

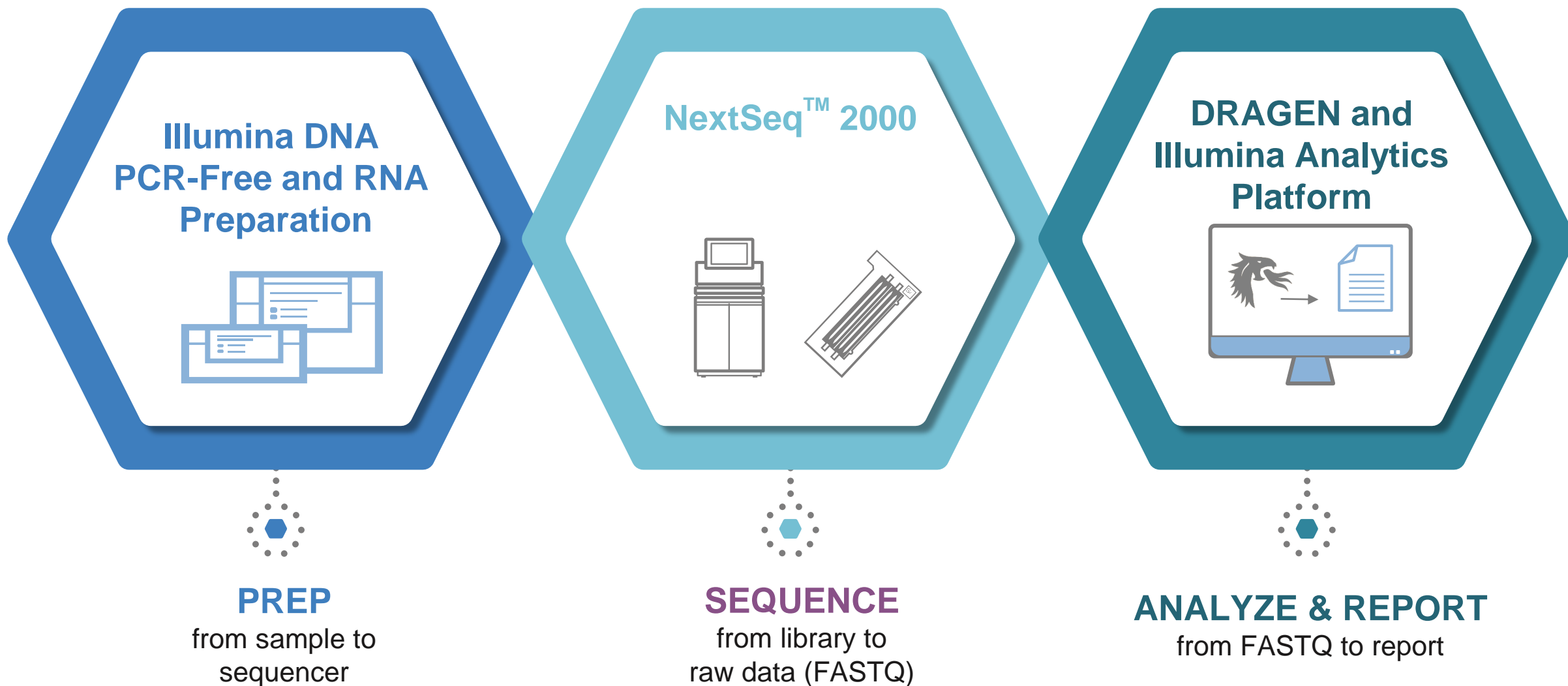
Expanding Sequencing Capabilities with the Illumina NextSeq2000

Ryan Hegarty | Sales Specialist, Sequencing

November 12, 2020



Improvements and New Releases Across the Illumina Portfolio



Introducing the new Illumina RNA Library Prep Suite

Built for quality, scalability and flexibility

Preliminary Specifications

16 & 96

Sample kit sizes

384

Unique Dual Indexes³

	1	2	3
	Illumina Stranded Total RNA Ligation with Ribo-Zero Plus	Illumina Stranded mRNA Ligation	Illumina RNA Prep with Enrichment (L) Tagmentation
Detection	Coding & Non-coding regions	Coding transcriptome w/ Poly A tail	Targeted coding region ⁴
FFPE Compatibility	✓	-	✓
Input	1-1,000 ng ¹ 10ng for optimal quality & FFPE	25-1,000 ng	10ng non-FFPE 20ng FFPE
Total Time (hours) ²	7	6.5	< 9
Hands-on time (hours) ²	< 3	< 3	< 2
Automation Friendly	✓	✓	✓
	<ul style="list-style-type: none">Includes Ribo-Zero Plus for multi-species rRNA depletionIncludes cDNA synthesis reagents	<ul style="list-style-type: none">Includes Illumina Poly A capture kitIncludes cDNA synthesis reagents	<ul style="list-style-type: none">Illumina Tested with Illumina Exome & Illumina Respiratory Viral Panel

1. Minimum input for high-quality RNA shown, 10ng minimum recommended for optimal quality and FFPE for Total RNA workflow

2. Hands-on and total time based on manual processing of up to 24 samples for Total & mRNA workflows and 1 sample on Enrichment workflow

3. Up to 192 UDIs available at launch, up to 384 available later in 2020

4. Note new Illumina RNA Prep with Enrichment does not provide strand information (is non-stranded)

Ribo-Zero Plus Enzymatic Depletion Methodology

Robust method allows for increased flexibility for mix sample labs

Total RNA



1 Hybridize probes



2 Deplete rRNA



3 Remove probes



— RNA of interest — Ribo-Zero Probes — Abundant rRNA

One Tube to deplete multiple species with Ribo-Zero Plus
Provides seamless study flexibility for mixed sample labs

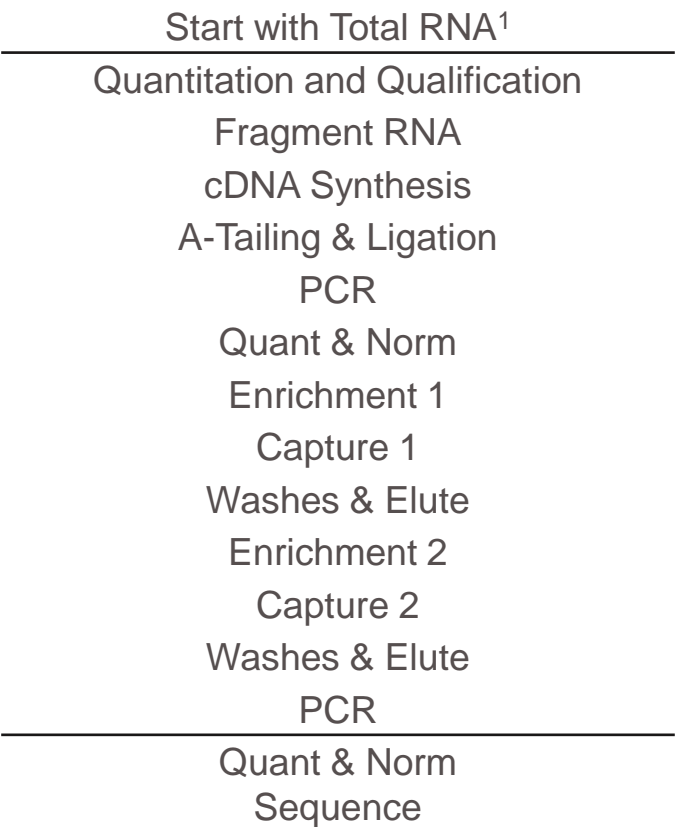


- Cytoplasmic and mitochondrial rRNA – human/mouse/rat
- Globin transcripts
- Bacterial rRNA (Gram+, Gram-)

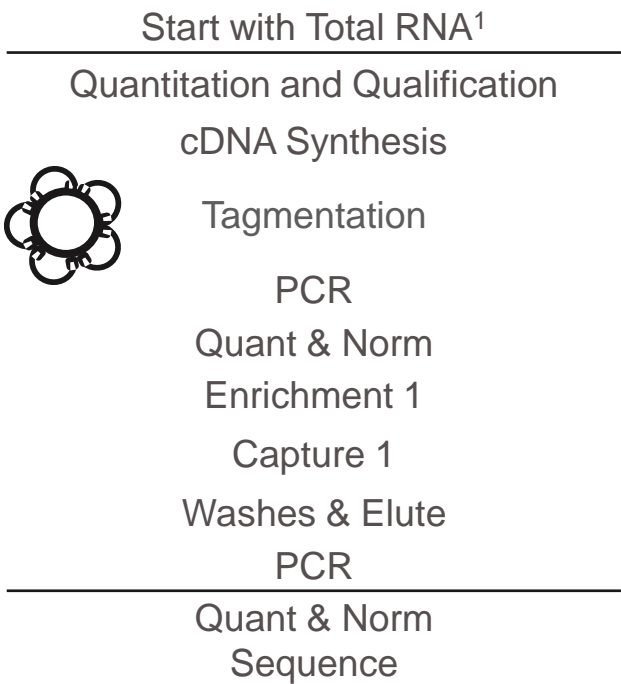
Depletion target	rRNAs targeted
Human Cytoplasmic rRNAs	28S, 18S, 5.8S, 5S
Human Mitochondrial rRNAs	12S, 16S
Human Beta Globin transcripts	HBA1, HBA2, HBB, HBG1, HBG2
Mouse and Rat rRNA	16S, 28S
Gram(-) Bacterial rRNAs	<i>E.coli</i> : 5S, 16S, 23S
Gram(+) Bacterial rRNAs	<i>B. subtilis</i> : 5S, 16S, 23S

NEW Illumina RNA Prep with Enrichment workflow

Legacy TruSeq RNA Exome



NEW Illumina RNA Prep with Enrichment



- No Fragmentation
- Tagmentation replaces A-Tailing & Ligation
- Single Hybridization

>50% Faster²
>80% Less Hands-on²

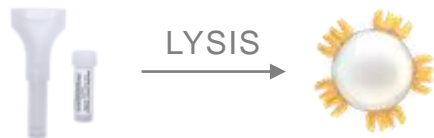
¹ Before starting the workflow, quantify the total RNA using standard methods and assess quality using a fragment analysis method
² Compared to TruSeq RNA Exome based on 2+ days total time and 10+ hours hands-on time vs. <9 hours total time and <2 hours hands-on time for New Illumina RNA prep with Enrichment

Illumina DNA PCR-Free Library Prep

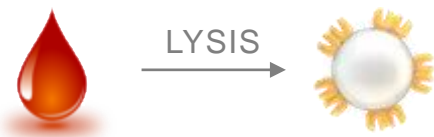
Optional front-end extraction to direct input into PCR-free workflow

Direct sample input

30ul saliva



10ul blood



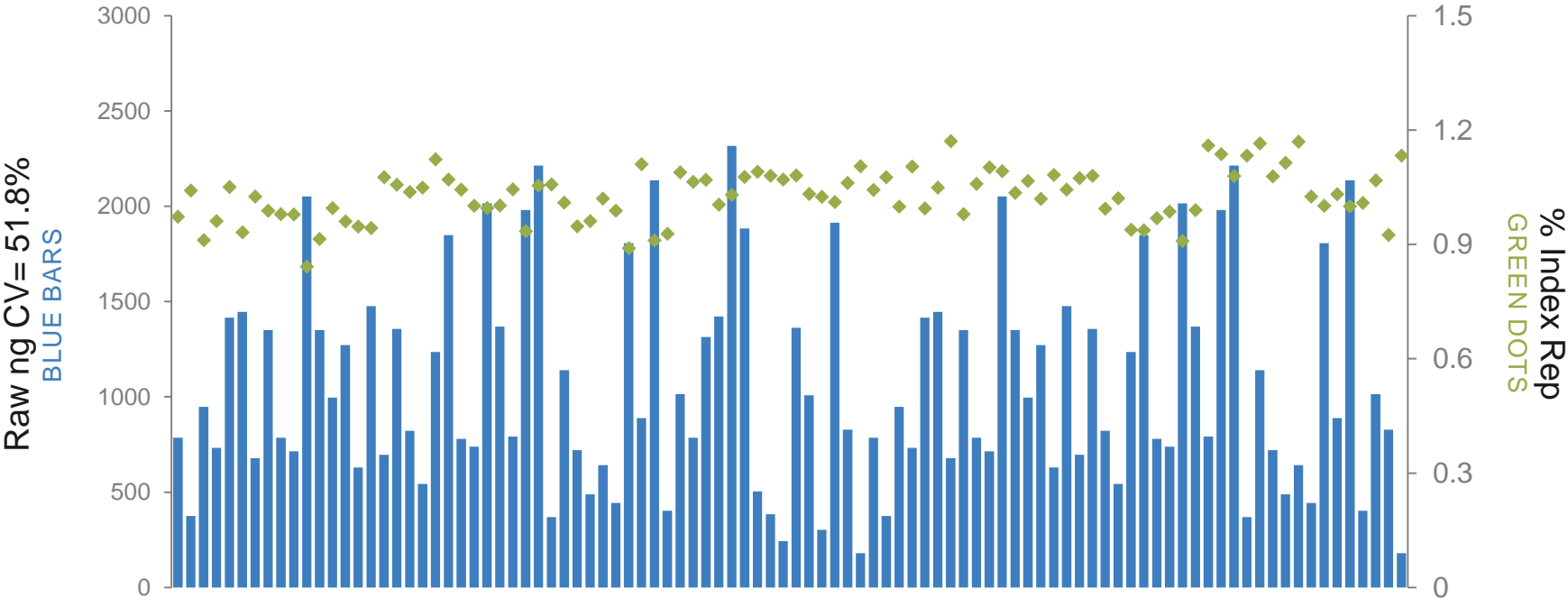
1/2 loop from culture



*Internal data on file

Sample input versus index representation after data analysis

Index rep CV = 6.5%



NextSeq 2000

Deeper, larger studies, on a benchtop platform.



**NextSeq 1000/2000
P2 Flow Cell**
120 Gb
Max output



**NextSeq 2000
P3 Flow Cell**
330 Gb
Max output

RELEASED TODAY

Current NextSeq 2000 Customers

Access new features through simple customer installable upgrades



NextSeq™ 2000
system
(ICS v1.1.0)

ICS v1.2.0

DRAGEN v3.7.4

Custom primers
& recipes



New P2 & P3
consumables

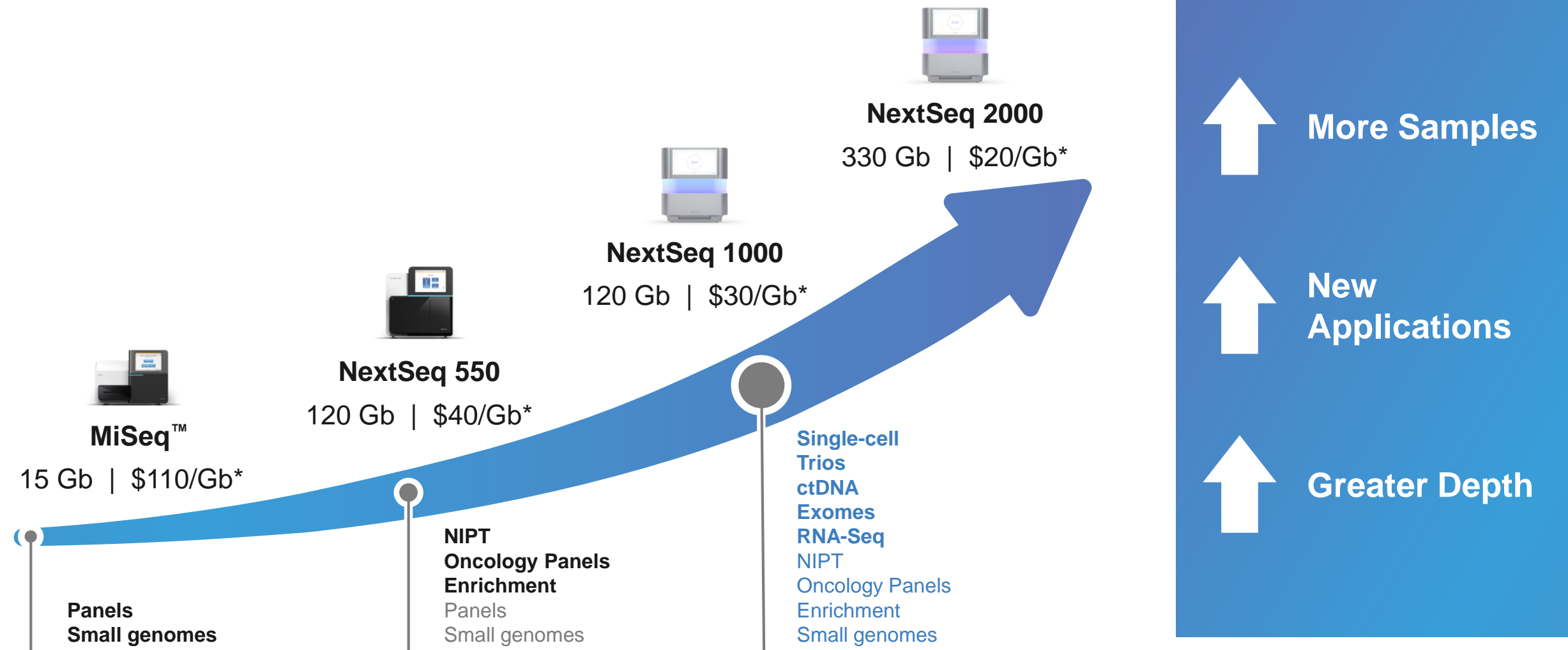


On board data
compression



New DRAGEN
analysis pipelines

Nextseq 1000 And Nextseq 2000 Provide Sequencing Power for High-throughput Applications



Addition of P3 50 Cycle Kit

Kit	List	\$/G (list)	\$/M read
P2 100 cycles (40G)	\$1,420	\$35.50	\$3.55
P2 200 cycles (80G)	\$2,670	\$33.38	\$6.68
P2 300 cycles (120G)	\$3,540	\$29.50	\$8.85
P3 50 cycles (55G)	\$2,250	\$40.91	\$2.05
P3 100 cycles (110G)	\$3,250	\$29.55	\$2.95
P3 200 cycles (220G)	\$4,500	\$20.45	\$4.09
P3 300 cycles (330G)	\$6,000	\$18.18	\$5.45

Applications

Infectious disease:

Small genome characterization (Currently not validated for COVIDSeq)

Proteomics:

Antibody-linked oligo tags BioLegend Cite Seq.

Spatial transcriptomics:

NanoString's GeoMx (27bp)

Small RNA analysis:

50 bp reads will sequence most small RNAs

Other counting applications

NextSeq 2000 is the First Systems to Integrate DRAGEN Bio-IT Platform On-Board

DRAGEN Bio-IT platform:

- Fast
- Accurate
- Cost efficient
- Industry standard pipelines
- Great for both novice and expert users.

Pipelines available on-board at launch:

- Dragen Enrichment pipeline
- Dragen RNA pipeline
- Dragen Germline
- Generate FASTQ
- *Additional pipelines available in BSSH*



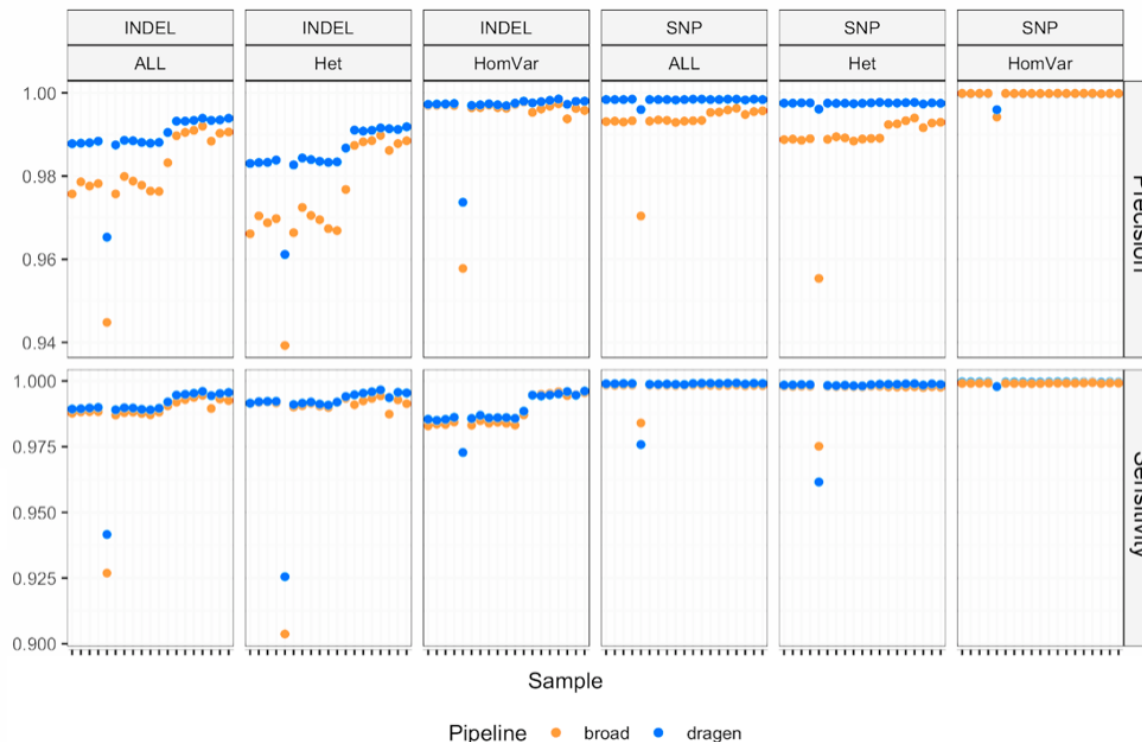
DRAGEN™ wins PrecisionFDA Truth Challenge V2 for Difficult-to-Map regions and All Benchmark Regions on Illumina sequencing data





The GATK team's evaluation confirmed DRAGEN™ accuracy gains

”..we wanted to do our own evaluation ... and long story short, we saw the same overall improvements in sensitivity and specificity, which you can see for yourself in the figure below.”



“finally got a taste of that famous acceleration -- yeah it is *fast*, no kidding”

THE GATK TEAM

Content published on Feb. 19, 2020 DRAGEN demonstrated gains in indel and heterozygous SNP calling precision & Increase in indel sensitivity
<https://gatk.broadinstitute.org/hc/en-us/articles/360039984151-DRAGEN-GATK-Update-Let-s-get-more-specific>

DRAGEN Single Cell RNA on NextSeq2000

Outputs functional starting point for downstream single cell analysis: Cell x Gene Expression Matrix

FEATURES

Ultra-rapid

Accurate analysis in < 2 hours for a full NextSeq 2000 P3 flow cell

Widely compatible

Supports a range of input library prep types for compatibility with downstream analysis tools

Set up and walk away

Goes from run set up to quantified expression per cell with a single touch point

BENEFITS

Accelerate your research

On-board QC of single cell expression libraries and single cell analysis pipeline executed in a fraction of the time – without the need for additional compute hardware or sacrificing accuracy

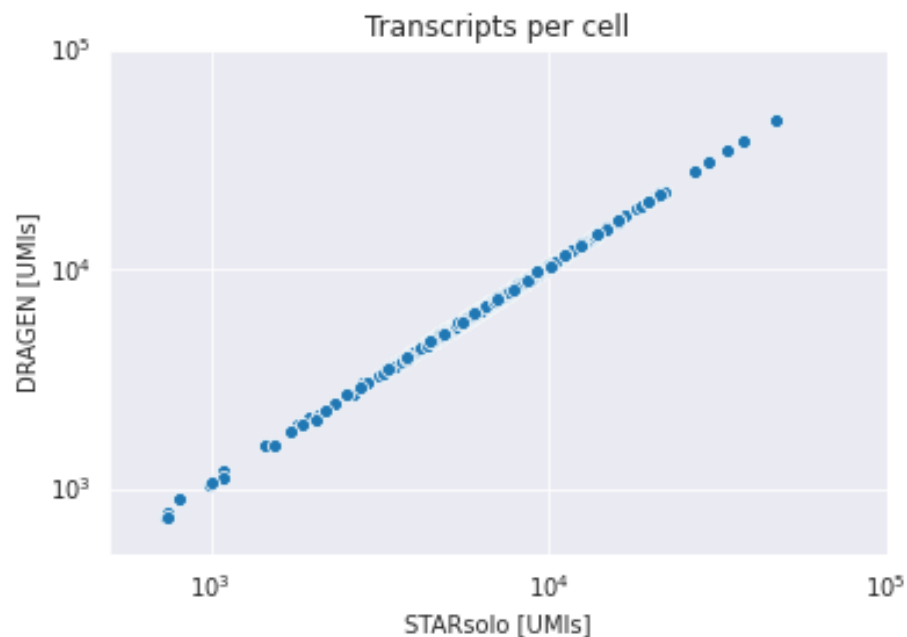
Choose your tools

Cell x gene matrix in an open data format supports a variety of library prep and downstream single cell analysis tools

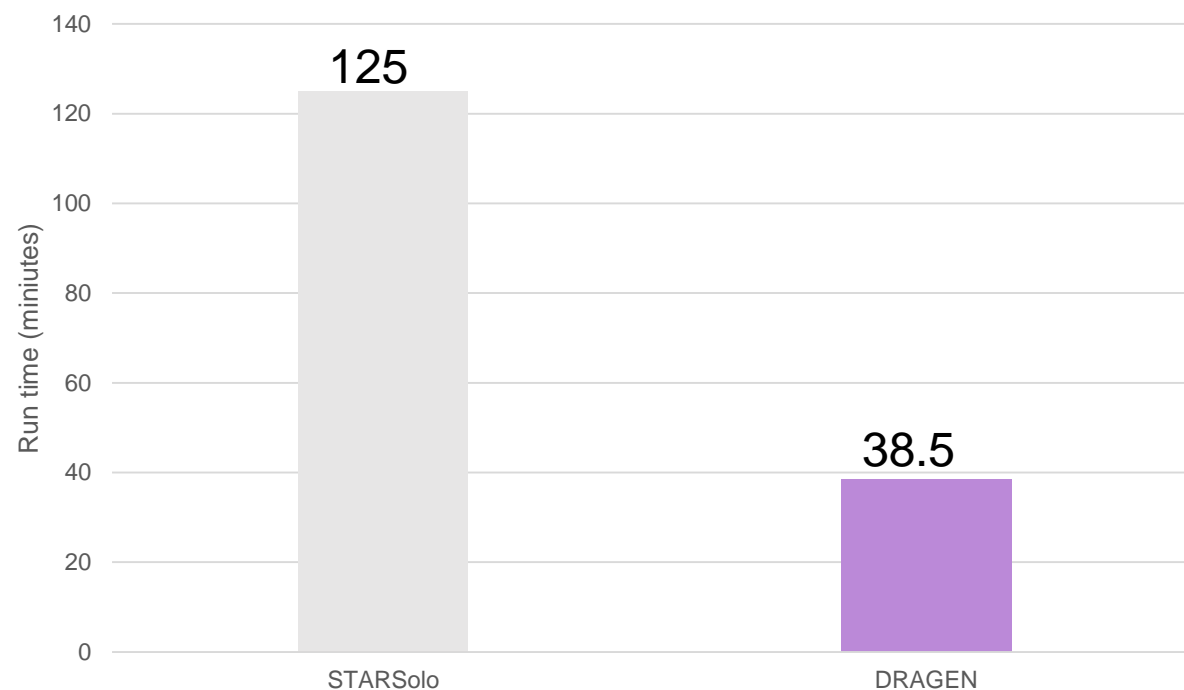
Streamline your work

Eliminate the set up & execution of a separate pipeline post-sequencing

Consistent with Established Tools, Yet Much Faster

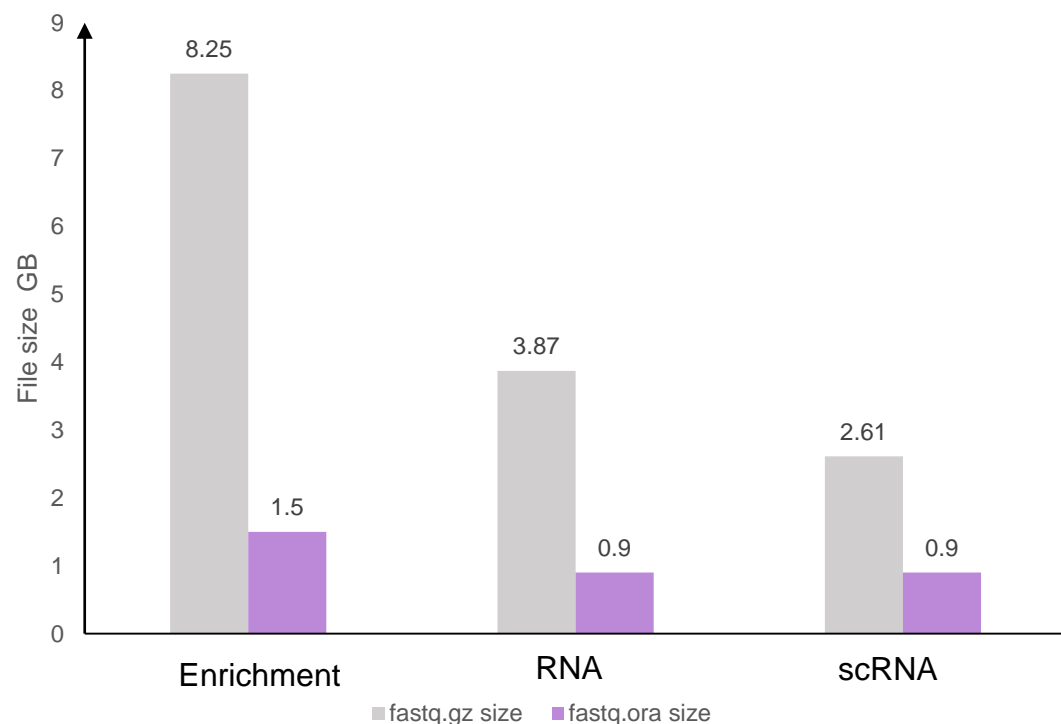


Correlation of per-cell gene expression
results with STARsolo: $r > 0.985$
857 / 858 cells overlap



8018 cell sample with 1.4 billion reads

New: DRAGEN Compression Reduces Data Storage Costs



Save ~ \$3,000 - \$10,000 on FASTQ storage costs*

*Compared to storing fastq.gz. Assumptions: 36 to 125 runs per year, 300 GB fastq.gz file sizes, compression ratio of 5, files stored 1 year in hot storage + 2 years in cold storage on AWS.



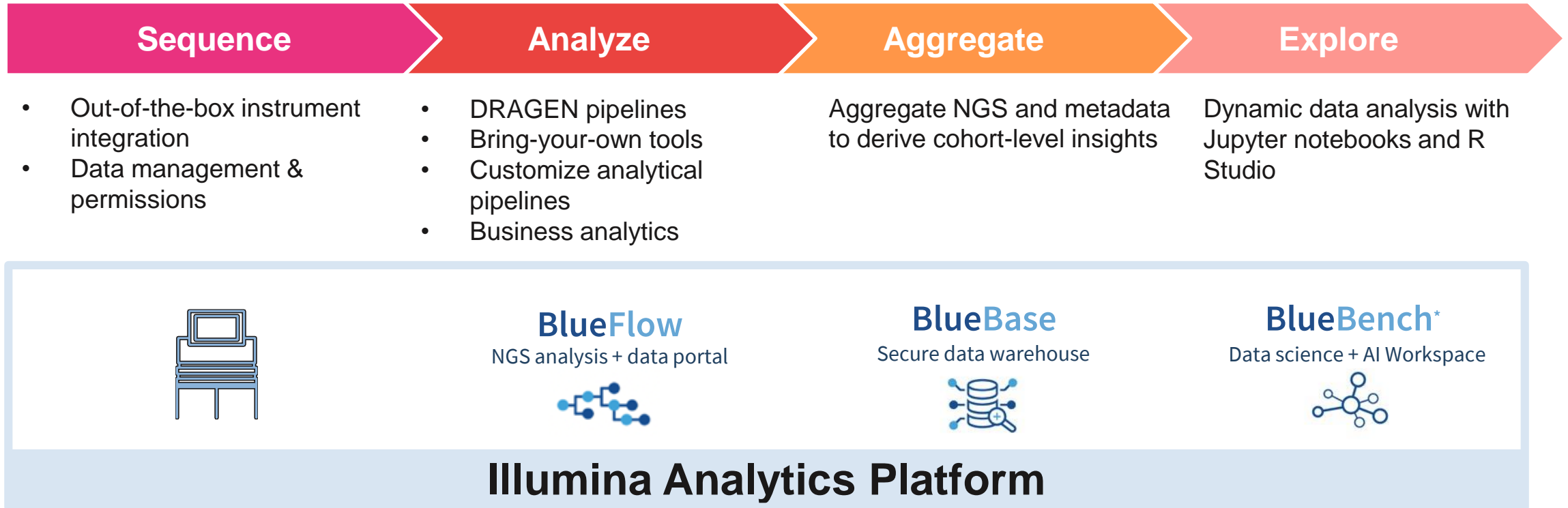
Reduce FASTQ file transfer times from 25 minutes to 5 minutes**

**Compared to transferring fastq.gz. Assumptions: 200 MB/s file transfer speeds and 300 GB fastq.gz file

Software license included in NextSeq 1000/2000 purchase

Illumina Analysis Platform (IAP) is a Comprehensive Platform to Drive Insights

Flexible, interoperable components supporting life sciences research



Data Production

Data Trending & Mining

Data Science

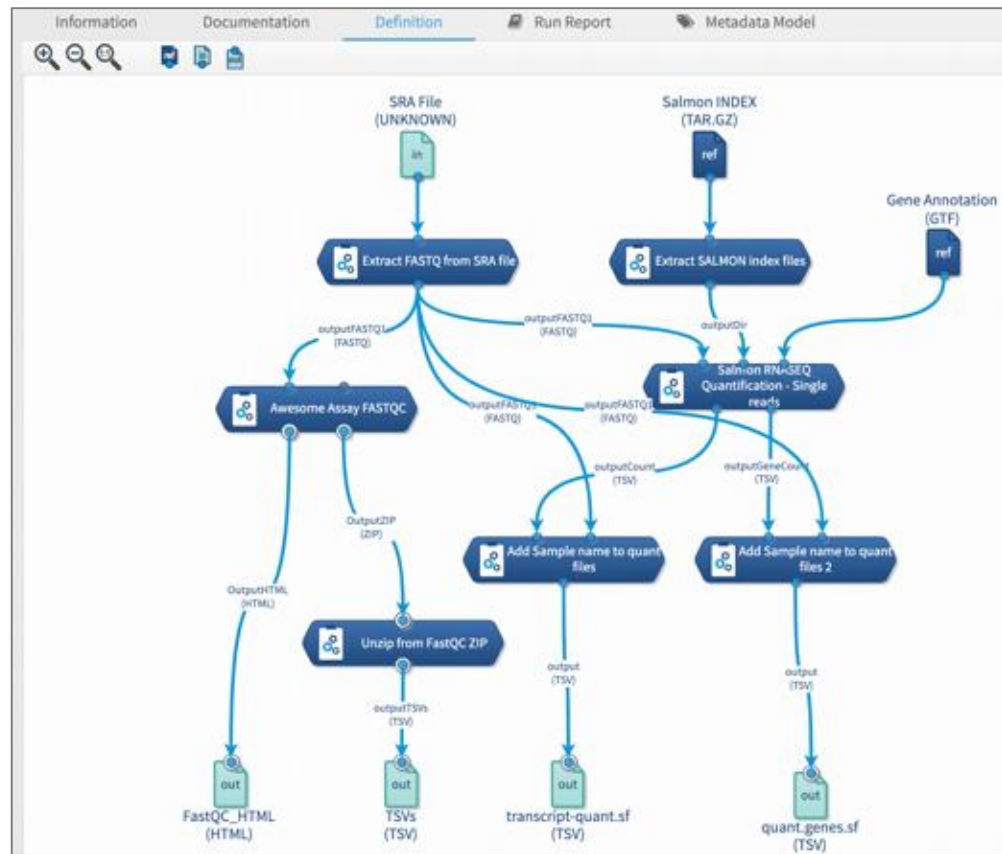
For Research Use Only. Not for use in diagnostic procedures.

illumina[®]

Interactive Portals & Data Science Workbench

Analyze

Drag & Drop Pipeline Creator



Aggregate

SQL Query of Data Warehouse

The screenshot shows the BlueBench Beta SQL Query interface. The query is as follows:

```
1 SELECT q.Sample, a.gene_name, a.gene_id, q.Length, q.EffectiveLength, q.TPM, q.NumReads, t.clinical_condition, t.source_name, t.gender
2 FROM gene_quant AS q
3 LEFT JOIN ( SELECT DISTINCT gene_name, gene_id
4             FROM gene_annotations ) AS a
5           ON a.gene_id = q.gene_id
6 LEFT JOIN SraRunTable t
7 ON t.Run = q.sample
8 WHERE q.NumReads > 0
9 ORDER BY q.TPM DESC;
```

The query returned 132738 rows taking 3s to process. The results are displayed in a table with the following columns: #, Sample, gene_name, gene_id, Length, EffectiveLength, TPM, NumReads, clinical_condition, source_name, and gender.

#	Sample	gene_name	gene_id	Length	EffectiveLength	TPM	NumReads	clinical_condition	source_name	gender
1	SRR3714714	MT-ATP8	ENSG00000228253.1	207.0	10.766	257215.0	107548.0	Crohn's disease	leftcolon	male
2	SRR3714758	MT-ATP8	ENSG00000228253.1	207.0	10.766	161077.0	74606.0	control	leftcolon	male
3	SRR3714658	MT-ATP8	ENSG00000228253.1	207.0	10.766	101859.0	70337.0	control	leftcolon	male
4	SRR3714641	MT-ATP8	ENSG00000228253.1	207.0	10.766	89564.3	36621.0	Crohn's disease	leftcolon	male
5	SRR3714714	MT-ND4L	ENSG00000212907.2	297.0	49.689	47925.6	92487.0	Crohn's disease	leftcolon	male
6	SRR3714714	MT-CO1	ENSG00000198804.2	1542.0	1293.0	41558.4	2086940.0	Crohn's disease	leftcolon	male
7	SRR3714714	MT-CO2	ENSG00000198712.1	684.0	435.0	38893.6	657082.0	Crohn's disease	leftcolon	male
8	SRR3714758	MT-CO3	ENSG00000198938.2	784.0	535.0	34639.5	797281.0	control	leftcolon	male
9	SRR3714714	MT-CO3	ENSG00000198938.2	784.0	535.0	34291.2	712506.0	Crohn's disease	leftcolon	male
10	SRR3714641	IGKC	ENSG00000211592.8	523.0	274.0	28907.4	300816.0	Crohn's disease	leftcolon	male
11	SRR3714758	MT-CO1	ENSG00000198804.2	1542.0	1293.0	28223.2	1569970.0	control	leftcolon	male
12	SRR3714758	MT-ND4L	ENSG00000212907.2	297.0	49.689	26989.7	57696.0	control	leftcolon	male
13	SRR3714658	MT-CO3	ENSG00000198938.2	784.0	535.0	26832.7	920760.0	control	leftcolon	male
14	SRR3714714	MT-ND3	ENSG00000198840.2	346.0	97.006	26698.3	100585.0	Crohn's disease	leftcolon	male
15	SRR3714758	MT-CO2	ENSG00000198712.1	684.0	435.0	24685.7	461977.0	control	leftcolon	male
16	SRR3714658	MT-ND3	ENSG00000198840.2	346.0	97.006	24438.1	152053.0	control	leftcolon	male
17	SRR3714714	MT-ATP6	ENSG00000198899.2	681.0	432.0	23926.6	401436.0	Crohn's disease	leftcolon	male

Interactive Portals & Data Science Workbench

Explore

Data Science Workbench with Notebooks

Cancer_Research > Workspaces

Press Enter to search

+ New Workspace

Image analysis using FASTAI

BlueBee - Data science for Python and R - 0.97

A workspace dedicated to doing deep learning model building on the MRI images we have for these patients. This is using the restricted access mode, since I do need to install some packages on the basic image, but cannot allow internet access once the image is running. It is using the FASTAI deep learning python library.

Large GPU 1000 GB

Starting

Output file analysis

BlueBee - Data science for Python and R - 0.97

A workspace created to analyse the output files from our pipeline. Used a closed workspace as there is personal information in the output files. As this requires quite some resources, we've chosen the "large" resource model but did not need a GPU.

Large 10 GB

Running

Tool Builder

BlueBee - Builder for Python and R - 0.97

This workspace is used to build the various tools and packages we need

VCF analysis using HAIL

BlueBee - HAIL for Python - 0.97

A workspace dedicated to the GWAS analysis of our genotype data using

Cancer_Research > Workspaces > VCF analysis using HAIL

Details Access Build Activity

File Edit View Run Kernel Tabs Settings Help

Launcher

Hail_on_1000genomes_samj

Python 3

Running on Apache Spark version 2.4.1
SparkUI available at http://bbench.tst-testplatform-w10551879.tst-testplatform-w10551879.svc.cluster.local:4040
Welcome to
HAIL version 0.2.34-914bd8a10ca2
LOGGING: writing to /data/hail-20200324-0007-0.2.34-914bd8a10ca2.log

Download public 1000 Genomes data

The workshop materials are designed to work on a small (~20MB) downsampled chunk of the public 1000 Genomes dataset.
You can run these same functions on your computer or on the cloud!

...

Explore genetic data with Hail

Read 1KG into Hail Like tables, matrix tables can be imported from a variety of formats: VCF, (B)GEN, PLINK, TSV, etc. Matrix tables can also be read from a "native" matrix table format. Let's read a sample of prepared 1KG data.
We represent genetic data as a Hail MatrixTable, and name our variable mt to indicate this.

(7) mt = hl.read_matrix_table('1kg.mt')

(8) mt.rows().show()

locus	alleles	rsid	qual	filters	info.AC	info.AF	info.AN	info.BaseQRankSum	info.ClippingRankSum	info.DP	info.DS	info.FS	info.Ha
locus<GRCh37>	array<str>	str	float64	set<str>	array<int32>	array<float64>	int32	float64	float64	int32	bool	float64	
1:904165	["G","A"]	NA	5.23e+04	NA	[518]	[1.03e-01]	5020	-3.39e+00	-1.70e-01	17827	false	2.23e+00	
1:909917	["G","A"]	NA	1.58e+03	NA	[18]	[3.73e-03]	4830	-1.48e+00	1.26e-01	14671	false	5.52e+00	
1:986963	["C","T"]	NA	3.98e+02	NA	[5]	[1.09e-03]	4588	1.25e+00	-3.77e+00	12398	false	8.34e-01	

0 Python 3 | idle Mode: Command Ln 1, Col 1 Hail_on_1000genomes_sample.ipynb

Thank You!



Frankie Gwynne

Account Manager

fgwynne@illumina.com



Paige Burres

Inside Sales

pburres@illumina.com



Ryan Hegarty

Sequencing Specialist

rhegarty@illumina.com