Введение о платформах для секвенирования

Сравнение, недостатки/ограничения “железа”.
Сравнение панелей генов разных производителей,
ограничения “мокрых” технологий

Александр Лавров

Медико-генетический научный центр
им. Н.П. Бочкова

Три этапа высокопроизводительного секвенирования

1. Подготовка библиотеки
2. Амплификация библиотеки
3. Секвенирование библиотеки

Платформа = 3 + 2




	iSeq 100 System	MiniSeq System	MiSeq Series ⊕	NextSeq Series ⊕
Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)				●
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●
Exome Sequencing				●
Targeted Gene Sequencing (amplicon, gene panel)	●	●	●	●
Whole-Transcriptome Sequencing				●
Gene Expression Profiling with mRNA-Seq				●
Targeted Gene Expression Profiling	●	●	●	
Long-Range Amplicon Sequencing*	●	●	●	
miRNA & Small RNA Analysis	●	●	●	●
DNA-Protein Interaction Analysis			●	●
Methylation Sequencing				●
16S Metagenomic Sequencing		●	●	●
Run Time	9–17.5 hrs	4–24 hours	4–55 hours	12–30 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb
Maximum Reads Per Run	4 million	25 million	25 million †	400 million
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp

NextSeq Series ✪

HiSeq 4000 System

HiSeq X Series‡

NovaSeq 6000 System

Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)	●	●	●	●
Small Whole-Genome Sequencing (microbe, virus)	●	●		●
Exome Sequencing	●	●		●
Targeted Gene Sequencing (amplicon, gene panel)	●	●		●
Whole-Transcriptome Sequencing	●	●		●
Gene Expression Profiling with mRNA-Seq	●	●		●
Targeted Gene Expression Profiling				
Long-Range Amplicon Sequencing*				
miRNA & Small RNA Analysis	●	●		●
DNA-Protein Interaction Analysis	●	●		●
Methylation Sequencing	●	●		●
16S Metagenomic Sequencing	●	●		●
Run Time	12–30 hours	< 1–3.5 days	< 3 days	~13 - 38 hours (dual SP flow cells) ~13–25 hours (dual S1 flow cells) ~16–36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
Maximum Output	120 Gb	1500 Gb	1800 Gb	6000 Gb
Maximum Reads Per Run	400 million	5 billion	6 billion	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 x 250**

Ion GeneStudio S5 System
Turnaround time: 19 hr*

Ion GeneStudio S5 Plus System
Turnaround time: 10 hr*

Ion GeneStudio S5 Prime System
Turnaround time: 6.5 hr*

Sequencing throughput to support multiple applications of NGS

From gene panels to exomes, gene expression profiling to transcriptomes, or microbial genomes to microbiomes, Ion Torrent semiconductor sequencing optimizes cost and throughput for small and large NGS projects.

Cancer research:
gene panels for SNPs, indels, copy number, gene expression, and gene fusion analysis

Inherited disease research:
panels for targeted gene or whole exome analysis

Gene expression analysis:
whole transcriptome RNA-Seq, targeted RNA sequencing, and small RNA sequencing

Reproductive health research:
aneuploidy detection

Microbiology/infectious disease research:
microbial whole genomes, microbial typing, and metagenomics

Ion GeneStudio S5 System
Turnaround time: 19 hr*

Ion GeneStudio S5 Plus System
Turnaround time: 10 hr*

Ion GeneStudio S5 Prime System
Turnaround time: 6.5 hr*

Chip type	Number of reads	Read length (output*)	Ion GeneStudio™ S5 System	Ion GeneStudio™ S5 Plus System	Ion GeneStudio™ S5 Prime System
			Turnaround time (sequencing run** plus analysis time)		
Ion 510 Chip	2–3 million	200 bp (0.3–0.5 Gb)	4.5 hr	3 hr	3 hr
		400 bp (0.6–1 Gb)	10.5 hr	5 hr	5 hr
Ion 520 Chip	4–6 million	200 bp (0.6–1 Gb)	7.5 hr	3.5 hr	3 hr
		400 bp (1.2–2 Gb)	12 hr	5.5 hr	5.5 hr
	3–4 million	600 bp (0.5–1.5 Gb)	12 hr	5.5 hr	5.5 hr
Ion 530 Chip	15–20 million	200 bp (3–4 Gb)	10.5 hr	5 hr	4 hr
		400 bp (6–8 Gb)	21.5 hr	8 hr	6.5 hr
	9–12 million	600 bp (1.5–4.5 Gb)	21 hr	8 hr	7 hr
Ion 540 Chip	60–80 million	200 bp (10–15 Gb)	19 hr	10 hr	6.5 hr
		200 bp (20–30 Gb) 2 runs in 1 day	NA	20 hr	10 hr†
Ion 550 Chip	100–130 million	200 bp (20–25 Gb)	NA	11.5 hr	8.5 hr
		200 bp (40–50 Gb) 2 runs in 1 day	NA	NA	12 hr†



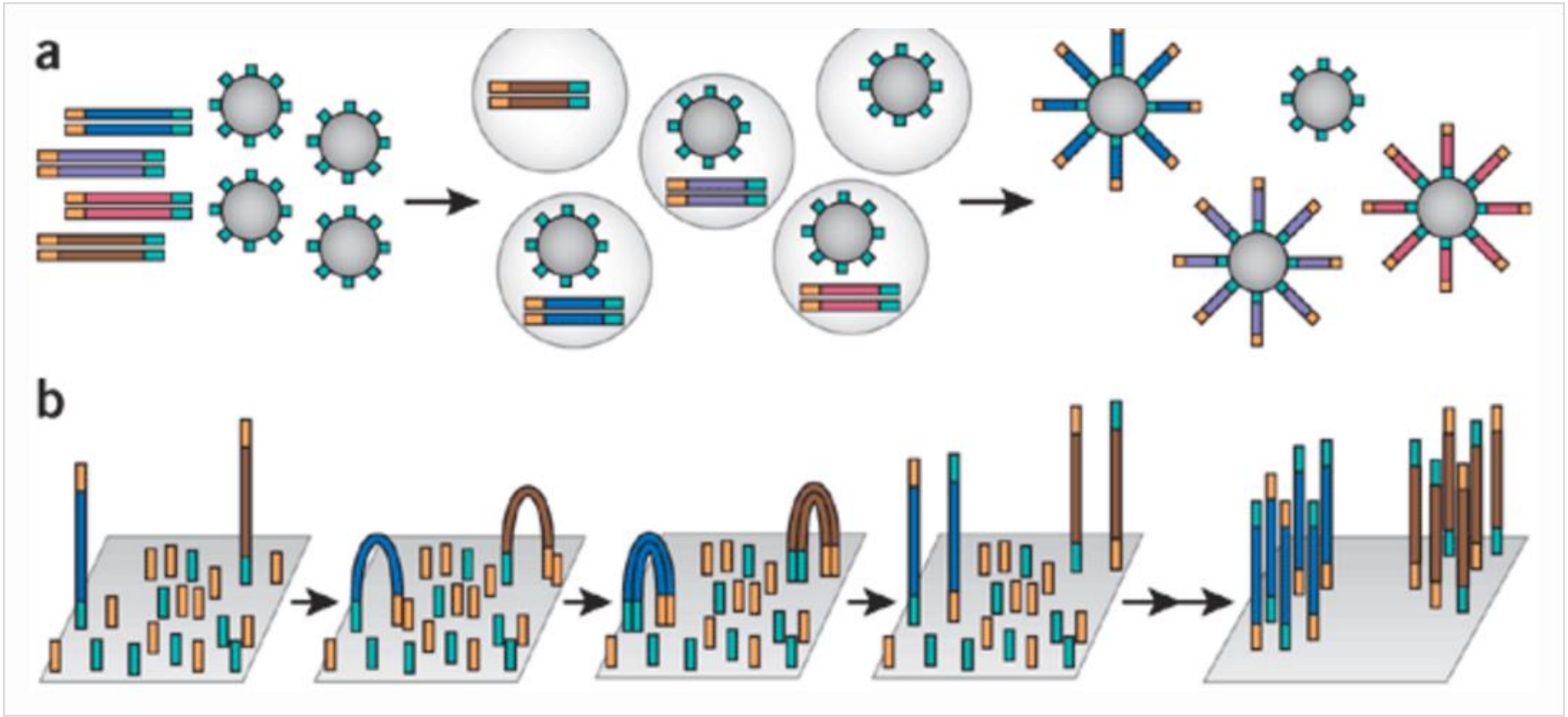


DNBSEQ-T7	DNBSEQ-G400 （previously MGISEQ-2000)	DNBSEQ-G50 （previously MGISEQ-200)	BGISEQ-500	BGISEQ-50
Ultra-high Throughput	Adaptive	Effective	Reliable	Fast
Whole Genome Sequencing,Deep Exome Sequencing,Transcriptome Sequencing,and Targeted Panel Projects.	WGS, WES, Transcriptome sequencing and more	Targeted DNA, RNA, Microbial sequencing	Targeted DNA, RNA, Epigenetics and clinical applications	Pathogen Rapid tests, NIPT, PGS and CNV tests
FC	FCL & FCS	FCS	FCL	FCS
——	4 lane & 2 lane	1 lane	2 lane	2 lane
Ultra-high Throughput	High Throughput	Medium Throughput	High Throughput	Low Throughput
6Tb	1440Gb	60Gb	520Gb	225Gb
5000M	1500-1800M	280-300M	1300M	375M
PE150 within 24 hours	~38hours	15-48hours	<9days	<15hours
SE50	SE50	SE50	SE50	SE50

Run Time	12–30 hours	
Maximum Output	120 Gb	60Gb
Maximum Reads Per Run	400 million	280-300M
Maximum Read Length	2 × 150 bp	15-48hours
		SE50

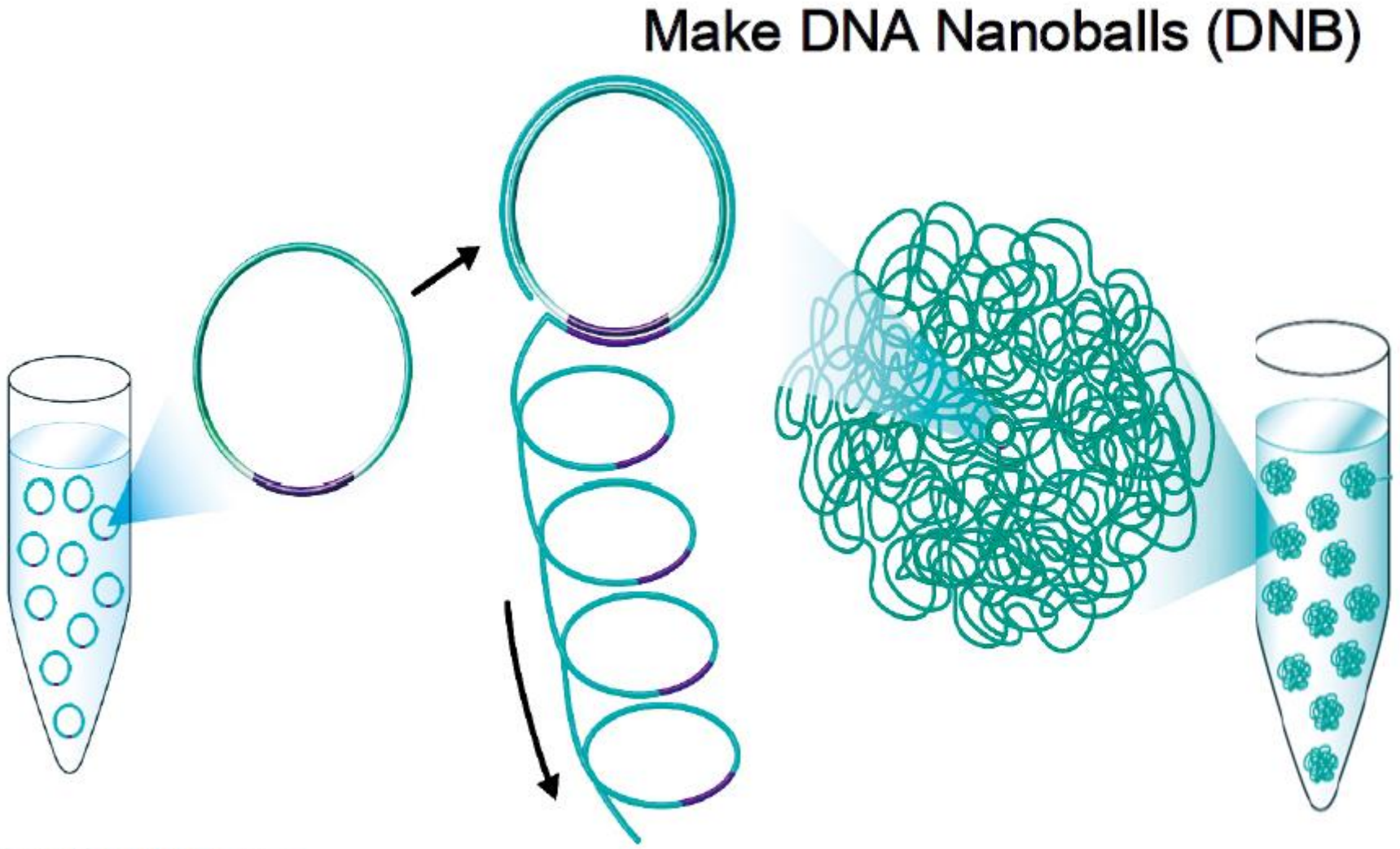
Ion 540 Chip	60–80 million	200 bp (10–15 Gb)	19 hr	10 hr	6.5 hr
		200 bp (20–30 Gb) 2 runs in 1 day	NA	20 hr	10 hr[1]

Клональная амплификация



Next-generation DNA sequencing
Jay Shendure & Hanlee Ji
Nature Biotechnology **26**, 1135 - 1145 (2008) Published online: 9 October 2008
doi:10.1038/nbt1486

Клональная амплификация



Oxford Nanopore Technologies:
> 200 Kb; до 42 Gb; 4.4 M

Oxford NANOPORE Technologies

PRODUCTS SERVICES APPLICATIONS GET STARTED PUBLICATIONS

Publications Posters Tools Data Releases Video White papers

Search Keywords

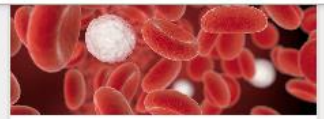
Search

Sequencing techniques

- Whole genome
- Targeted
- Epigenetics
- Metagenomics
- RNA/cDNA

Areas of research

- Environmental
- Informatics
- Human
- Clinical
- Cancer
- Basic Genome
- Microbiome



20TH AUGUST 2017

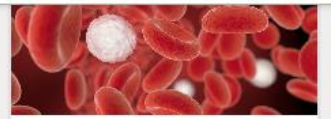
Accurate typing of class I human leukocyte antigen by Oxford nanopore ...

BioRxiv

31ST JULY 2017

Linear Assembly of a Human Y Centromere using Nanopore Long Reads

BioRxiv



27TH JULY 2017

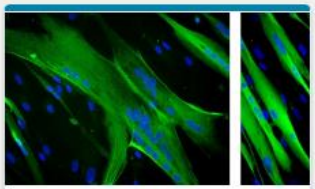
Multiplexed nanopore sequencing of HLA-B locus in Māori and Polynesian...

BioRxiv

26TH JULY 2017

The potential impact of nanopore sequencing on human genetics.

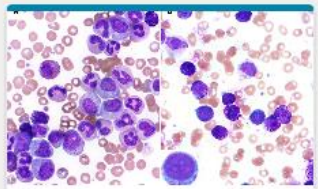
Human Molecular Genetics



28TH JUNE 2017

Nanopore-based single molecule sequencing of the D4Z4 array responsibl...

BioRxiv



27TH JUNE 2017

Mutational analysis in BCR-ABL1 positive leukemia by deep sequencing b...

Experimental and Molecular Pathology

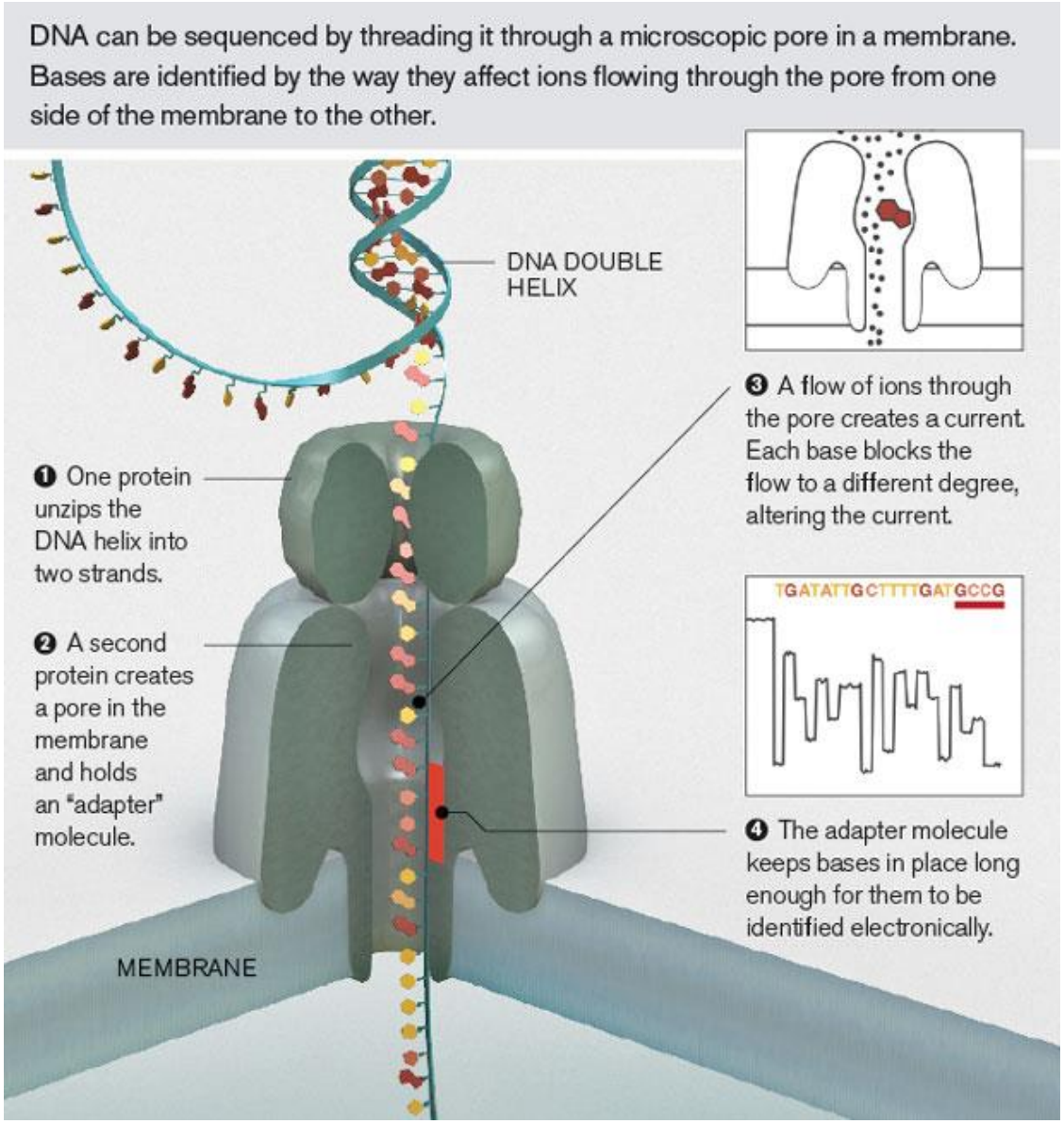
	Mk 1 MinION Commercially available	**GridION X5** **(5x MinIONs + Compute)**
Number of channels available for sequencing	Up to 512	up to 2,560
Your Sample		
Sample input requirement PCR free	10 pg - 1 µg	10 pg - 1 µg
Flow cell input volume	75 µl	75 µl
Sample preparation time 1D	10 minutes	10 minutes
System Operation		
Run time	1 minute - 48 hours	1 minute - 48 hours
Flow cell lifetime	~72 hrs	~72 hrs
Time to first usable read (in real time)	2 minutes	2 minutes
Number of reads at 10 kb at standard speed (250 bps)	Up to 4.4 M	Up to 4.4 M
Read length	The longest reported read by a MinION user is now approaching 1Mb	
1D Yield at 450 bps in 48 hours	Up to 40 Gb (Theoretical)	Up to 40 Gb (Theoretical)
1D Yield at 450 bps (Feb 2017 release)	10-20 Gb	50-100 Gb
Base calling accuracy	Up to 99%	Up to 99%
Raw data available	Yes	Yes
Modified base detection	Yes	Yes
Data analysis	Local offline/online	Local offline/online
Price on Application		
Reagent cost per run	$ 99	$99
Flow cell cost	$500 - $900	$299-900

SmidgION: nanopore sensing for use with mobile devices

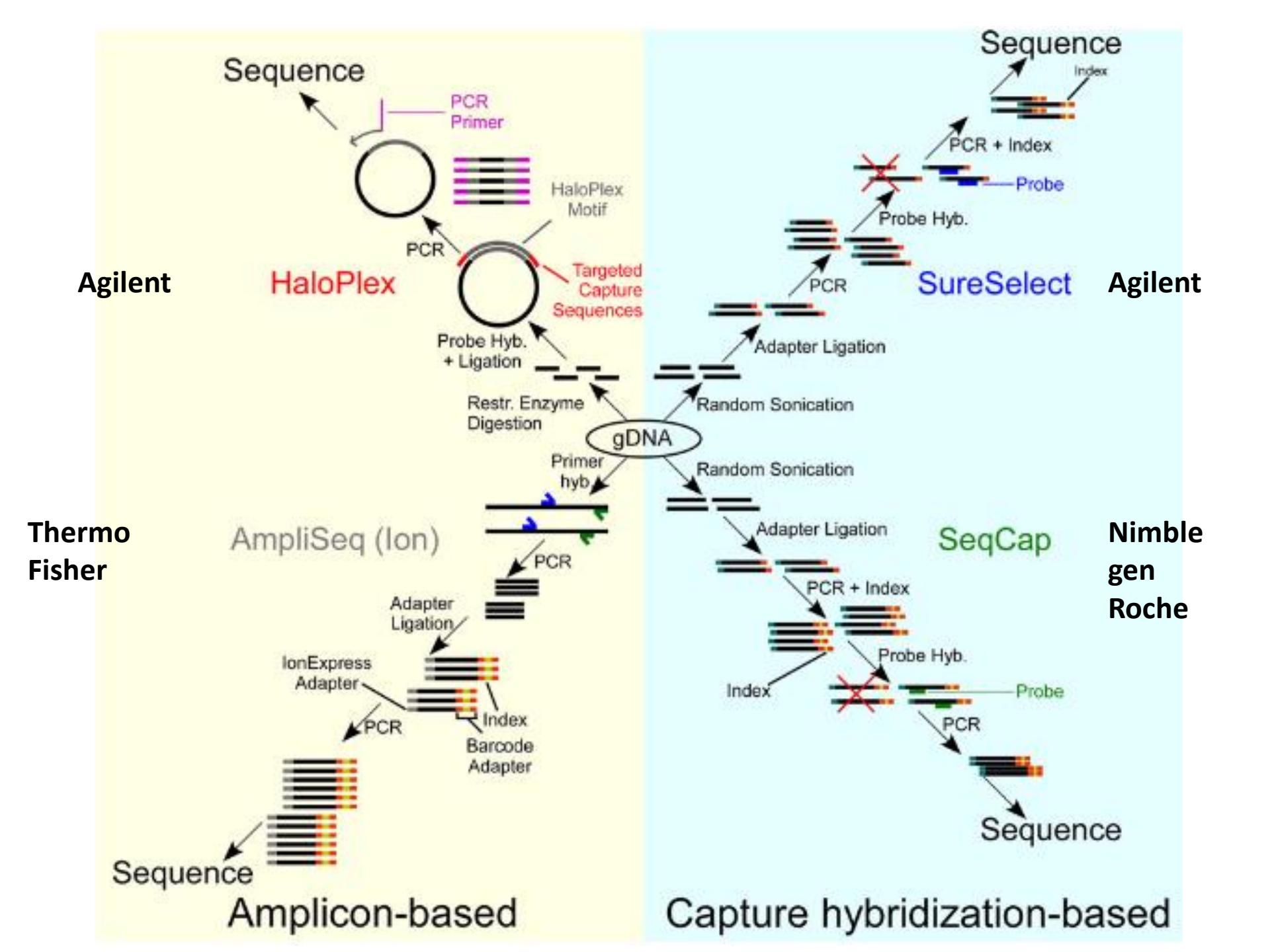
Using the same core technology as the handheld MinION device, we are now starting to develop an even smaller device.

Register your interest ›

Oxford Nanopore



Что выбрать?


Agilent
HaloPlex
Sequence
PCR Primer
HaloPlex Motif
PCR
Targeted Capture Sequences
Probe Hyb. + Ligation
Restr. Enzyme Digestion
gDNA
Primer hyb
AmpliSeq (Ion)
PCR
Adapter Ligation
IonExpress Adapter
PCR
Index
Barcode Adapter
Sequence
Amplicon-based
Thermo Fisher
Random Sonication
Adapter Ligation
PCR
Probe Hyb.
PCR + Index
Probe
Index
Sequence
SureSelect
SeqCap
Capture hybridization-based
Agilent
Nimble gen Roche

METHODS

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

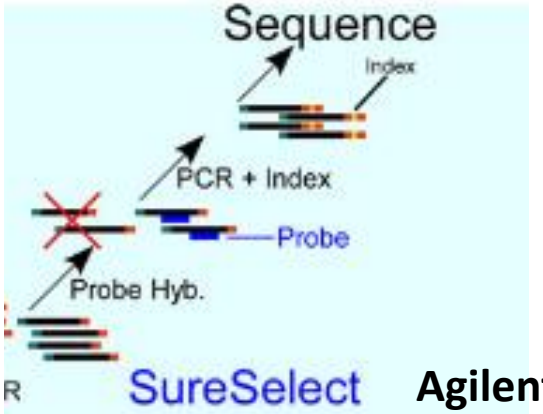
Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing

Eric Samorodnitsky,[1] Benjamin M. Jewell,[1] Raffi Hagopian,[1] Jharna Miya,[1] Michele R. Wing,[1] Ezra Lyon,[1] Senthilkumar Damodaran,[1,2] Darshna Bhatt,[1] Julie W. Reeser,[1] Jharna Datta,[1] and Sameek Roychowdhury[1,2,3]*

[1]Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio 43210; [2]Division of Medical Oncology, Department of Internal Medicine, The Ohio State University, Columbus, Ohio 43210; [3]Department of Pharmacology, The Ohio State University, Columbus, Ohio 43210

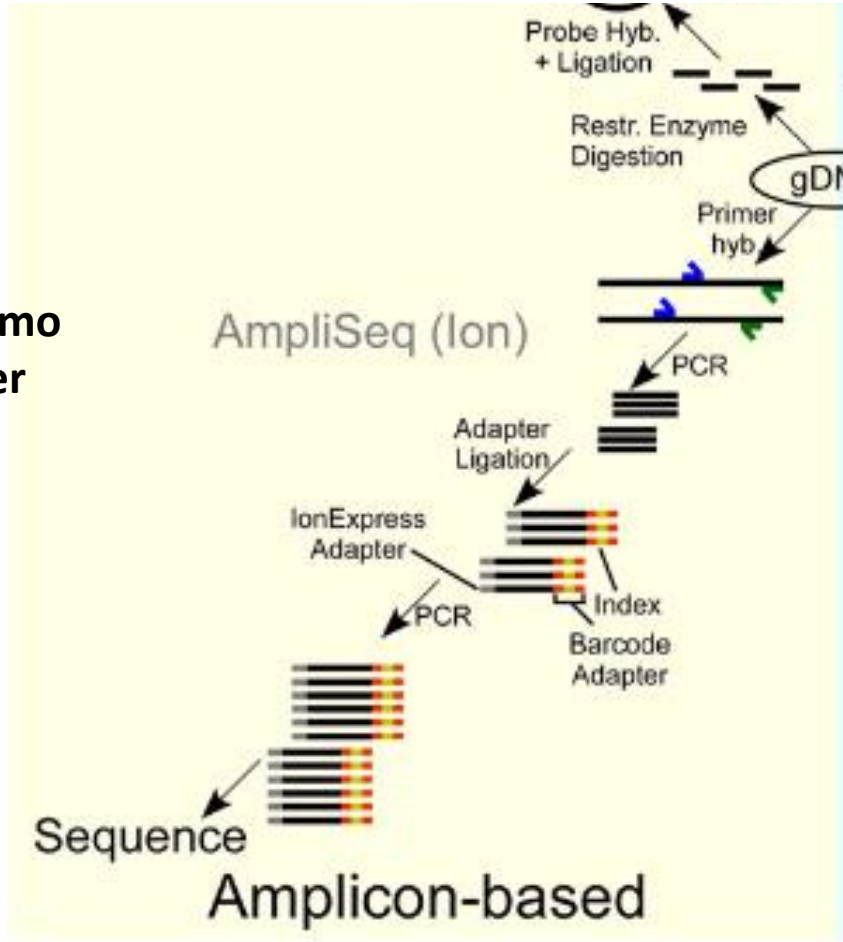
Communicated by Stephanie B. Taylor
Received 3 February 2015; accepted revised manuscript 11 June 2015.
Published online 25 June 2015 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22825

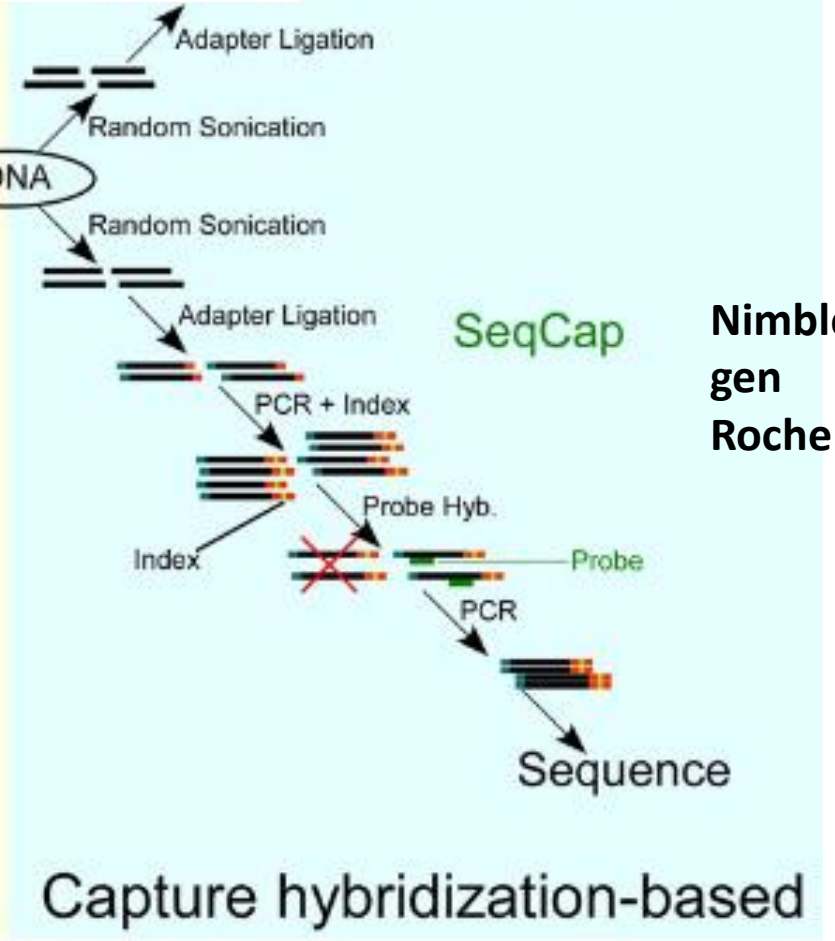
Agilent



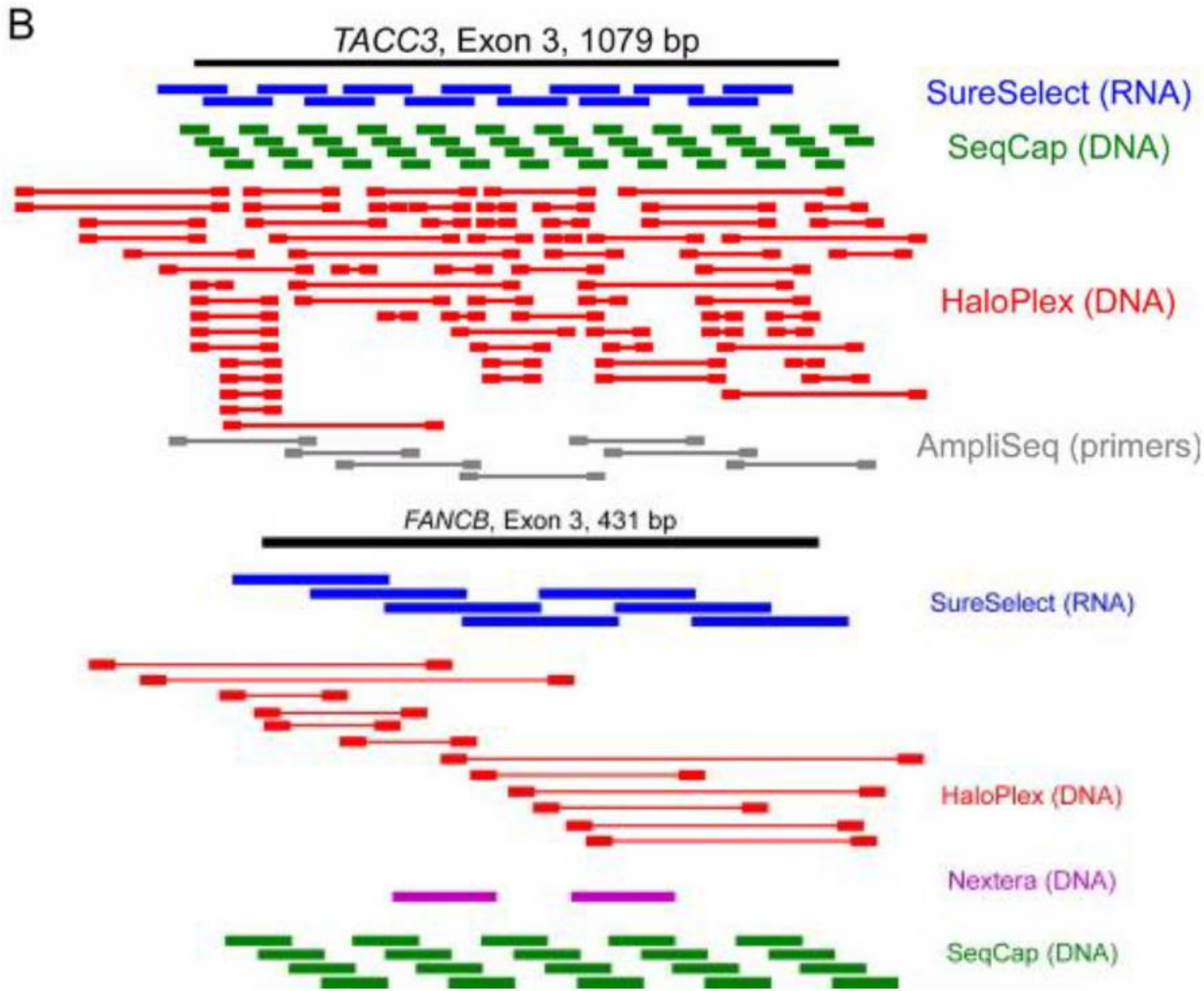
Agilent

Thermo Fisher

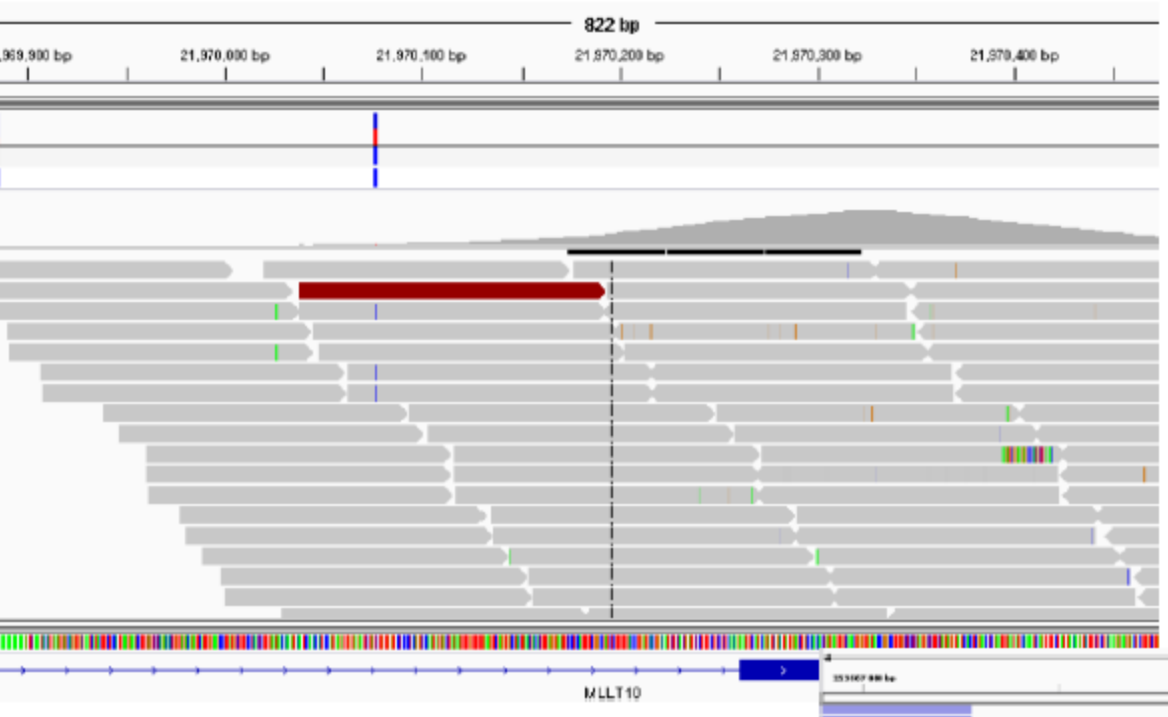


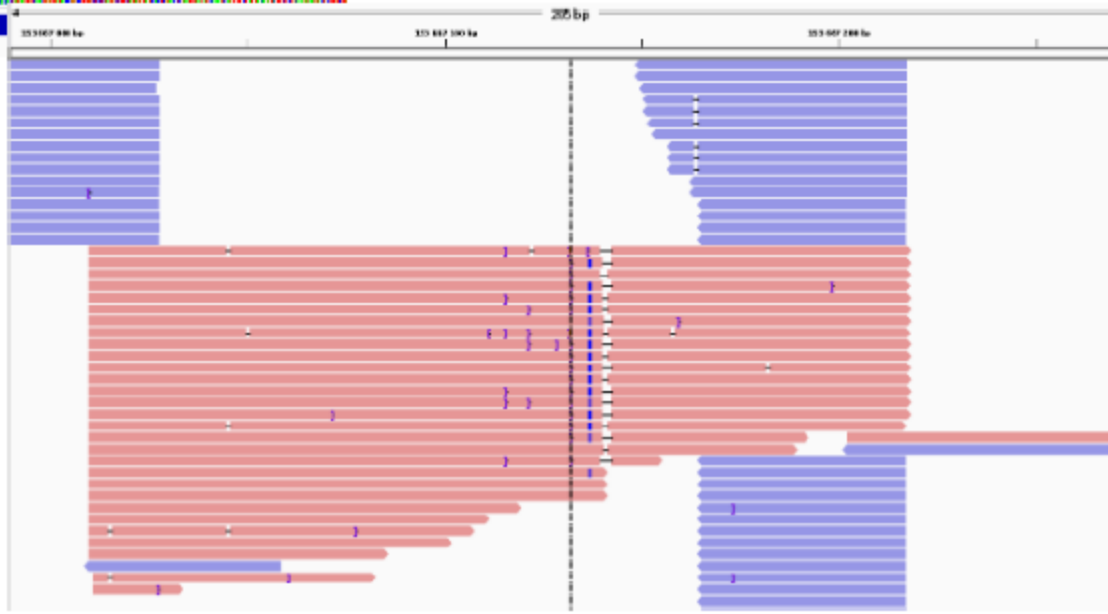


Nimble gen Roche


B
TACC3, Exon 3, 1079 bp
SureSelect (RNA)
SeqCap (DNA)
HaloPlex (DNA)
AmpliSeq (primers)
FANCB, Exon 3, 431 bp
SureSelect (RNA)
HaloPlex (DNA)
Nextera (DNA)
SeqCap (DNA)

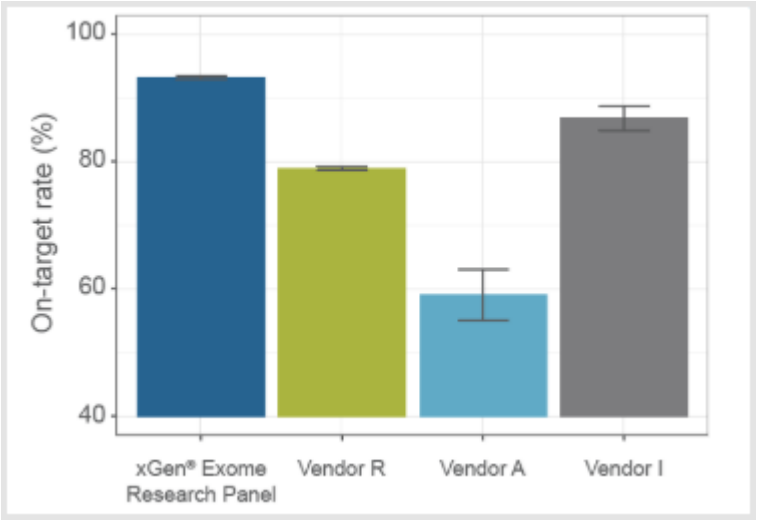
Амплификация или гибридизация?

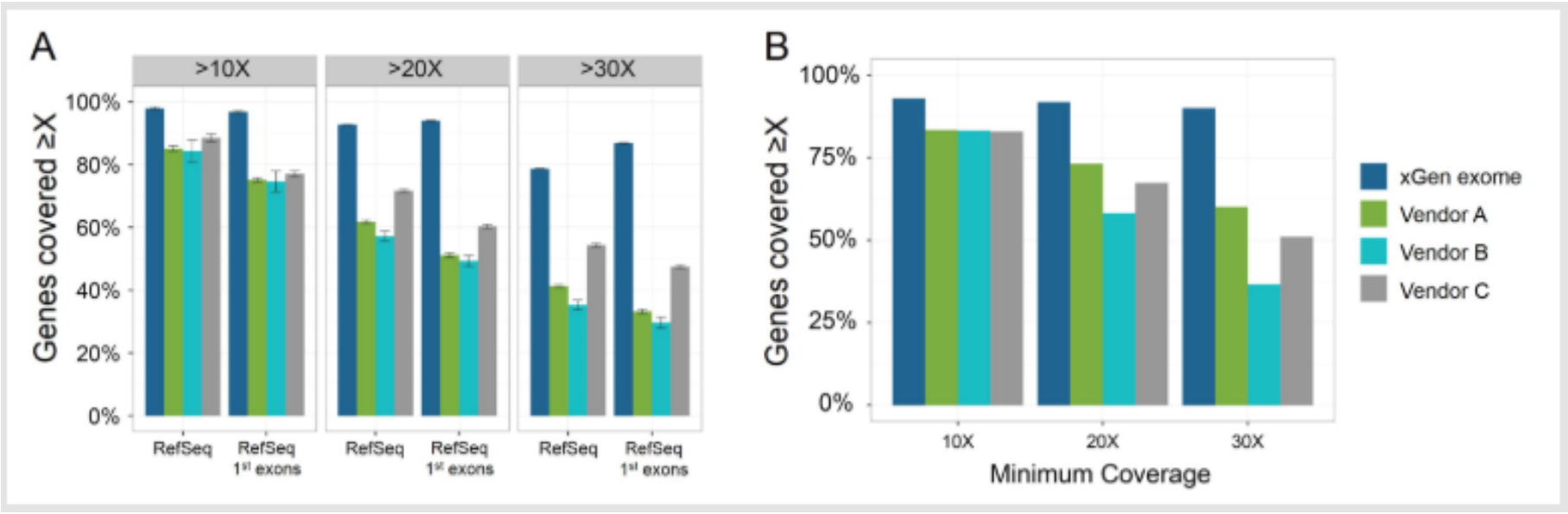




xGen® Exome Research Panel

- 429,826 xGen Lockdown® Probes
- 39 Mb target region (19,396 genes)
- covers 51 Mb of end-to-end tiled probe space
- GMP standards
- QC of each probe





SureSelect Human All Exon V7 is the most comprehensive exome

	RefSeq	GENCODE	CCDS	UCSC Known Genes
SureSelect Human All Exon V7	**99.3%**	**99.6%**	**99.6%**	**99.6%**
Vendor ID	97.3%	97.1%	98.3%	94.5%
Vendor R	96.9%	97.2%	97.5%	96%
Vendor I	98.8%	99.1%	99.5%	98.3%
Vendor T	96.9%	97.1%	99.9%	93.7%

Table 2. Protein coding regions from each database were downloaded from UCSC Genome Browser.

Nimblegen / Roche

Catalog (predesigned)
SeqCap EZ MedExome Kit 47 Mb whole exome enrichment design
SeqCap EZ Exome v3.0 Kit 64 Mb sequence capture design
SeqCap EZ Exome Plus Kit 64 Mb exome capture with the ability to add up to 200 Mb of custom targets
SeqCap EZ Exome +UTR Enrichment Kit 96 Mb

Agilent

SureSelect

HaloPlex

SureSelect Focused Exome – 12 Mb

HaloPlex Exome – 37Mb 21,522 genes

SureSelect Human All Exon V6 – 60 Mb

SureSelect Clinical Research Exome V2 – 67.3 Mb

ClearSeq Inherited Disease
2,742 genes known to cause inherited disorders

OneSeq Constitutional Research Panel:
300 kb functional copy number resolution genome-wide, with higher resolution of 25-50 kb in disease-associated ClinGen regions, copy-neutral LOH as small as 5 Mb
PLUS SureSelect Focused Exome

ThermoFisher

Ion AmpliSeq panel	Genes covered	Average gene coverage	Panel uniformity	Percentage reads on target
Inherited Disease Panel	325	97%	91%	96%
Cardiovascular Research Panel	404	99%	90%	97%
Hematology Research Panel	394	99%	94%	97%
Neurological Research Panel	751	99%	92%	97%
Ophthalmic Research Panel	316	99%	89%	98%
Deafness Research Panel v2	128	99%	94%	95%
Dermatology Research Panel v2	222	99%	96%	96%
Dysmorphia-Dysplasia Research Panel v2	519	99%	96%	95%
Endocrine Research Panel v2	340	99%	93%	96%
Gastrointestinal Research Panel v2	194	99%	97%	95%
Inborn Errors of Metabolism Research Panel v2	594	99%	96%	97%
Primary Immune Deficiency Research Panel v2	264	99%	95%	98%
Pulmonary Research Panel v2	131	98%	96%	95%
Renal Research Panel	155	99%	96%	95%
Epilepsy Research Panel	386	99%	91%	98%
Autism Research Panel v2	236	99%	94%	97%
Inherited Cancer Research Panel	134	99%	95%	96%
Cardiac Arrythmias and Cardiomyopathy Research Panel	92	99%	97%	98%
Hearing Loss Research Panel v1	63	96%	96%	91%
Dementia Research Gene Panel	17	99%	95%	87%
Noonan Research Panel	14	100%	93%	98%
TP53 Research Panel	1	100%	88%	97%
BRCA1 & *BRCA2* Research Panel	2	100%	97%	98%
Ovarian Cancer Research Panel	41	99%	99%	97%
CFTR Research Panel	1	100%	98%	94%
Pharmacogenomics Research Panel	40	100%	98%	82%
Exome RDY Panel	19,072	96%	92%	90%

Illumina

	Nextera Rapid Capture Exome and Expanded Exome Kits	TruSeq Exome	TruSeq Rapid Exome Library Prep Kit
Assay Time	1.5 days	2.5 days	1 day
Hands-On Time	< 5 hours	6 hours	3 hours
Input Quantity	50 ng genomic DNA	100 ng genomic DNA	50 ng genomic DNA
Content Specifications	Nextera Rapid Capture Exome: CDS. 45 Mb, Nextera Rapid Capture Expanded Exome: CDS, UTRs, and miRNAs. 62 Mb	CDS. 45 Mb (≥98% of RefSeq, CCDS, and Ensembl coding content)	CDS. 45 Mb (≥98% of RefSeq, CCDS, and Ensembl coding content)
Mechanism of Action	Transposase-based fragmentation and exome enrichment with biotinylated capture probes	Covaris fragmentation and exome enrichment with biotinylated capture probes	Transposase-based fragmentation and exome enrichment with biotinylated capture probes
Specialized Sample Types	Low Input	FFPE	Low Input

TruSight One Sequencing Panels – 4813 – 6700 genes associated with human disease
TruSight One – ~12 Mb
TruSight One Expanded – ~16.5 Mb

«Клинический» экзом

– совокупность генов, патогенные варианты в которых приводят к развитию наследственных (моногенных) заболеваний

2017 год

Type of lesion in HGMD	Variant class	Number of gene entries
Disease-causing mutations only	DM or DM?	4562
Disease-associated/functional polymorphisms only	DP, FP or DFP	1425
Disease-causing mutations and disease-associated/functional polymorphisms	DM or DM? and DP, FP or DFP	1765

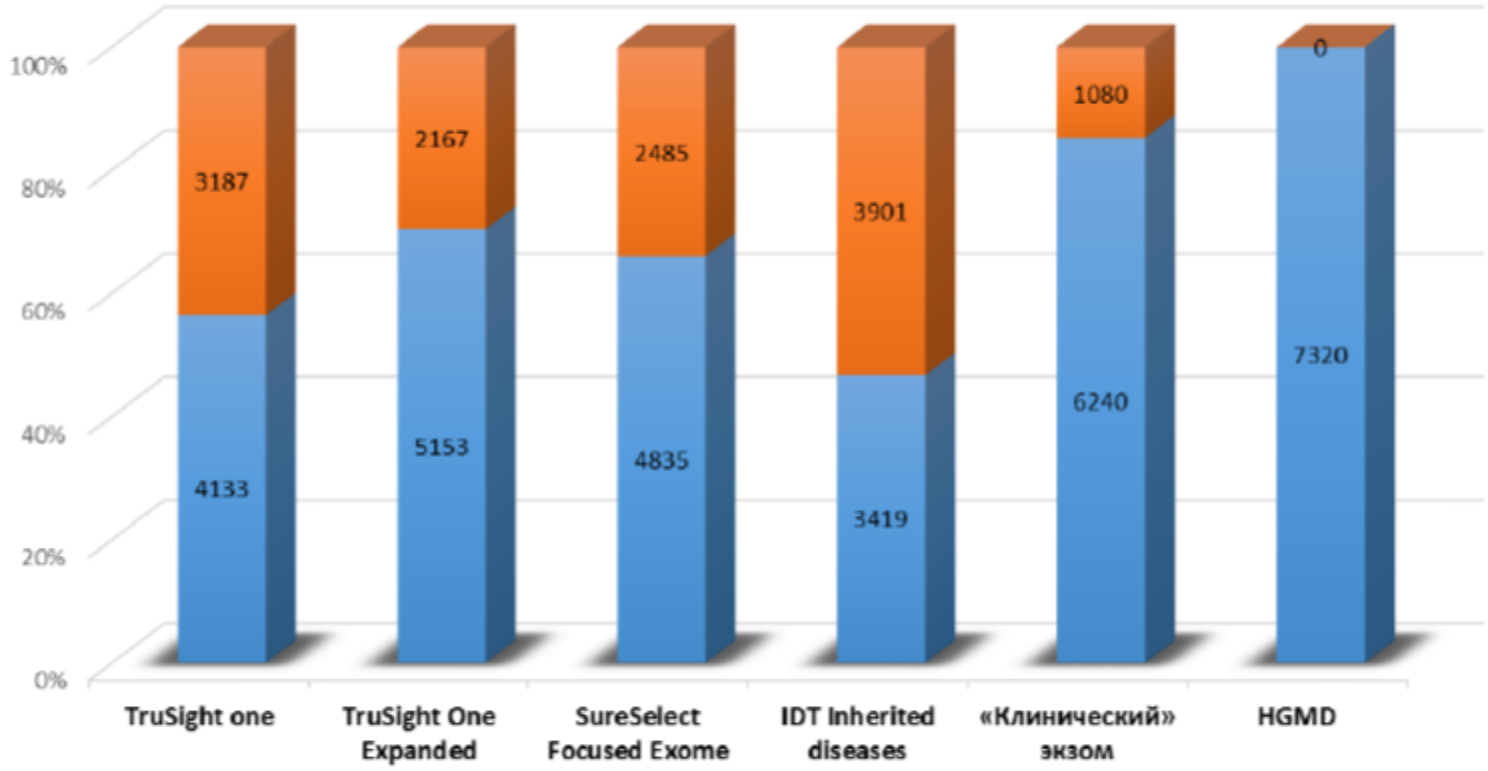
6327

2018 год

Type of lesion in HGMD	Variant class	Number of gene entries
Disease-causing mutations only	DM or DM?	5276
Disease-associated functional polymorphisms only	DP, FP or DFP	1421
Disease-causing mutations and disease-associated/functional polymorphisms	DM or DM? and DP, FP or DFP	2044

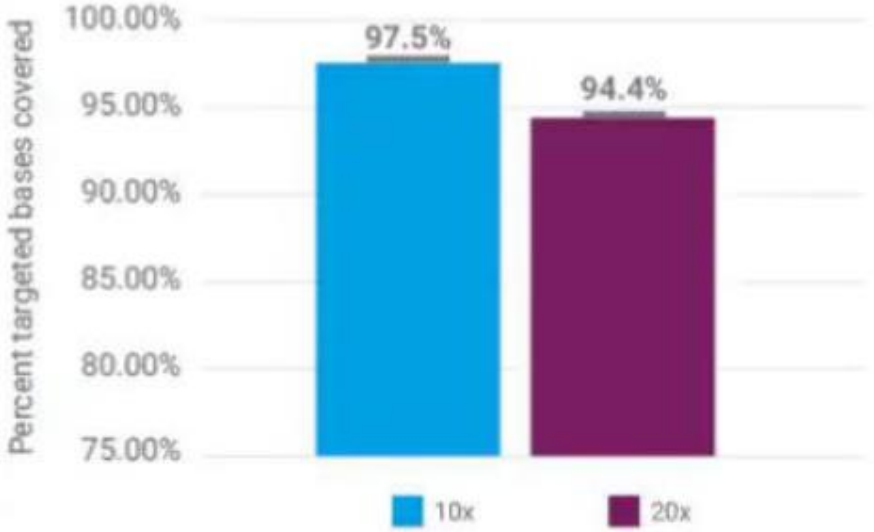
7320

Сравнение КЭ с базой HGMD



751 ген из базы HGMD (2018) не представлен ни в одном экзоме

SureSelect Human All Exon V7 provides superior coverage of targeted content



Design Size	48.2 Mb
Target Size	35.7 Mb
Sequencing cost	$151

Sequencing cost based on 5.3Gb sequencing on HiSeq3000

От панелей к геному

- Полный экзом
- «Клинический» экзом
- Небольшие панели
- Онкопанели

- Геном
  - 600 $ -> 100 $
  - CNV
  - Любые перестройки
  - Идеальное покрытие
  - Любые мутации

Через 5 лет останутся геном + онкопанели +метагеномика?

Проблемы и ограничения при анализе омных данных:

1. Allele dropout – при обогащении методом ПЦР, гетерозиготная мутация, например, рядом с праймером, праймер не работает => ложный гомозиготный немутированный вариант, а мутацию потеряем (или наоборот).

2. Пробелы в покрытии - все наборы для экзомного обогащения не поднимают полностью те участки, которые должны поднять. Мутации в таком участке теряем.

3. Неисчерпывающие списки генов в наборах обогащения. «Клинический экзом», в несколько тысяч генов, но интересные гены, потенциально связанные с заболеванием могут быть в "клинический экзом" не включены.

4. Ошибки пробоподготовки, заметные на QC: сбитый GC-content, плохое качество прочтений

5. Повторы и псевдогены

6. Strand bias: разные варианты на прямых и обратных прочтениях и вообще представленность цепей

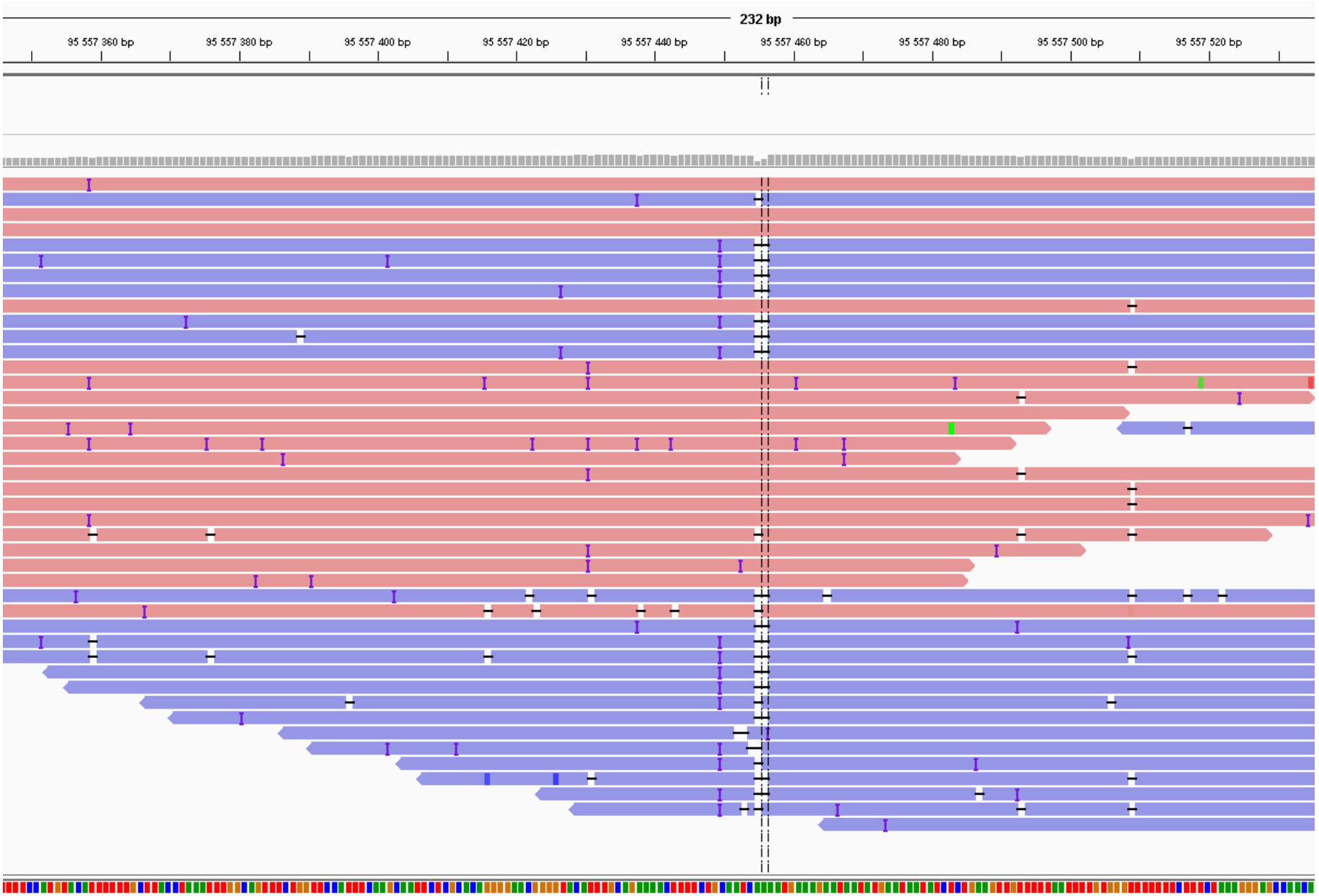
7. Ошибки на гомополимерах (Thermosifher) и падение качества прочтения к концу прочтения и сложность чтения GC (Illumina, BGI)

8. CNV и тандемные повторы

Strand bias


100 bp
106 857 310 bp
106 857 320 bp
106 857 330 bp
106 857 340 bp
106 857 350 bp
106 857 360 bp
106 857 370 bp
106 857 380 bp
106 857 390 bp

Strand bias


232 bp
95 557 360 bp
95 557 380 bp
95 557 400 bp
95 557 420 bp
95 557 440 bp
95 557 460 bp
95 557 480 bp
95 557 500 bp
95 557 520 bp