santoso Kalim[1], Dyah Ayu Puspitarani[1], Kemala Athollah[1], Eggi Arguni[11], Titik Nuryastuti[3] and Tri Wibawa[3]

Gunadi et al. BMC Med Genomics (2021) 14:144
https://doi.org/10.1186/s12920-021-00990-3

BMC Medical Genomics

RESEARCH Open Access

Molecular epidemiology of SARS-CoV-2 isolated from COVID-19 family clusters

Gunadi[1*†], Hendra Wibawa[2†], Mohamad Saifudin Hakim[3], Marcellus[4], Ika Trisnawati[5], Riat El Khair[6], Rina Triasih[7], Irene[8], Afiahayati[9], Kristy Iskandar[10], Siswanto[11], Nungki Anggorowati[12], Edwin Widyanto Daniwijaya[13], Endah Supriyati[14], Dwi Aris Agung Nugrahaningsih[15], Eko Budiono[5], Heni Retnowulan[5], Yunika Puspadewi[6], Ira Puspitawati[6], Osman Sianipar[6], Dwiki Afandy[4], Susan Simanjaya[4], William Widitjiarso[4], Dyah Ayu Puspitarani[4], Fadil Fahri[4], Untung Riawan[4], Aditya Rifqi Fauzi[4], Alvin Santoso Kalim[4], Nur Rahmi Ananda[5], Amalia Setyati[7], Dwikisworo Setyowireni[7], Ida Safitri Laksanawati[7], Eggi Arguni[7], Titik Nuryastuti[3] and Tri Wibawa[3] on behalf of the Yogyakarta-Central Java COVID-19 study group

2021
BMC Medical Genomics
IF: 3.063
Family

2020
PeerJ
IF: 2.98
D614G

2021
Scientific Reports
IF: 4.379
Multiple S protein mutations

www.nature.com/scientificreports

scientific reports

OPEN

Association between prognostic factors and the outcomes of patients infected with SARS-CoV-2 harboring multiple spike protein mutations

Gunadi[2*], Mohamad Saifudin Hakim[2], Hendra Wibawa[3], Marcellus[4], Ika Trisnawati[5], Endah Supriyati[6], Afiahayati[7], Riat El Khair[8], Kristy Iskandar[9], Siswanto[10], Irene[11], Nungki Anggorowati[12], Edwin Widyanto Daniwijaya[13], Dwi Aris Agung Nugrahaningsih[14], Yunika Puspadewi[6], Susan Simanjaya[4], Dyah Ayu Puspitarani[6], Hana Fauzyyah Hanifin[4], Alvina Alexandra Setiawan[4], Irene Tania[4], Cita Shafira Amalia[4], I. Putu Aditio Artayasa[4], Haries Rachman[4], Herdiyanto Mulyawan[4], Nur Rahmi Ananda[5], Eggi Arguni[15], Titik Nuryastuti[2] & Tri Wibawa[2]

2021
Frontiers in Medicine
IF: 5.091
Delta

frontiers in Medicine

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fmed.2021.780611

Is the Infection of the SARS-CoV-2 Delta Variant Associated With the Outcomes of COVID-19 Patients?

Gunadi[1], Mohamad Saifudin Hakim[2], Hendra Wibawa[3], Marcellus[1], Vivi Setiawaty[4], Siamet[4], Ika Trisnawati[5], Endah Supriyati[6], Riat El Khair[7], Kristy Iskandar[8], Afiahayati[9], Siswanto[10], Irene[11], Nungki Anggorowati[12], Edwin Widyanto Daniwijaya[3], Dwi Aris Agung Nugrahaningsih[13], Yunika Puspadewi[7], Dyah Ayu Puspitarani[1], Irene Tania[1], Khanza Adzkia Vujira[1], Muhammad Buston Ardiyamustaqim[1], Gita Christy Gabriela[1], Laudria Stella Eryvinka[1], Bunga Citta Nirmala[1], Esensi Tarian Geometri[1], Abirafdi Amajida Darutama[1], Anisa Adityarini Kuswandani[1], Lestari[3], Sri Handayani Irianingsih[3], Siti Khoiriyah[14], Ina Lestari[14], Nur Rahmi Ananda[5], Eggi Arguni[6], Titik Nuryastuti[2] and Tri Wibawa[2] on behalf of the Yogyakarta-Central Java COVID-19 Study Group*

[1] Pediatric Surgery Division, Department of Surgery/Genetics Working Group, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia, [2] Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia, [3] Disease Investigation Center, Ministry of Agriculture Indonesia, Yogyakarta, Indonesia, [4] National Institute of Health Research and Development, Ministry of Health, Jakarta, Indonesia, [5] Pulmonology Division, Department of Internal Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/Dr. Sardjito Hospital, Yogyakarta, Indonesia, [6] Centre of Tropical Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia, [7] Department of Clinical Pathology and Laboratory Medicine,

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

Comparative analysis of the outcomes of COVID-19 between patients infected with SARS-CoV-2 Omicron and Delta variants: a retrospective cohort study

Gunadi, Mohamad Saifudin Hakim, Hendra Wibawa, Khanza Adzkia Vujira, Dyah Ayu Puspitarani, Endah Supriyati, Ika Trisnawati, Kristy Iskandar, Riat El Khair, Afiahayati, Siswanto, Yunika Puspadewi, Irene, Sri Handayani Irianingsih, Edwin Widyanto Daniwijaya, Dwi Aris Agung Nugrahaningsih, Gita Christy Gabriela, Esensi Tarian Geometri, Laudria Stella Eryvinka, Fadila Dyah Trie Utami, Edita Mayda Devana, Lanang Aditama, Nathania Christi Putri Kinasih, Verrell Christopher Amadeus, Yekti Hediningsih, Nur Rahmi Ananda, Eggi Arguni, Titik Nuryastuti, Tri Wibawa the Yogyakarta-Central Java COVID-19 study group

doi: https://doi.org/10.1101/2022.04.30.22274532

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

2022
Preprint
Omicron vs. Delta

2022
Genes
IF: 4.141
Bioinforma pipelines

Congratulations on the acceptance of your manuscript, and thank you for submitting your work to Genes:

Manuscript ID: genes-1818069
Type of manuscript: Article
Title: A Comparison of Bioinformatics Pipelines for Enrichment Illumina Next Generation Sequencing Systems in Detecting SARS-CoV-2 Virus Strains
Authors: FNU Afiahayati *, Stefanus Bernard, FNU Gunadi, Hendra Wibawa, Mohamad Saifudin Hakim, FNU Marcellus, Arli Aditya Parikesit, Chandra Kusuma Dewa, Yasubumi Sakakibara
Received: 30 June 2022

OPEN ACCESS

Edited by:
Bin Luo,
Lanzhou University, China

Reviewed by:
Nur Izzah Ismail,
The Chinese University of Hong Kong, China
Zongliang Yuan,
Southeast University, China

**Correspondence:*
Gunadi

HEPATITIS UNKNOWN ETIOLOGIES

CASE#1

Viral DNA material detected

Virus name	Detected
Adenoviruses (HAdV)	No
Herpes simplex virus 1 (HSV-1)	No
Herpes simplex virus 2 (HSV-2)	No
Varicella zoster virus (VZV)	No
Epstein-Barr virus (EBV)	No
Cytomegalovirus (CMV)	Yes
Human herpesvirus 6 (HHV-6)	No
Human herpesvirus 7 (HHV-7)	No

NUSANTICS

Viral DNA material composition

NUSANTICS

Bacterial DNA material composition

Component	Ratio (compared to healthy samples)
Fold change total bacterial DNA load	290%
Proteobacteria	24%
Actinobacteria	7%
Bacteroidetes	15%
Firmicutes	23%

Less than that of healthy samples

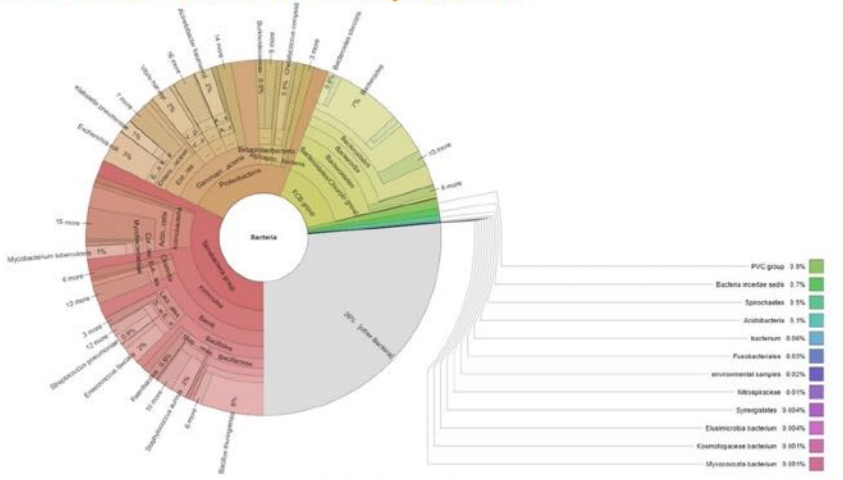
More than that of healthy samples

Similar to that of healthy samples

This bacterial composition indicates bacterial dysbiosis and bacteremia.

NUSANTICS

Bacterial DNA material composition



NUSANTICS

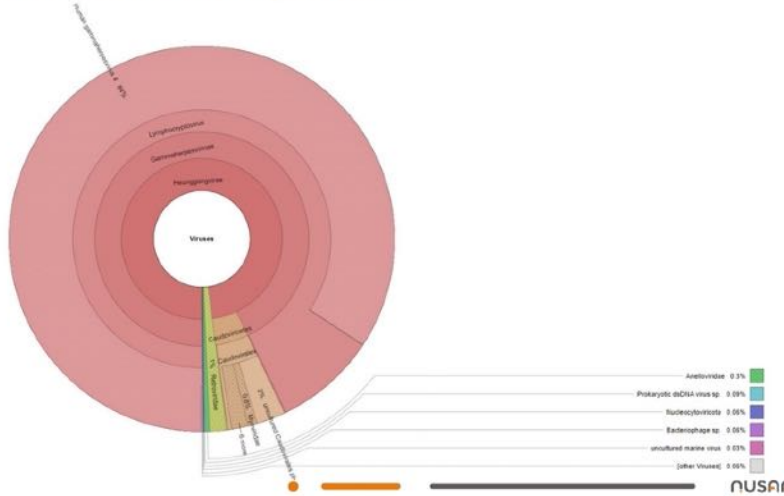
CASE#2

Viral DNA material detected

Virus name	Detected
Adenoviruses (HAdV)	No
Herpes simplex virus 1 (HSV-1)	No
Herpes simplex virus 2 (HSV-2)	No
Varicella zoster virus (VZV)	No
Epstein-Barr virus (EBV)	Yes
Cytomegalovirus (CMV)	No
Human herpesvirus 6 (HHV-6)	No
Human herpesvirus 7 (HHV-7)	No

NUSANTICS

Viral DNA material composition



NUSANTICS

Bacterial DNA material composition

Component	Ratio (compared to healthy samples)
Fold change total bacterial DNA load	36%
Proteobacteria	79%
Actinobacteria	2%
Bacteroidetes	4%
Firmicutes	3%

Less than that of healthy samples

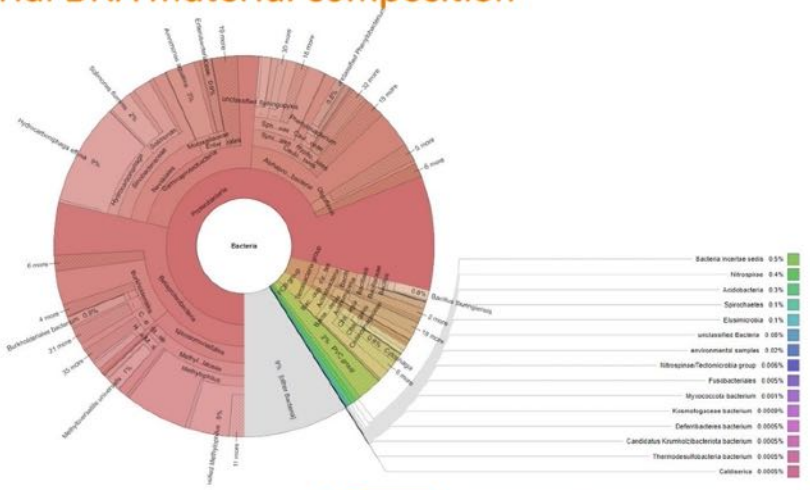
More than that of healthy samples

Similar to that of healthy samples

This bacterial composition indicates normal bacterial composition.

NUSANTICS

Bacterial DNA material composition

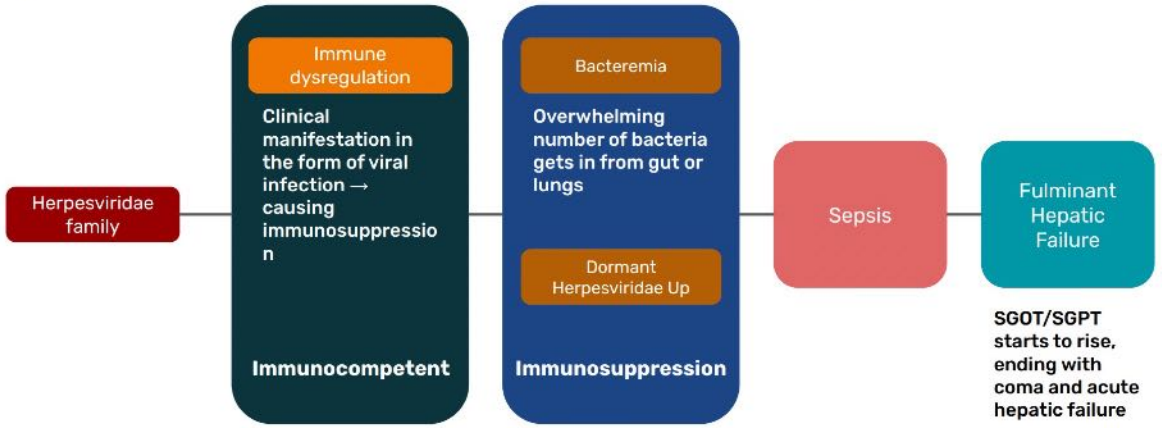


NUSANTICS

HYPOTHESIS

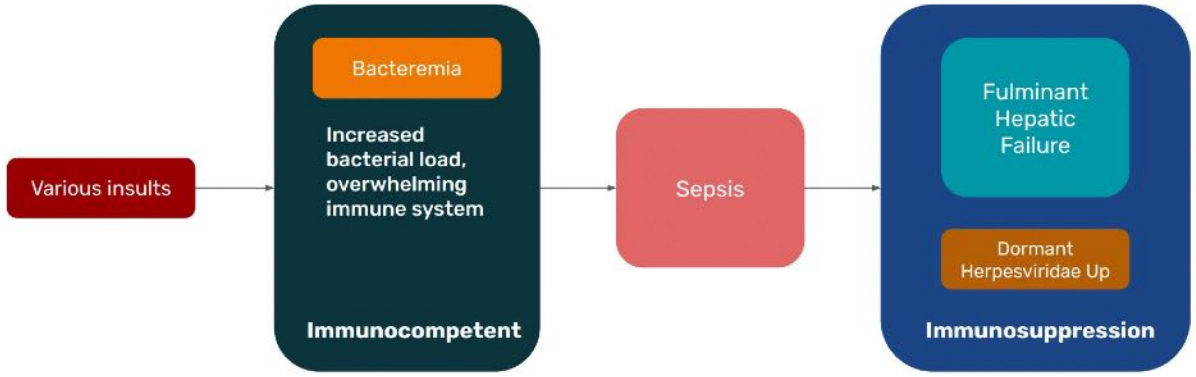
Plausible Hypotheses

Hypothesis 1: Herpesvirus infection caused immunosuppression and sepsis-induced hepatic failure



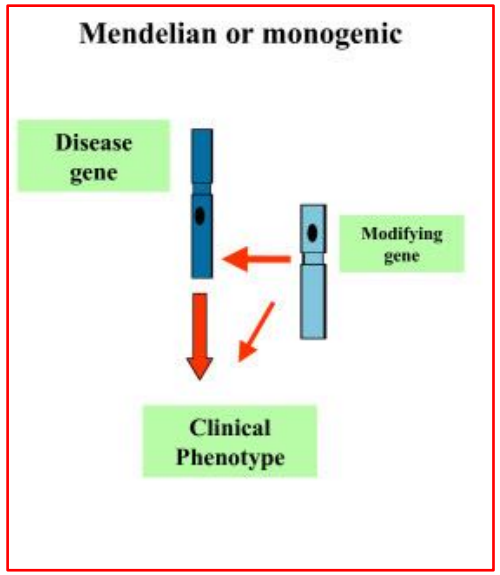
Plausible Hypotheses

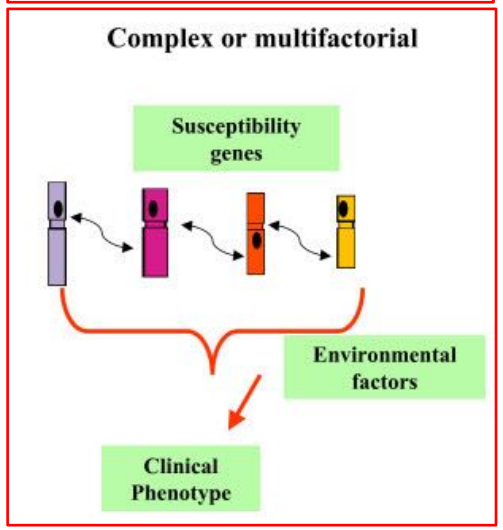
Hypothesis 2: Increased Herpesviridae level were caused by sepsis-induced immunosuppression

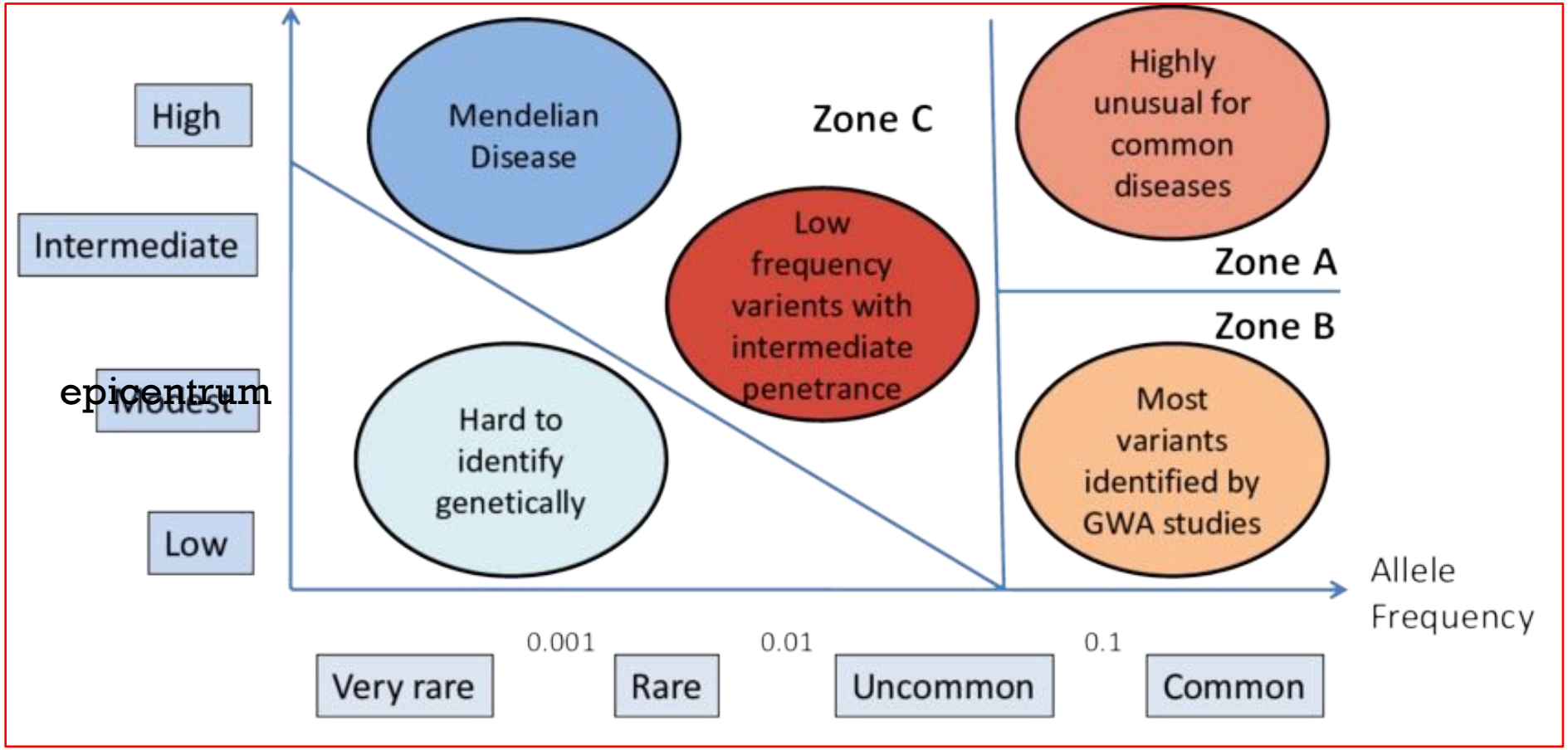


Sepsis-induced immunosuppression causing pathogens to still persist in the blood and causing hepatic failure (Hotchkiss, Monneret and Payen, 2013).

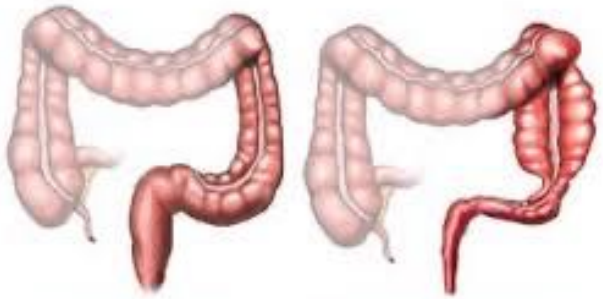
Mendelian *vs.* Complex Genetic Disorder







Hirschsprung Disease



Functional obstruction

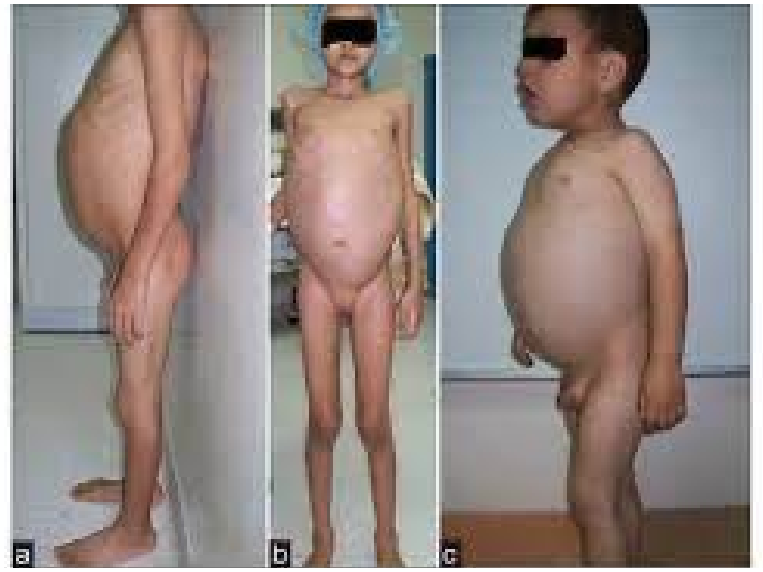
Indonesia 1:3,250

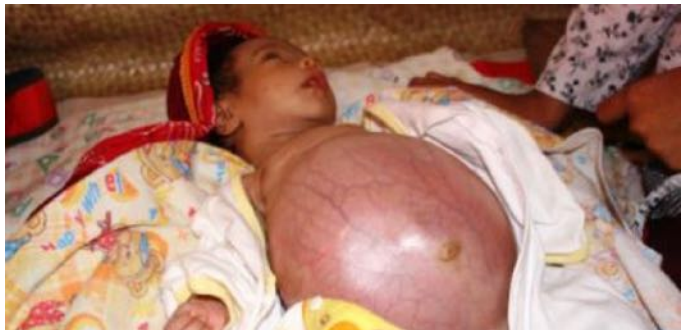
~ 24 Genes miRNA

Hirschsprung disease (HSCR)

European 1:5000
Asian 1:3500

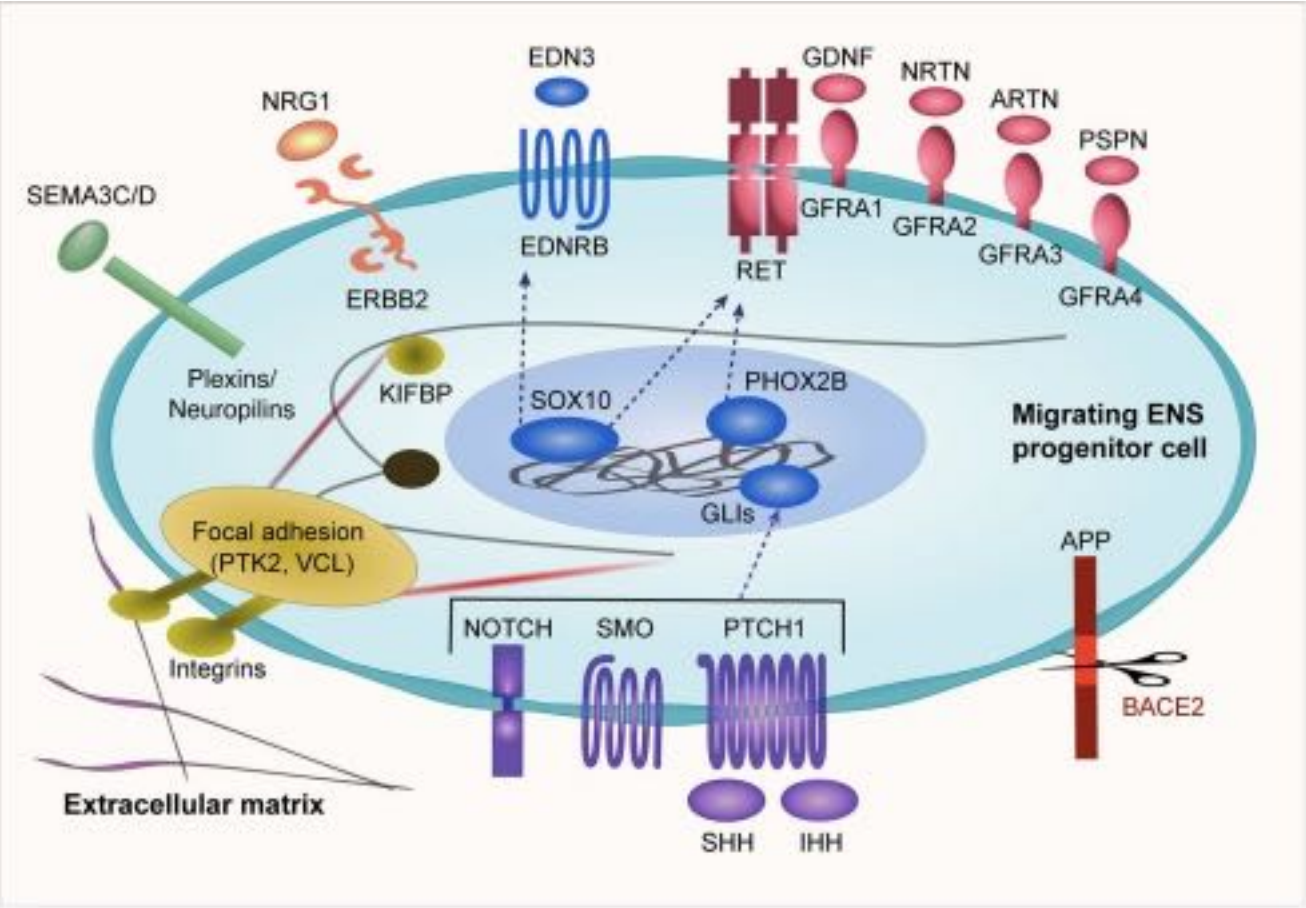
4:1 male





Tilghman et al., 2019; Gunadi et al., 2019a; 2019b; 2018a, 2018b; Parahita et al., 2018; Setiadi et al., 2017; Alves et al., 2013

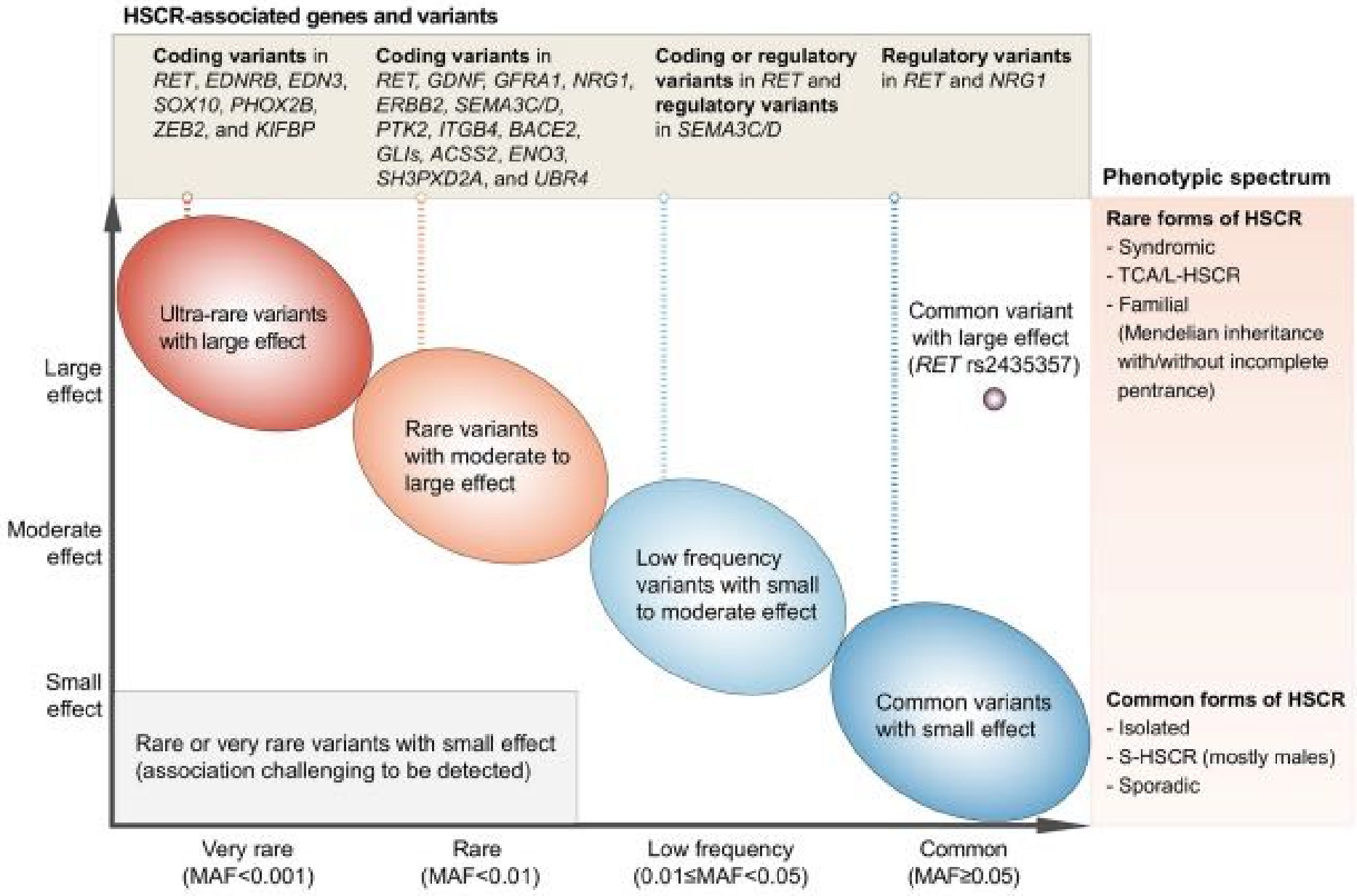
ROLE OF NGS IN GENETICS OF HSCR



~70% → **Next-Generation Sequencing**

Karim et al., 2021; Alves et al., 2013; Amiel et al., 2008; Gunadi et al. 2019; 2018; 2016a; Ped Surg Int 2016b; J Ped Surg 2014

HOW DO VARIANTS CAUSE THE HSCR?



(Karim et al., 2021)

JOURNEY OF *NGS WES* HSCR

14 May 2022
Submitted

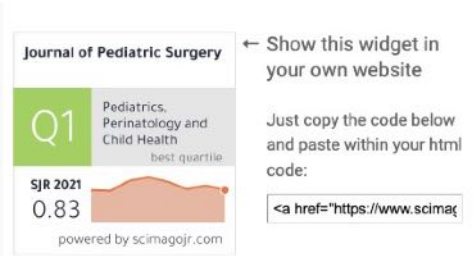
10 Sep 2022
1st Revision

26 Oct 2022
2nd Revision

12 Nov 2022
3rd Revision

16 Nov 2022

Manuscript Accepted!



ScienceDirect

Journal of Pediatric Surgery
Supports *open access*

4.1 CiteScore | 2.549 Impact Factor

H-INDEX

131

WHOLE-EXOME SEQUENCING *for* HSCR

39 sporadic non-syndromic HSCR patients and
16 non-HSCR control

Whole-exome sequencing

Bioinformatic Analysis

Population Allele Frequency (MAF <1%)

In silico Prediction tools
(PhyloP, CADD score, Grantham score)

FIRST VERSIONS

VARIANTS OF HSCR PATIENTS WITH *IN-SILICO* PREDICTION TESTS RESULTS

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Grantham score	Sample ID	Aganglionosis type	Associated disease
SORL1	c.1916G>T; p.Arg639Leu	rs200869993	Chr11: 121416003-121416003	0.002294	0.001925	5.336	21.4	102	HSCR12 HSCR21	Short Long	Alzheimer's disease, Parkinson's disease, dementia
ASTN1	c.3824C>T; p.Thr1275Met	rs201817286	Chr1: 176833481-176833481	0.002294	0.003277	6.938	20.2	81	HSCR16 HSCR19 HSCR24	Short Short Short	Hepatocellular carcinoma
APC	c.310T>A; p.Ser104Thr	N/A	Chr5: 112102975-112102975	N/A	N/A	2.939	15.8	58	HSCR11	Long	Adenomatous polyposis coli, colon cancer
APC	c.3445G>A; p.Glu1149Lys	rs371117193	Chr5: 112174736-112174736	0	0.0001089	3.557	18.14	56	HSCR21	Long	N/A
MYOF	c.1027G>A; p.Val343Met	N/A	Chr10: 95161265-95161265	N/A	N/A	4.021	20.6	21	HSCR22	Short	Muscular dystrophy and cardiomyopathy
MDN1	c.4118C>T; p.Ala1373Val	rs114919514	Chr6: 90455052-90455052	0.003992	0.003675	4.373	23.1	64	HSCR6	Short	Breast cancer
WWOX	c.211G>A; p.Gly71Arg	rs767880120	Chr16: 78143713-78143713	0.000998	0.0006229	4.821	27.3	125	HSCR14	Long	Hepatocellular carcinoma, breast cancer, lung cancer, encephalopathy, ataxia, seizures
HERC1	c.9392C>T; p.Thr3131Ile	rs558716578	Chr15: 63951967-63951967	0.000	0.000	7.689	27.3	89	HSCR07	Short	Intellectual disability

FIRST VERSIONS

KNOWN HIRSCHSPRUNG-RELATED VARIANTS

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Grantham score	Sample ID	Aganglionosis type
RET	c.1597G>A; p.Gly533Ser	rs75873440	Chr10: 43607621-43607621	0.000998	0.00096	9.097	34	56	HSCR97	Short
RET	c.2914A>T; p.Arg972Trp	rs76534745	Chr10: 43619231-43619231	N/A	N/A	2.363	22.7	101	HSCR95	Short

THIS IS
Second Version

KNOWN HIRSCHSPRUNG-RELATED VARIANTS

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Grantham score	Sample ID	Aganglionosis type
RET	c.1597G>A; p.Gly533Ser	rs75873440	Chr10: 43607621-43607621	0.000998	0.00096	9.097	34	56	HSCR97	Short
~~***RET***~~	~~c.2914A>T; p.Arg972Trp~~	~~rs76534745~~	~~Chr10: 43619231-43619231~~	~~N/A~~	~~N/A~~	~~2.363~~	~~22.7~~	~~101~~	~~HSCR95~~	~~Short~~

FIRST VERSIONS

NOVEL VARIANTS IN KNOWN HIRSCHSPRUNG-RELATED GENES

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Grantham score	Sample ID	Aganglionosis type
UBR4	c.5497G>A; p.Gly1833Arg	rs770317755	Chr1: 19486685-19486685	N/A	0.000	7.484	29.4	125	HSCR13	Short
GDNF	c.349G>A; p.Gly117Ser	N/A	Chr5: 37816040-37816040	N/A	N/A	7.456	25.7	56	HSCR13	Short
CELSR3	c.6695G>A; p.Arg2232His	rs148825770	Chr3: 48686234-48686234	0.000998	0.00096	5.638	18.44	29	HSCR14 HSCR20 HSCR5	Long Short Long
EDNRB	c.1146C>A; p.Asn392Lys	N/A	Chr13: 7874042-78478042	N/A	N/A	1.554	18.29	94	HSCR90	Short
BDNF	c.650G>C; p.Arg217Pro	rs757598331	Chr11: 27679486-27679486	N/A	0.000	6.091	15.75	103	HSCR90	Short
RET	c.1907C>G; p.Thr636Arg	rs121913310	Chr10: 43609955-43609955	N/A	N/A	3.187	20.9	71	HSCR95	Short
RET	c.2656C>G; p.Arg886Gly	rs146838520	Chr10: 43615577-43615577	N/A	0.000	2.536	18.47	125	HSCR93	Short
ZEB2	c.98A>C; p.Asp33Ala	N/A	Chr2: 145187569-145187569	N/A	N/A	7.586	16.94	126	HSCR108	Long
ECE1	c.2278C>G; p.Pro760Ala	rs186453862	Chr1: 21546474-21546474	0.00499	0.002498	5.696	19.45	27	HSCR12	Short
NRTN	c.167A>T; p.Gln56Leu	rs757355174	Chr19: 5824343-5824343	N/A	0.000	4.527	19.66	113	HSCR89	Short
NRTN	c.23C>A; p.Ala8Asp	rs768503585	Chr19: 5824199-5824199	N/A	0.000	5.066	23.3	126	HSCR06	Short

THIS IS
Second Version

NOVEL VARIANTS IN KNOWN HIRSCHSPRUNG-RELATED GENES

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Grantham score	Sample ID	Aganglionosis type
UBR4	c.5497G>A; p.Gly1833Arg	rs770317755	Chr1: 19486685-19486685	N/A	0.000	7.484	29.4	125	HSCR13	Short
GDNF	c.349G>A; p.Gly117Ser	N/A	Chr5: 37816040-37816040	N/A	N/A	7.456	25.7	56	HSCR13	Short
CELSR3	c.6695G>A; p.Arg2232His	rs148825770	Chr3: 48686234-48686234	0.000998	0.00096	5.638	18.44	29	HSCR14 HSCR20 HSCR5	Long Short Long
~~*EDNRB*~~	~~c.1146C>A; p.Asn392Lys~~	~~N/A~~	~~Chr13: 7874042-78478042~~	~~N/A~~	~~N/A~~	~~1.554~~	~~18.29~~	~~94~~	~~HSCR90~~	~~Short~~
~~*BDNF*~~	~~c.650G>C; p.Arg217Pro~~	~~rs757598331~~	~~Chr11: 27679486-27679486~~	~~N/A~~	~~0.000~~	~~6.091~~	~~15.75~~	~~103~~	~~HSCR90~~	~~Short~~
~~*RET*~~	~~c.1907C>G; p.Thr636Arg~~	~~rs121913310~~	~~Chr10: 43609955-43609955~~	~~N/A~~	~~N/A~~	~~3.187~~	~~20.9~~	~~71~~	~~HSCR95~~	~~Short~~ -
~~*RET*~~	~~c.2656C>G; p.Arg886Gly~~	~~rs146838520~~	~~Chr10: 43615577-43615577~~	~~N/A~~	~~0.000~~	~~2.536~~	~~18.47~~	~~125~~	~~HSCR93~~	~~Short~~
~~*ZEB2*~~	~~c.98A>C; p.Asp33Ala~~	~~N/A~~	~~Chr2: 145187569-145187569~~	~~N/A~~	~~N/A~~	~~7.586~~	~~16.94~~	~~126~~	~~HSCR108~~	~~Long~~
ECE1	c.2278C>G; p.Pro760Ala	rs186453862	Chr1: 21546474-21546474	0.00499	0.002498	5.696	19.45	27	HSCR12	Short
~~*NRTN*~~	~~c.167A>T; p.Gln56Leu~~	~~rs757355174~~	~~Chr19: 5824343-5824343~~	~~N/A~~	~~0.000~~	~~4.527~~	~~19.66~~	~~113~~	~~HSCR89~~	~~Short~~
NRTN	c.23C>A; p.Ala8Asp	rs768503585	Chr19: 5824199-5824199	N/A	0.000	5.066	23.3	126	HSCR06	Short

FIRST VERSIONS

POSSIBLE COMPOUND HETEROZYGOUS VARIANTS

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Sample ID	Aganglionosis type
MUTYH	c.1354G>T; p.Glu452Ter	rs376790729	Chr1: 45796895-45796895	0.000	0.000	1.048	22.7	HSCR18	Short
	c.116C>T; p.Ala39Val	N/A	Chr1: 45799236-45799236	N/A	N/A	1.003	12.99		
BAZ1A	c.2677G>T; p.Glu893Ter	N/A	Chr14: 35245185-35245185	N/A	N/A	6.997	45	HSCR90	Short
	c.1027G>T; p.Glu343Ter	rs1133283	Chr14: 35269531-35269531	N/A	N/A	3.825	42		
DICER1	c.2386G>T; p.Glu796Ter	N/A	Chr14: 95574711-95574711	N/A	N/A	7.556	43	HSCR93	Short
	c.1232C>A; p.Ser411Ter	N/A	Chr14: 95589677-95590677	N/A	N/A	9.187	42		
HYDIN	c.9088G>T; p.Glu3030Ter	N/A	Chr16: 70929944-70929944	N/A	N/A	7.37	50	HSCR93	Short
	c.5611G>T; p.Glu1871Ter	N/A	Chr16: 71004431-71004431	N/A	N/A	7.484	58		
SACS	c.6493G>T; p.Glu2165Ter	N/A	Chr13: 23911522-23911522	N/A	N/A	7.487	55	HSCR94	Short
	c.520G>T; p.Glu174Ter	N/A	Chr13: 23922558-23932558	N/A	N/A	7.785	40		
CKAP5	c.2512G>T; p.Gly838Ter	N/A	Chr11: 46800071-46800071	N/A	N/A	4.071	41	HSCR94	Short
	c.2101G>T; p.Glu701Ter	N/A	Chr11: 46806091-46806091	N/A	N/A	5.948	40		
UBR1	c.1927G>T; p.Glu643Ter	N/A	Chr15: 43330066-43330066	N/A	N/A	2.889	39	HSCR89	Short
	c.1152G>A; p.Tyr384Ter	N/A	Chr15: 43350569-43350569	N/A	N/A	1.929	35		
BPTF	c.2598dup; p.Glu867fs	rs753044214	Chr17: 65899950-65899950	N/A	0.000	N/A	7.891	HSCR05	Long
	c.4138A>T; p.Lys1380Ter	N/A	Chr17: 65908138-65908138	N/A	N/A	2.09	37		

THIS IS
Second Version

POSSIBLE COMPOUND HETEROZYGOUS VARIANTS

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Sample ID	Aganglionosis type
MUTYH	c.1354G>T; p.Glu452Ter	rs376790729	Chr1: 45796895-45796895	0.000	0.000	1.048	22.7	HSCR18	Short
	c.116C>T; p.Ala39Val	N/A	Chr1: 45799236-45799236	N/A	N/A	1.003	12.99		
~~***BAZ1A***~~	~~c.2677G>T; p.Glu893Ter~~	~~N/A~~	~~Chr14: 35245185-35245185~~	~~N/A~~	~~N/A~~	~~6.997~~	~~45~~	~~HSCR90~~	~~Short~~
	~~c.1027G>T; p.Glu343Ter~~	~~rs1133283~~	~~Chr14: 35269531-35269531~~	~~N/A~~	~~N/A~~	~~3.825~~	~~42~~		
~~***DICER1***~~	~~c.2386G>T; p.Glu796Ter~~	~~N/A~~	~~Chr14: 95574711-95574711~~	~~N/A~~	~~N/A~~	~~7.556~~	~~43~~	~~HSCR93~~	~~Short~~
	~~c.1232C>A; p.Ser411Ter~~	~~N/A~~	~~Chr14: 95589677-95590677~~	~~N/A~~	~~N/A~~	~~9.187~~	~~42~~		
~~***HYDIN***~~	~~c.9088G>T; p.Glu3030Ter~~	~~N/A~~	~~Chr16: 70929944-70929944~~	~~N/A~~	~~N/A~~	~~7.37~~	~~50~~	~~HSCR93~~	~~Short~~
	~~c.5611G>T; p.Glu1871Ter~~	~~N/A~~	~~Chr16: 71004431-71004431~~	~~N/A~~	~~N/A~~	~~7.484~~	~~58~~		
~~***SACS***~~	~~c.6493G>T; p.Glu2165Ter~~	~~N/A~~	~~Chr13: 23911522-23911522~~	~~N/A~~	~~N/A~~	~~7.487~~	~~55~~	~~HSCR94~~	~~Short~~
	~~c.520G>T; p.Glu174Ter~~	~~N/A~~	~~Chr13: 23932558-23932558~~	~~N/A~~	~~N/A~~	~~7.785~~	~~40~~		
~~***CKAP5***~~	~~c.2512G>T; p.Gly838Ter~~	~~N/A~~	~~Chr11: 46800071-46800071~~	~~N/A~~	~~N/A~~	~~4.071~~	~~41~~	~~HSCR94~~	~~Short~~
	~~c.2101G>T; p.Glu701Ter~~	~~N/A~~	~~Chr11: 46806091-46806091~~	~~N/A~~	~~N/A~~	~~5.948~~	~~40~~		
~~***UBR1***~~	~~c.1927G>T; p.Glu643Ter~~	~~N/A~~	~~Chr15: 43330066-43330066~~	~~N/A~~	~~N/A~~	~~2.889~~	~~39~~	~~HSCR89~~	~~Short~~
	~~c.1152G>A; p.Tyr384Ter~~	~~N/A~~	~~Chr15: 43350569-43350569~~	~~N/A~~	~~N/A~~	~~1.929~~	~~35~~		
BPTF	c.2598dup; p.Glu867fs	rs753044214	Chr17: 65899950-65899950	N/A	0.000	N/A	7.891	HSCR05	Long
	c.4138A>T; p.Lys1380Ter	N/A	Chr17: 65908138-65908138	N/A	N/A	2.09	37		

POSSIBLE COMPOUND HETEROZYGOUS VARIANTS

Gene name	Variant	SNP id	Chromosomal position	MAF in 1000 Genomes	MAF in gnomAD	phyloP	CADD score	Sample ID	Aganglionosis type
MUTYH	c.1354G>T; p.Glu452Ter	rs376790729	Chr1: 45796895-45796895	0.000	0.000	1.048	22.7	HSCR18	Short
	c.116C>T; p.Ala39Val	N/A	Chr1: 45799236-45799236	N/A	N/A	1.003	12.99		
~~***BPTF***~~	~~c.2598dup; p.Glu867fs~~	~~rs753044214~~	~~Chr17: 65899950-65899950~~	~~N/A~~	~~0.000~~	~~N/A~~	~~7.891~~	~~HSCR05~~	~~Long~~
	~~c.4138A>T; p.Lys1380Ter~~	~~N/A~~	~~Chr17: 65908138-65908138~~	~~N/A~~	~~N/A~~	~~2.09~~	~~37~~		

Your Submission JPEDSURG-D-22-00481R3

External Inbox ×

Paul K. H. Tam
to me

Nov 16, 2022, 4:33 PM (7 days ago)

Ms. Ref. No.: JPEDSURG-D-22-00481R3
Journal of Pediatric Surgery

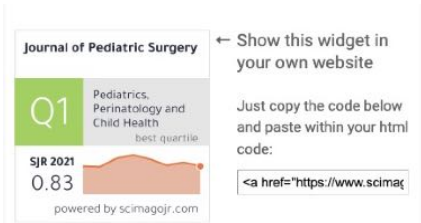
Dear Dr. . Gunadi,

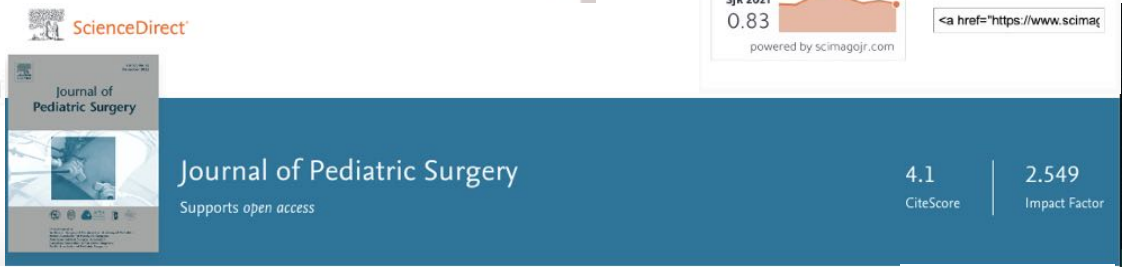
I am pleased to inform you that your manuscript JPEDSURG-D-22-00481R3 CR, titled, "Exome sequencing identifies novel genes and variants in patients with Hirschsprung disease", has been accepted for publication in the Journal of Pediatric Surgery. The manuscript will be published as an original paper. The editors wish to thank you and your co-authors for submitting this manuscript to the Journal.

If you have any questions regarding publication of this manuscript, please refer to your manuscript identification number when contacting us.

You will receive further information regarding galley proofs and the exact issue in as that determination is made.

Thank you for your support of the Journal of Pediatric Surgery.





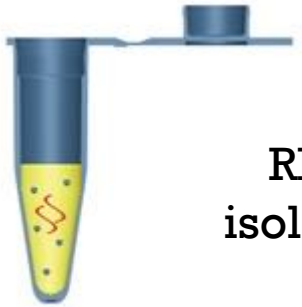
H-INDEX

131

RNASeq *for* HSCR

HSCR patients (Samples)

Non-HSCR patients (Controls)



RNA isolation

Next-Generation Sequencing (NGS)

- ✓Small RNA length ~35 base pairs, with >80% of reads
- ✓Total PF reads ranged from 229,972 to 1,354,232
- ✓Compared to controls, we found 11 unique miRNA families in HSCR

RNASeq *for* HSCR

Novel miRNA

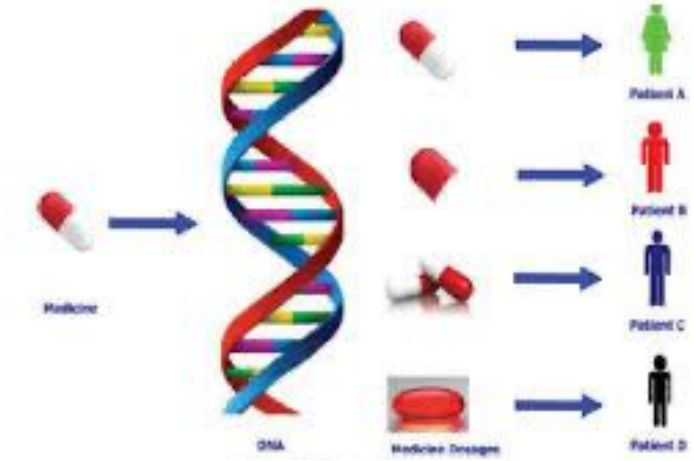
- ✓*miRNA-199*
- ✓*miRNA-221*
- ✓*miRNA-23*
- ✓*miRNA-26*
- ✓*miRNA-27*
- ✓*miRNA-3188*
- ✓*miRNA-4449*
- ✓*miRNA-8*

Established miRNA

- ✓*miRNA-143*
- ✓*miRNA-192*
- ✓*miRNA-770*

5

Precision Medicine



Personalized Medicine

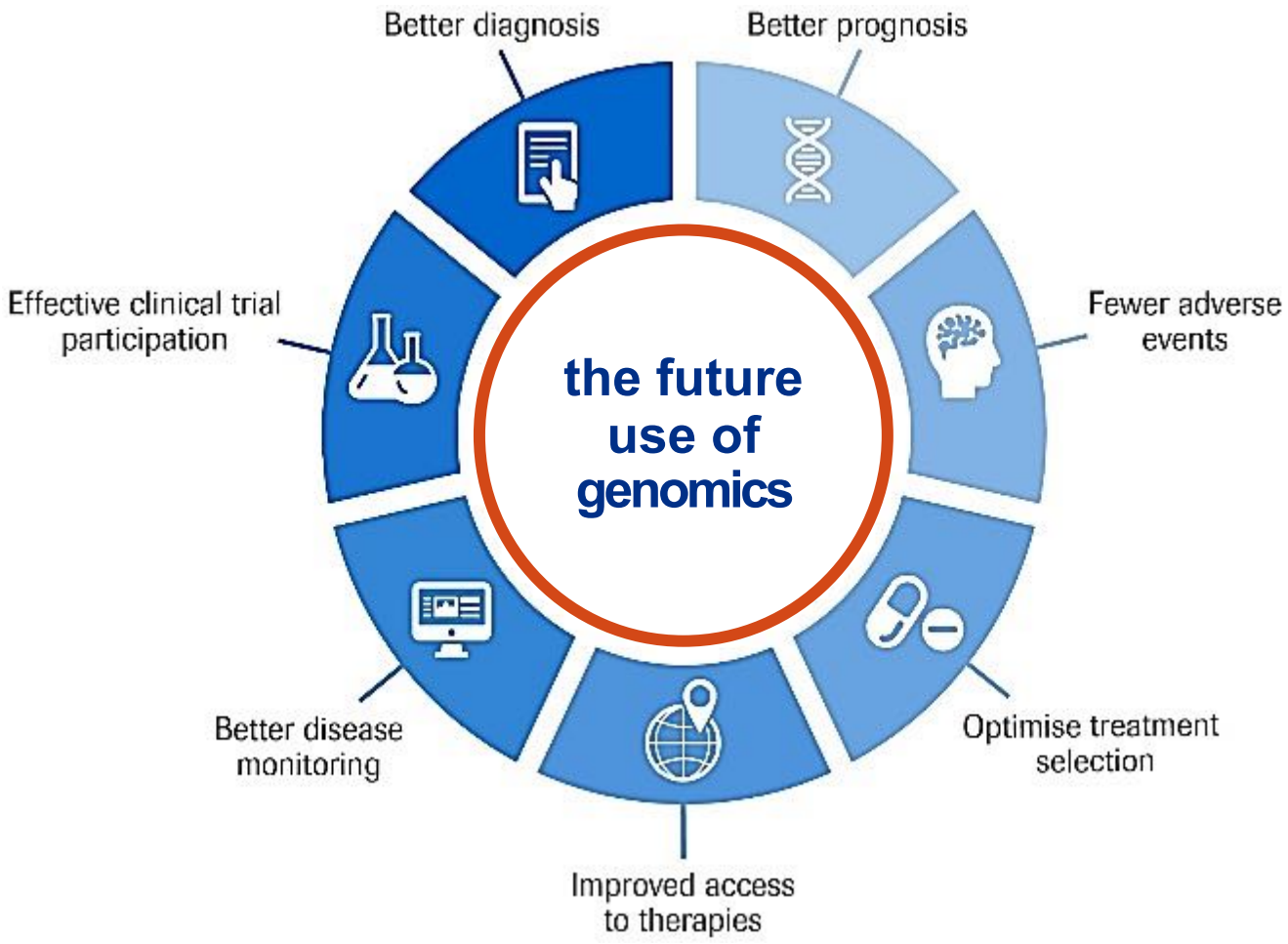
There is a lot of overlap between the terms "precision medicine" and "personalized medicine."

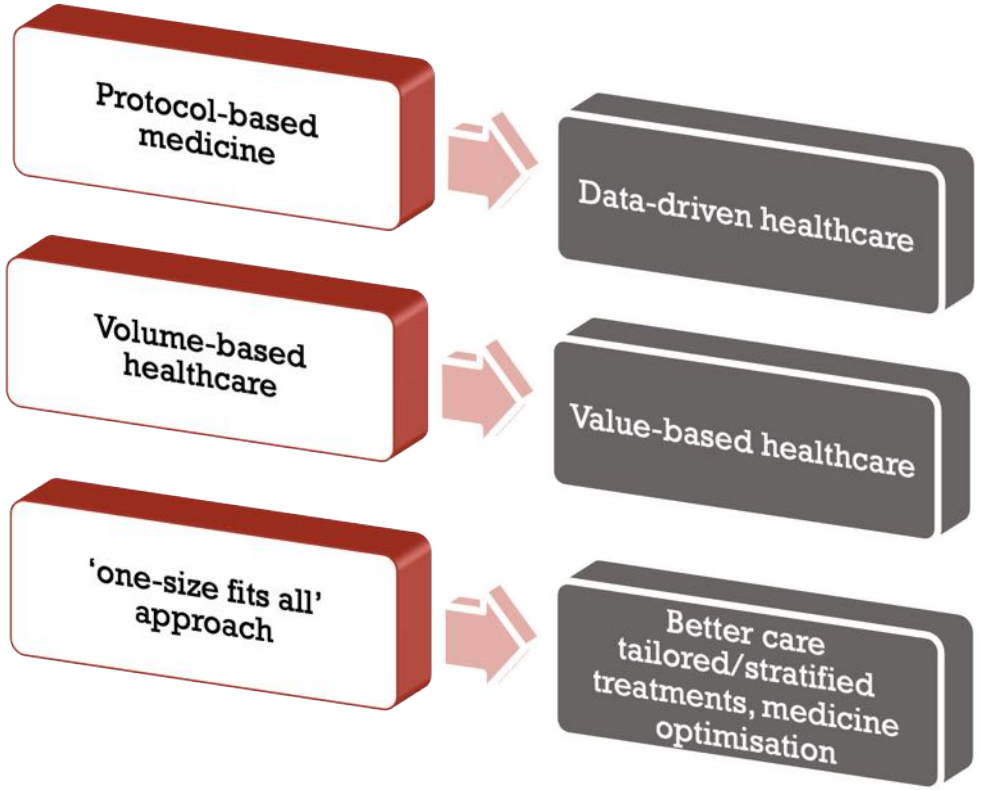
According to the National Research Council, "personalized medicine" is an older term with a meaning similar to "precision medicine."

However, there was concern that the word "personalized" could be misinterpreted to imply that treatments and preventions are being developed uniquely for each individual; in precision medicine, focus is on identifying which approaches will be effective for which patients based on genetic, environmental, and lifestyle factors.

The Council therefore preferred the term "precision medicine" to "personalized medicine." However, some people still use the two terms interchangeably.

Genomics & the Potential for Healthcare (Precision Medicine)



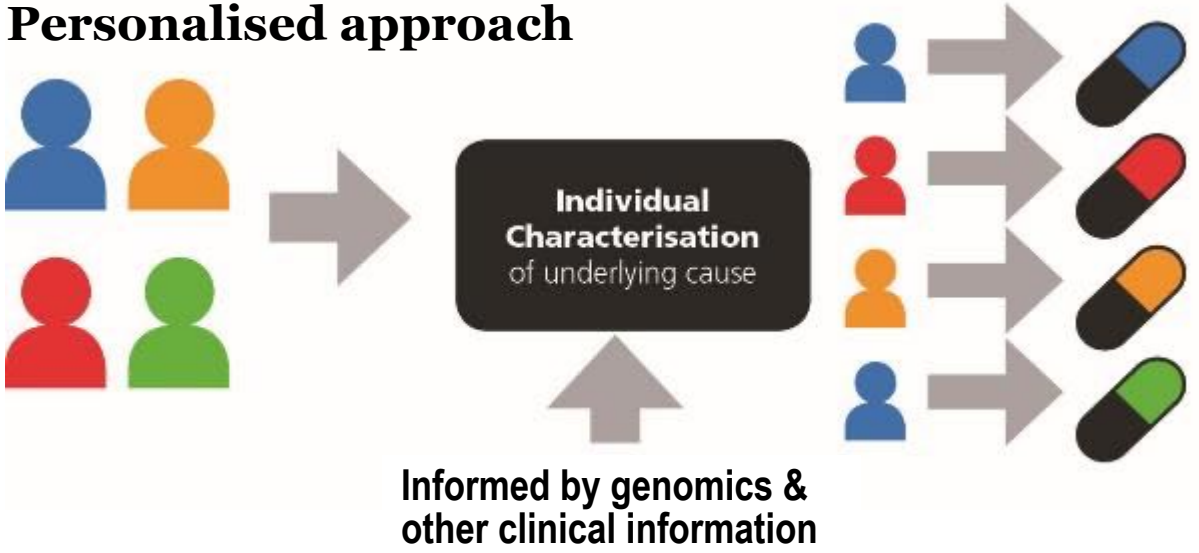


Using This Knowledge: Precision Medicine
The End Of The 'One Size Fits All' Era Of Medicine

Traditional approach

Everybody receives the same medicine – *typically only 30-60% effective*

Personalised approach



Tailored treatment to match an individual's makeup & response – *more effective and fewer side-effects*

Global Precision Medicine Landscape

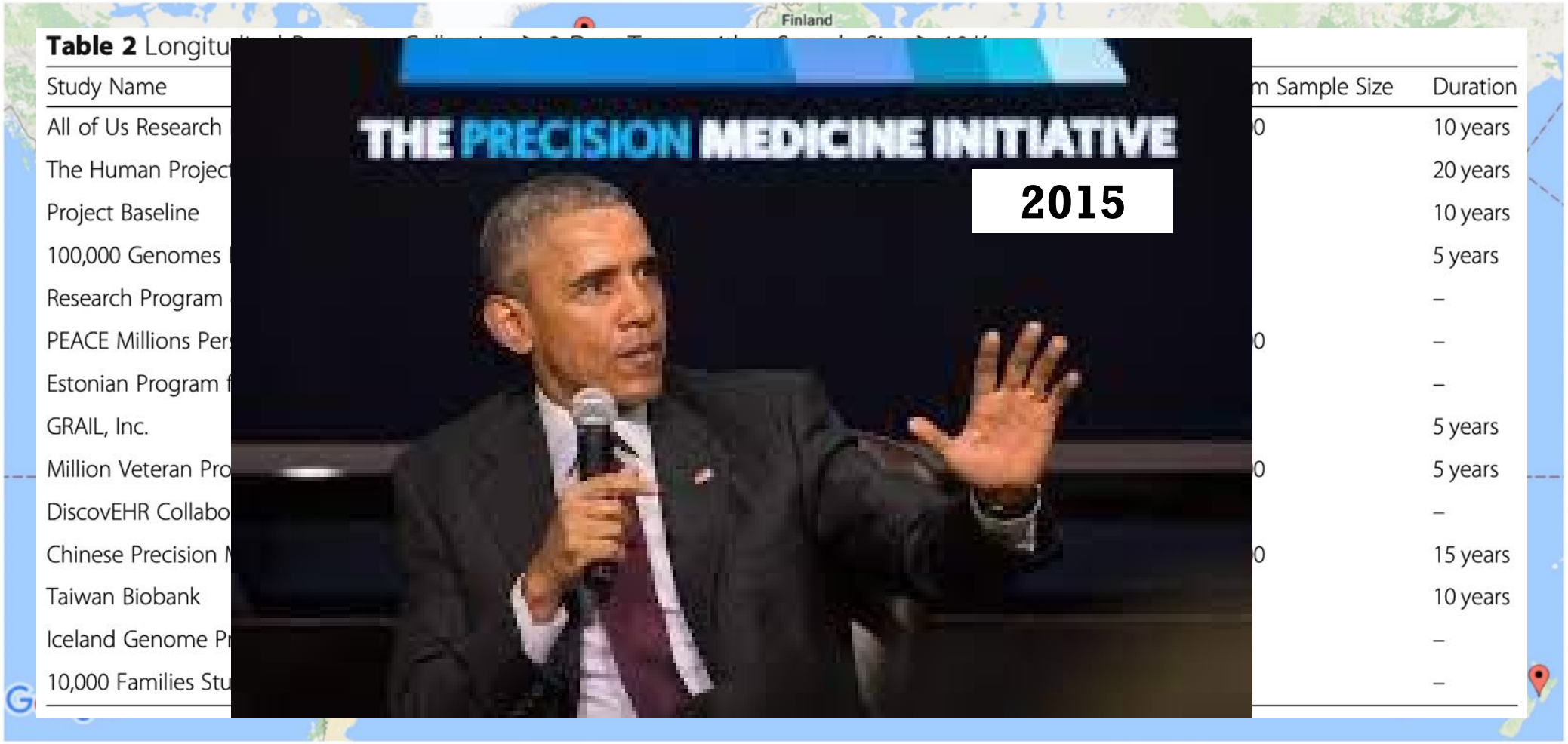


Table 2 Longitu

Study Name	m Sample Size	Duration
All of Us Research	0	10 years
The Human Project		20 years
Project Baseline		10 years
100,000 Genomes		5 years
Research Program		–
PEACE Millions Per	0	–
Estonian Program f		–
GRAIL, Inc.		5 years
Million Veteran Pro	0	5 years
DiscovEHR Collabo		–
Chinese Precision M	0	15 years
Taiwan Biobank		10 years
Iceland Genome Pr		–
10,000 Families Stu		–

Fig. 2 Global precision medicine efforts identified using search terms from Table I. The Google Map was generated through BatchGeo, an open source mapping tool

The World's Largest 10 Economics in 2030

Comparing 2017 vs. 2030

To create some additional context, we've compared these projections to the IMF's most recent data on GDP (PPP) for 2017. We've also added in potential % change for each country, if comparing these two data sets directly.

Here's how the numbers change:

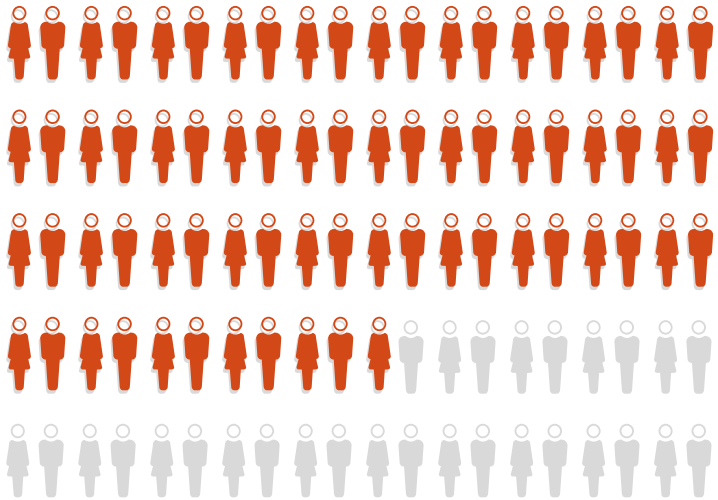
Rank	Country	Proj. GDP (2030, PPP)	GDP (2017, PPP)	% change
#1	China	$64.2 trillion	$23.2 trillion	+177%
#2	India	$46.3 trillion	$9.5 trillion	+387%
#3	United States	$31.0 trillion	$19.4 trillion	+60%
#4	Indonesia	$10.1 trillion	$3.2 trillion	+216%
#5	Turkey	$9.1 trillion	$2.2 trillion	+314%
#6	Brazil	$8.6 trillion	$3.2 trillion	+169%
#7	Egypt	$8.2 trillion	$1.2 trillion	+583%
#8	Russia	$7.9 trillion	$4.0 trillion	+98%
#9	Japan	$7.2 trillion	$5.4 trillion	+33%
#10	Germany	$6.9 trillion	$4.2 trillion	+64%

HTTPS://WWW.VISUALCAPITALIST.COM/WORLDS-LARGEST-10-ECONOMIES-2030/

Precision Medicine in Indonesia ?

Fact 1

2 million Indonesian go overseas for medical treatment every year



Fact 2

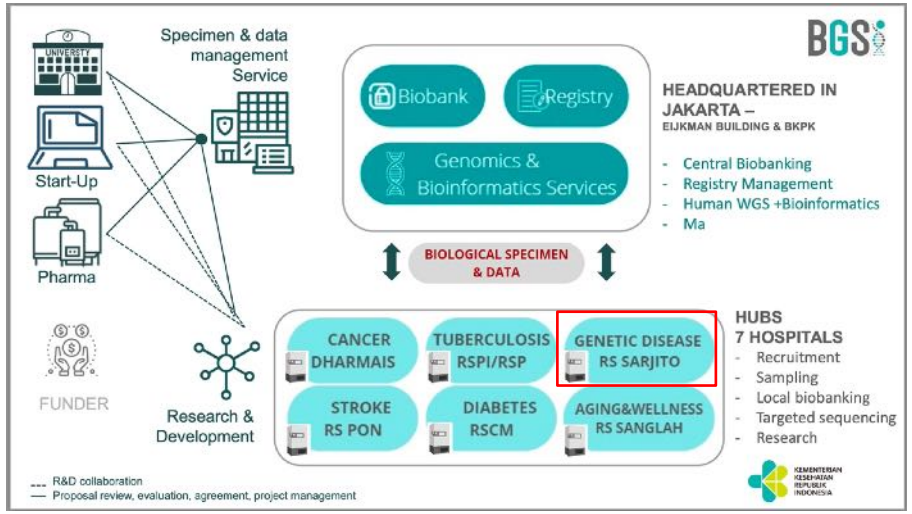
Human resources knowledge & skills varies among institutions

Collaboration is mandatory!

Indonesian Precision Medicine Initiative

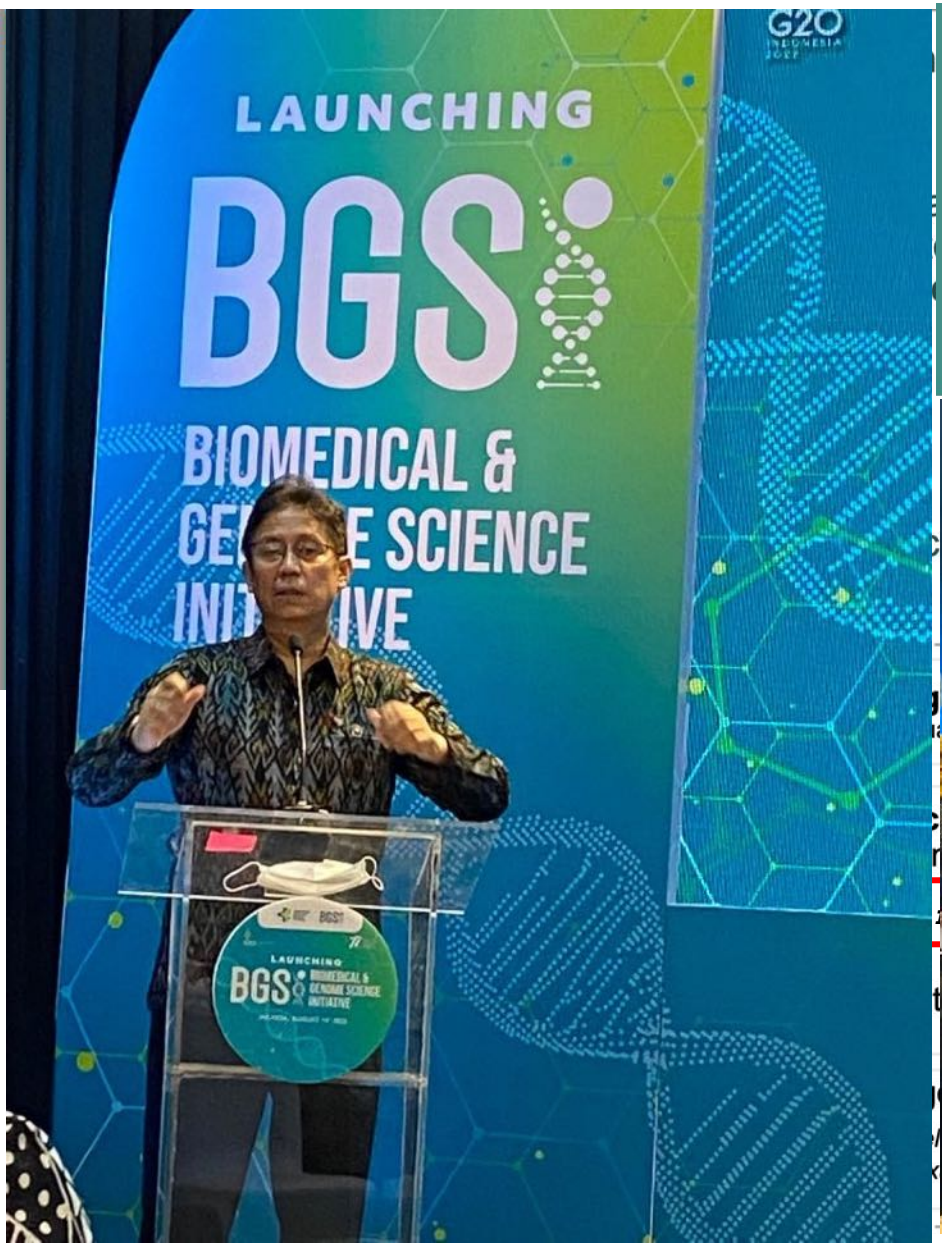
A selected Network to drive this initiative

- 6 hospitals will be first to drive this Precision Medicine targeted initiative
- On specific diseases
  - Diabetes (RSCM)
  - Cancer (Dharmais)
  - Tuberculosis (RSPI/RSP)
  - Stroke (RS PON)
  - Genetic Diseases (Sarjito)
  - Aging, beauty & wellness (RS Sanglah)



BGSi for Precision Medicine
14 August 2022

Indonesian Precision Medicine Initiative



Biomedical and Genomic Science Initiative

BGS i

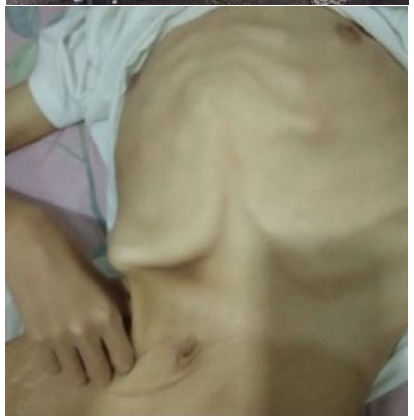
14 August 2022

Precision Medicine Approaches

Mendelian (e.g., DMD)	Complex Genetic Disorder (e.g., HSCR)
✓ Genetic testing and defining genetic etiology ✓ Genetic diagnosis guides treatment (gene therapy) ✓ Screening and counseling of family members ✓ Genomic screening in preventive health	✓ Disease risk stratification & polygenic risk score ✓ Network-based disease understanding & therapies ✓ Other novel therapies based upon key driver genes

Duchenne Muscular Dystrophy

- ✓ Incidence 1:3,500-5,000
- ✓ Indonesia: 5,000,000 live births/year → 500-700 DMD/year
- ✓ Progressive muscular weakness
- ✓ 10-12 yo wheelchair bound → Premature death within the 2nd decades

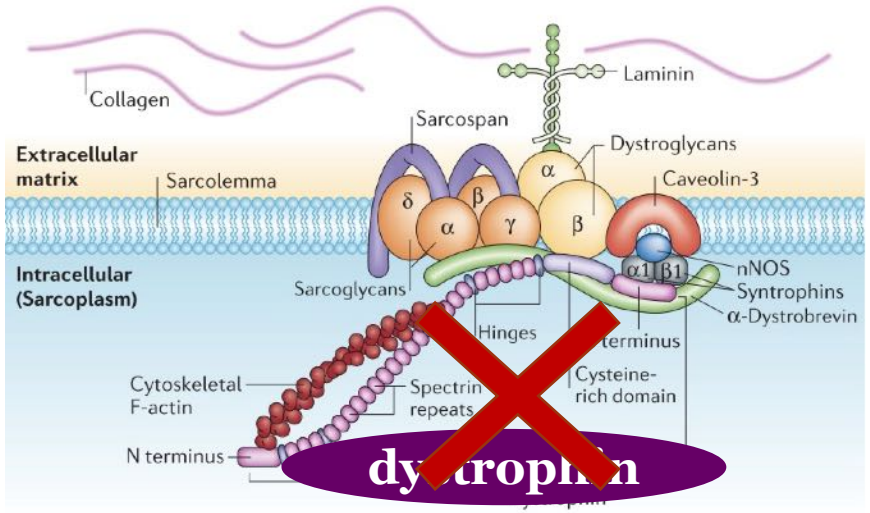


Dystrophin gene

VARIANTS

→ Dystrophin protein → Maintain skeletal muscle

- Deletions 65%
- Point mutation 26%
- Duplication 7%



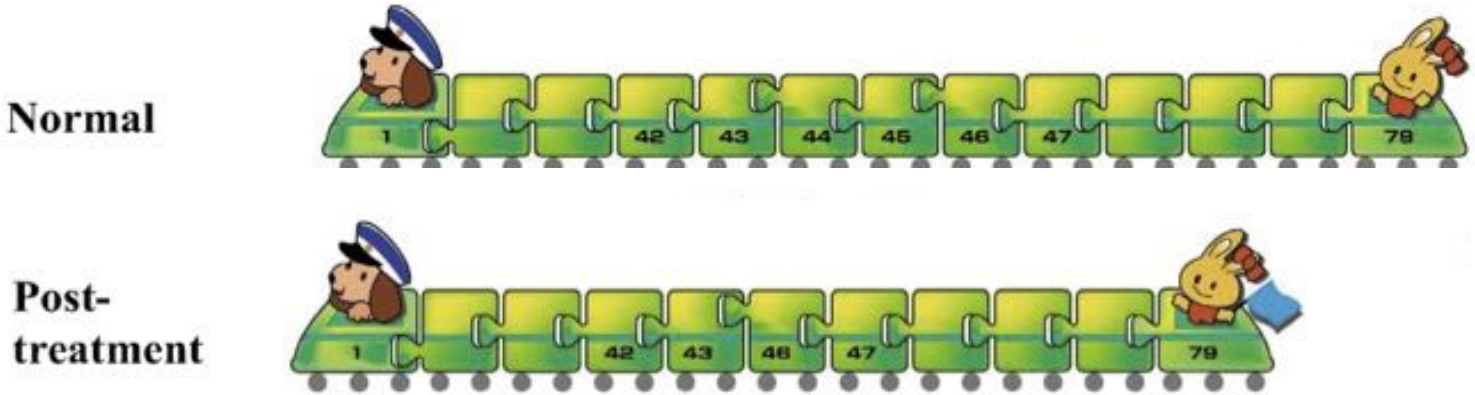
Duchenne Muscular Dystrophy (DMD)

Emerging Treatment for DMD

Gene therapy: **Exon skipping/ Read-through**

Lessen severity of DMD

Longer life expectancy and higher QoL



~80% can be treated with exon skipping strategy

Can this patient be treated with gene therapy?

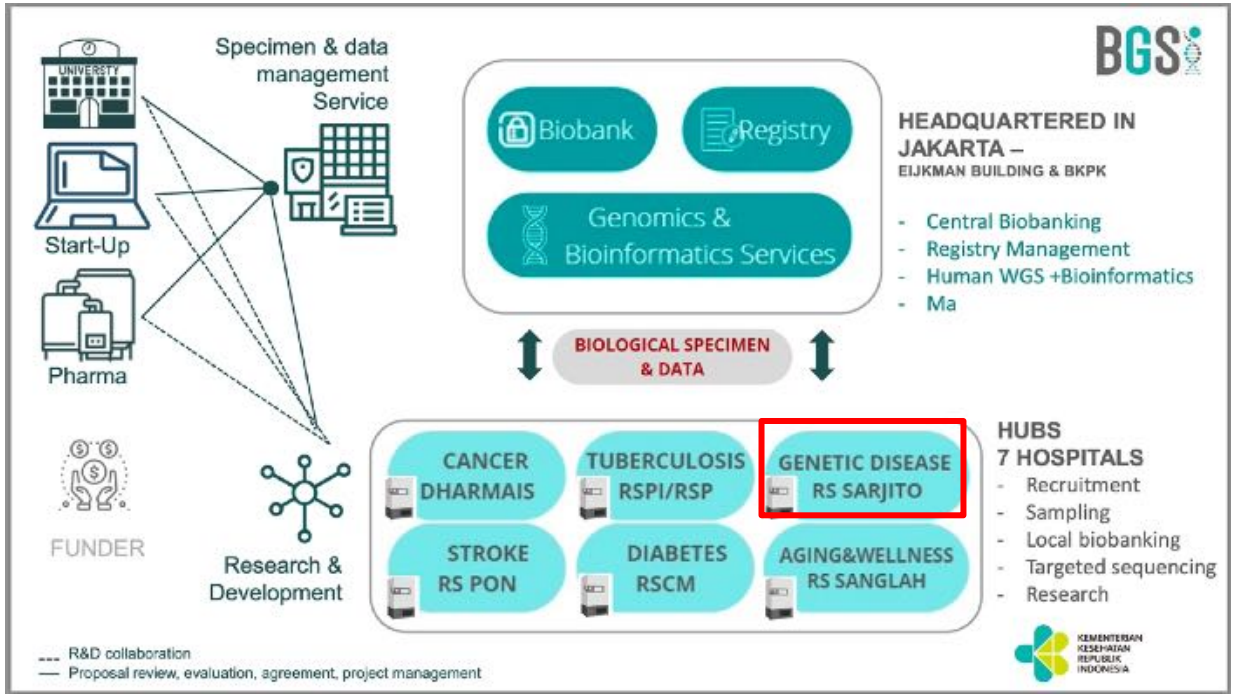
Molecular diagnostic tool to determine amenability

Echigoya Y, Yokota T. Nucleic Acid Ther. 2014, 24, 57–68.

Aartsma-Rus A, Straub V, Hemmings R et al (2017). *Nucleic Acid Ther* 27(5):251-259.

Matsuo M, Takeshima Y, Nishio H. Brain Dev. 2016 Jan;38(1):4-9

TARGETED SEQUENCING FOR DMD



- DMD → 79 exons
- EMD → 6 exons
- LMNA → 12 exons
- CAPN3 → 24 exons

Each amplicon consist a max. 1000 bp

✓DMD → 75 amplicons
✓EMD → 3 amplicons
✓LMNA → 7 amplicons
✓CAPN3 → 18 amplicons

Total: 103 amplicons

WHOLE-GENOME SEQUENCING *for* DMD

HUBS

HUBS :
1. Cancer
2. Tuberculosis
3. Stroke
4. Diabetes
5. Aging & Wellness
6. Genetic Diseases

- Study design
- Sample collection
- Recruitment
- Sample dispatch
- Data entry
- Targeted Sequencing
- Tertiary Analysis

REGISTRY

Biomedical Genome-Based Science Initiative

Resources

100K+ | 5 | 15+ | 10K+ | 30+ | 6+

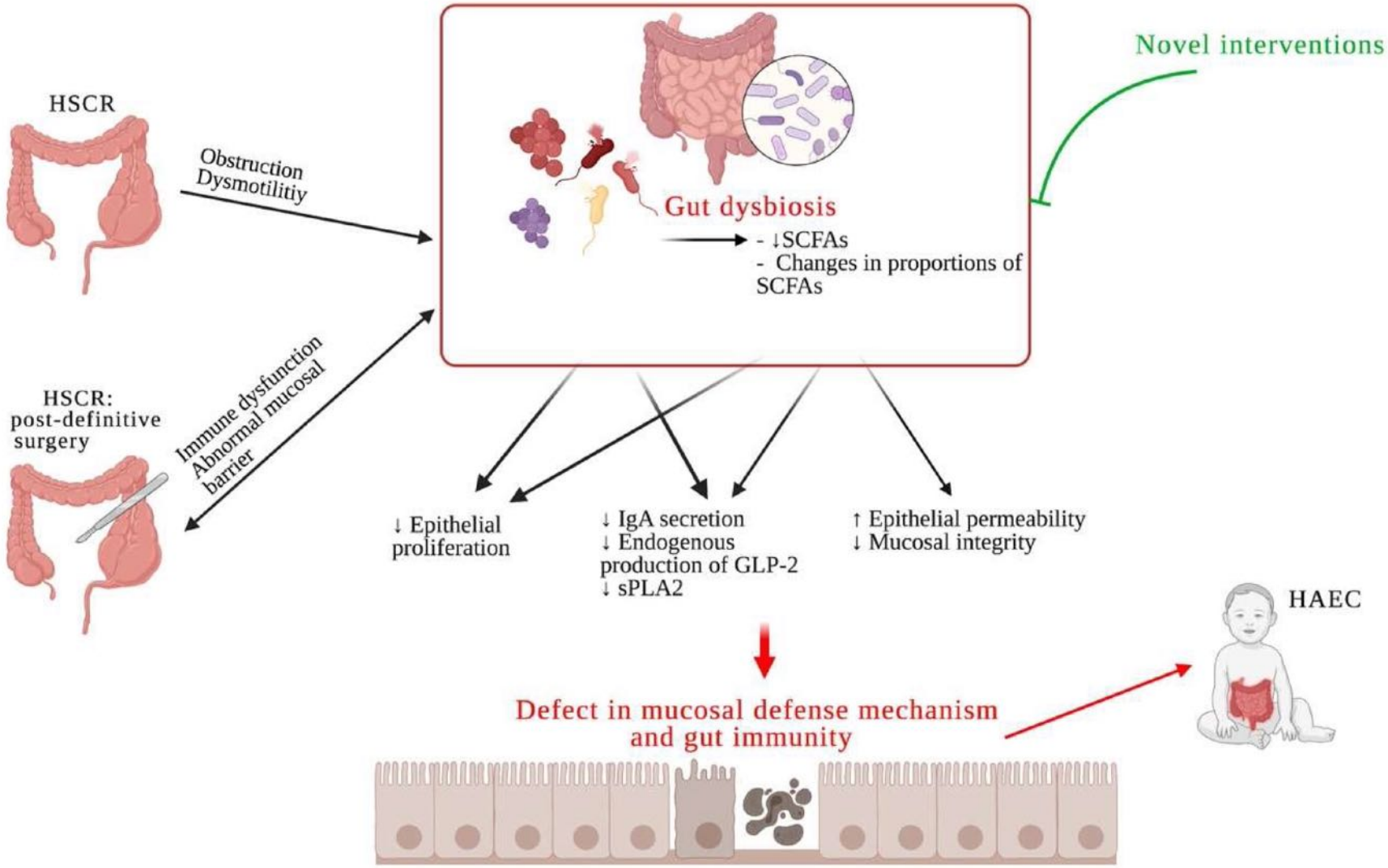
BioinformatiCS

Assembler, Genome browser, Blast, Phylogenetic, Gene annotation, Alignment, Drug-target prediction, etc.

BGSi

- Biobank
- WGS
- *INA Human Genome repository*

HIRSCHSPRUNG-ASSOCIATED ENTEROCOLITIS (HAEC)



Chantakhow S et al., 2021

PROPOSED KOCH'S POSTULATES FOR NGS

Original[A]	Molecular[B]	Next-generation sequencing[C]
• The microorganism is found in abundance in diseased but not in healthy individuals • The microorganism is able to be isolated from the diseased host and grown in pure culture • The cultured microorganism causes disease when inoculated into a healthy host • The microorganism can be re-isolated from the inoculated host and is the same as the original microorganism • Elimination of the microbe from the host alleviates disease	• The virulence gene is found in pathogenic but not nonpathogenic microbial strains • Deletion or inactivation of the virulence gene leads to loss of microbe pathogenicity • Reactivation or allelic replacement of the gene restores microbe pathogenicity	• Individual microorganisms or communities of microorganisms, as identified by sequencing, differ in abundance, organization, and/or function in diseased vs. healthy individuals • Community virulence or functional consequences may or may not depend on specific, well-defined microbe virulence genes • Community modification and/or elimination of specific community members alleviates the disease state

[A]The first four postulates are derived directly from Koch's original postulates, whereas the fifth is derived from Evans (138). Koch's original postulates do not account for viruses, parasites, unculturable bacteria, and/or the concept of host colonization with a potential pathogen. [B]Focuses on genes that make a microbe virulent (10, 139). [C]Focuses on use of nucleic acid sequences rather than culture or entire genes to find emerging pathogens (8, 9).

Wensel et al., 2022

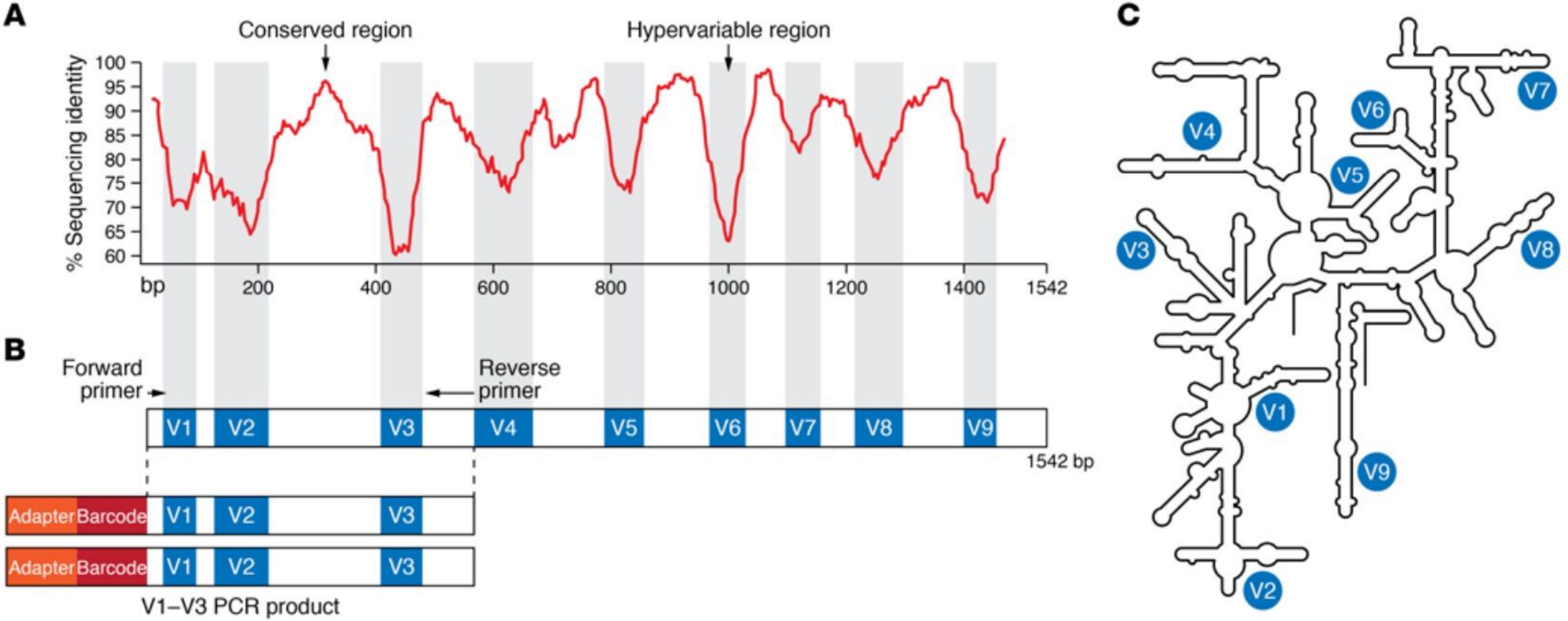
COMMON MICROBIOME SEQUENCING METHODS

	Method		
	Amplicon (16S, 18S, ITS[A])	**Shotgun metagenomics**	**RNA sequencing**
What is sequenced?	DNA coding for the 16S, 18S ribosomal subunit or ITS	Host and microbial DNA	Host and microbial RNA
What is the taxonomic resolution?	Phylum–genus, sometimes species	Species–strains	Species–strains
What is the taxonomic coverage?	Bacteria, archaea (16S); eukaryotes (18S)	Bacteria, archaea, eukaryotes, DNA viruses	Bacteria, archaea, eukaryotes, DNA and RNA viruses
Are appropriate reference databases available?	Over 3 million 16S gene sequences from humans and environmental sources are available	Over 100,000 genomes with a bias toward human microbiomes	Over 100,000 genomes with a bias toward human microbiomes
Does host contamination occur?	Limited	Yes, but can be mitigated by host DNA/rRNA depletion methods	Yes, but can be mitigated by host DNA/rRNA depletion methods
Can sequencing data yield a functional profile?	Not directly, but the functional profile can be predicted computationally	Yes, with appropriate computational expertise	Yes, with appropriate computational expertise
What is the minimum input for detection?	10 copies	1 ng[B]	1 ng[B]
What is the potential for false positives?	Lower due to extensive reference databases and error correction tools	Higher due to host DNA contamination of draft genomes	Higher due to host RNA/DNA contamination of draft genomes
What is the potential for bias?	Medium to high due to a dependence on primers, a targeted variable region, and PCR amplification	Lower due to the untargeted nature of the methodology	Lower due to the untargeted nature of the methodology
What level of computational skills is required?	Beginner–intermediate	Intermediate–advanced	Intermediate–advanced

[A]16S rRNA amplicons identify bacteria; 18S rRNA amplicons and internal transcribed spacer (ITS) sequences are most often used to identify fungi or parasites. [B]Although 1 ng is considered the minimum input, many sequencing facilities require at least 20 ng. Adapted from Zymo Research (140) and *Protein and Cell* (33).

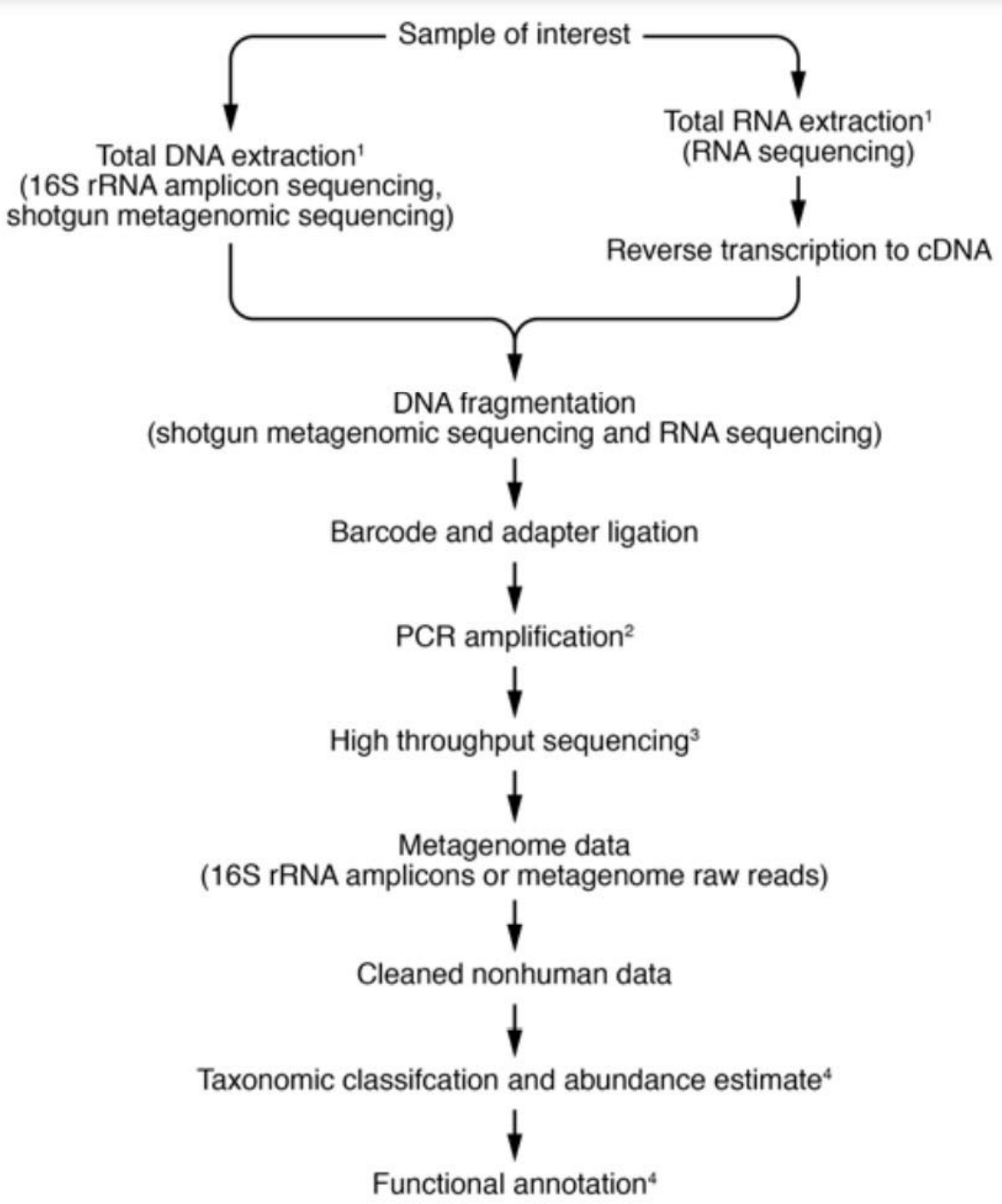
Wensel et al., 2022

BACTERIAL 16S rRNA GENE



(A) Percentage sequence identity of conserved and hypervariable regions of the bacterial 16S rRNA gene. (B) Illustration of conserved and hypervariable regions corresponding to A and PCR amplification of the V1–V3 region of the bacterial 16S rRNA gene. (C) Schematic of 16S rRNA gene structure with hypervariable regions (V1–V9) labeled. (Wensel et al., 2022)

Overview of key steps in 16s rRNA gene sequencing, shotgun metagenomic sequencing, & RNA sequencing processes



DEVELOPMENT OF THERAPEUTICS FROM NGS & MICROBIOME

A rich area of research, some important & ongoing work

1. Whole community transfer: fecal microbiota transplantation
2. Additions to the host microbiome: prebiotics and probiotic
3. Preclinical targeted NGS-linked tactics to modulate the microbiome
4. The microbiome as a source of new drugs

Wensel et al., 2022

- Advancement in Molecular Biology → NGS technology
- Most challenges in NGS research → Bioinformatics
- Accurate variant calling in NGS data is a critical step upon which virtually all downstream analysis & interpretation processes rely
- NGS has enabled a dramatic expansion of genomic research, including clinical genetic testing & Precision Medicine

Acknowledgement

Dr Kristy Iskandar
UGM

Prof Poh San Lai
NUS

Prof Christian Gilissen.
Radboud UMC

Dr Galuh Astuti
Radboud UMC

Prof Aravinda Chakravarti
New York Univ

NUSANTICS

PANDU BIOSAINS

UGM SARS-CoV-2
Genomic Surveillance

BBVET WATES

BBTKLPP YOGYAKARTA