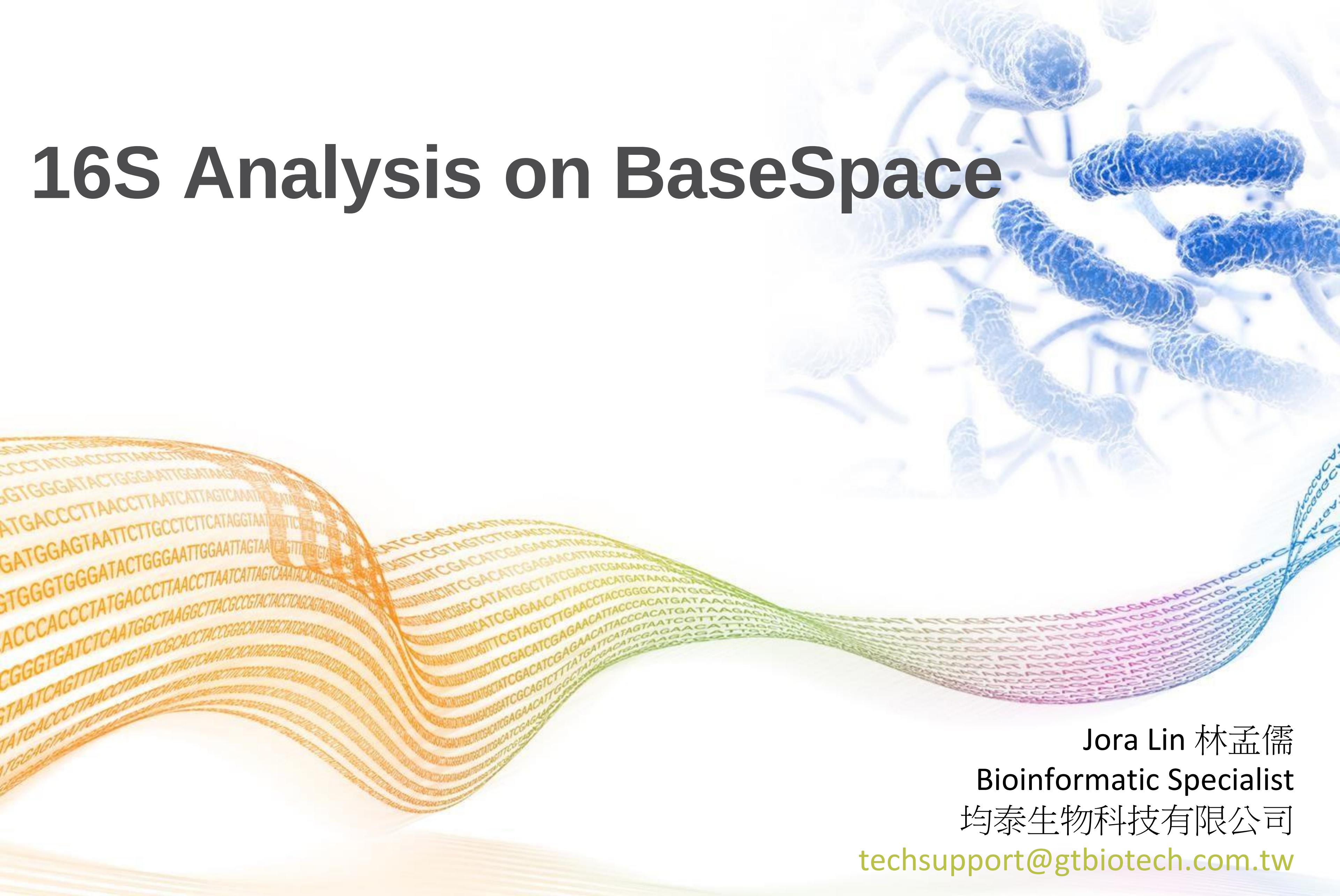
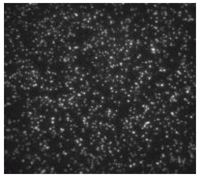
16S Analysis on BaseSpace



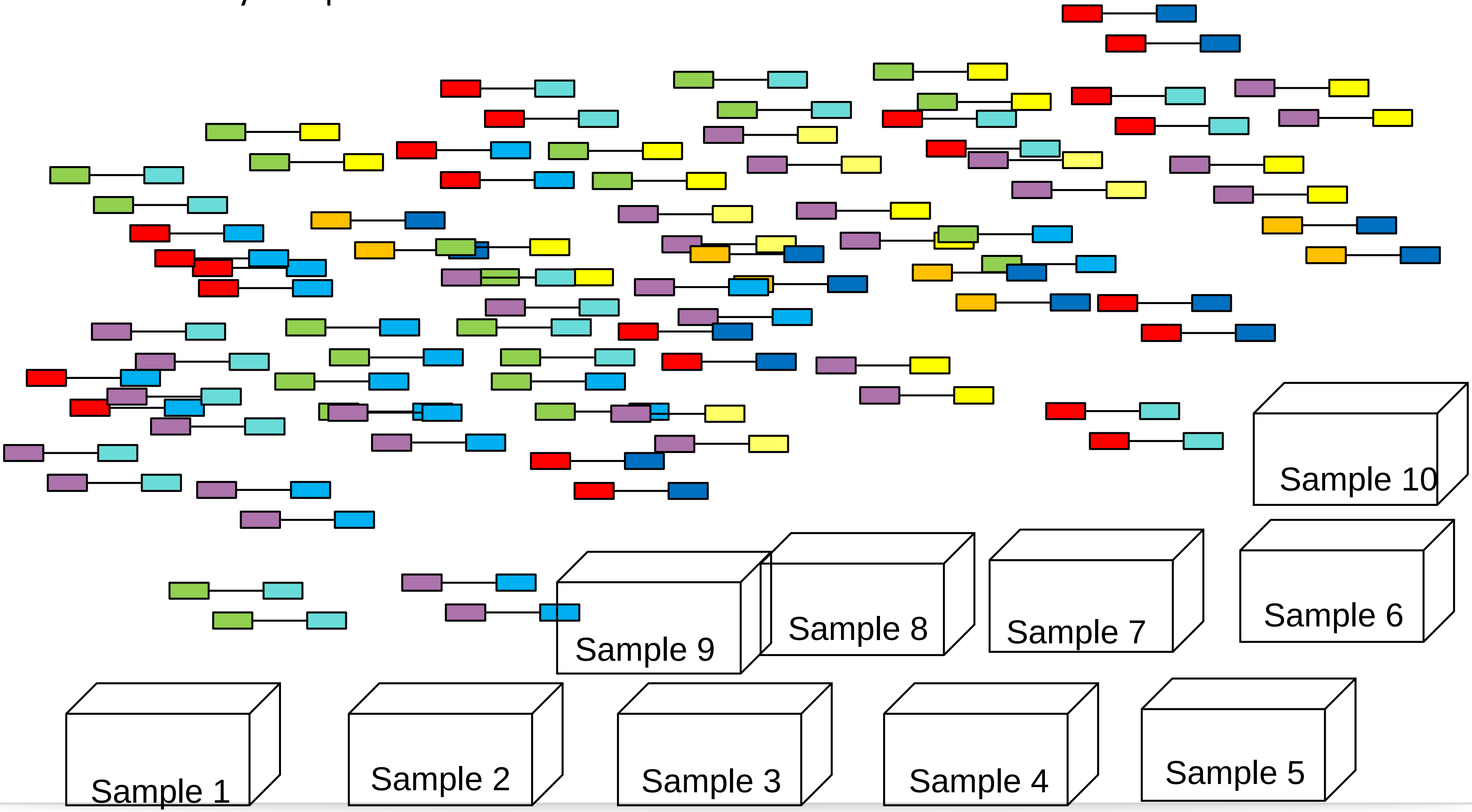
Jora Lin 林孟儒
Bioinformatic Specialist
均泰生物科技有限公司
techsupport@gtbiotech.com.tw

Sequencing Overview

Analysis Type	Software	Outputs
Control Software		  Images, Intensities and Base Calls
Analysis Software	BaseSpace	 Alignments, Variant Detection
Visualization Software	BaseSpace SEQUENCE HUB 	CSV 1,2,3 PDF Annotation, Filtering, Reports

Demultiplexing

Multiplexed samples are sequenced together
Unique index sequences allow sequencing data reads
to be sorted by samples



What is The BaseSpace?

BaseSpace is a website in the ‘**Cloud**’

- **BaseSpace** is hosted in **Amazon** Web Services
- Anyone can use **BaseSpace** – just go to basespace.illumina.com
    - Available for **Illumina** customers
- **BaseSpace** currently supports

1TB free storage for each account

The Benefit for BaseSpace

Online Monitor

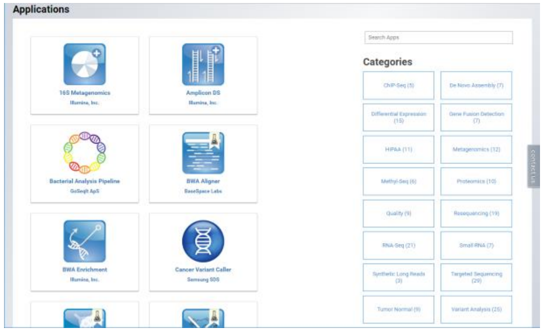
- ✓ Check your sequencing run at anytime with Internet

Backup

- ✓ Automatic Online Backup

Analysis App

- ✓ Very easy to use for anybody



Applications

16S Metagenomics
Illumina, Inc.

Amplicon DS
Illumina, Inc.

Bacterial Analysis Pipeline
GoSeqIt ApS

BWA Aligner
BaseSpace Labs

BWA Enrichment
Illumina, Inc.

Cancer Variant Caller
Samsung SDS

Search Apps

Categories

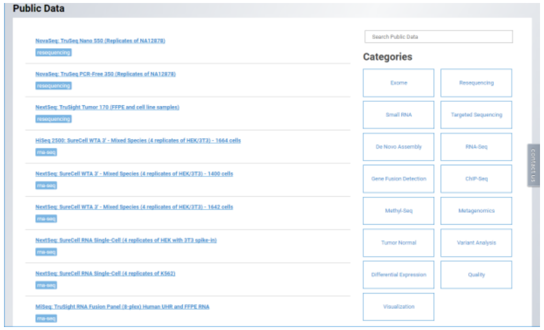
ChIP-Seq (5)
De Novo Assembly (7)
Differential Expression (15)
Gene Fusion Detection (7)
HIPAA (11)
Metagenomics (12)
Methyl-Seq (6)
Proteomics (10)
Quality (9)
Resequencing (19)
RNA-Seq (21)
Small RNA (7)
Synthetic Long Reads (3)
Targeted Sequencing (29)
Tumor Normal (9)
Variant Analysis (25)

Data Sharing

- ✓ No HDD or USB storage, just share data by account

Demo Data

- ✓ The Illumina data updated



Public Data

NovaSeq: TruSeq Nano 550 (Replicates of NA12878)
resequencing

NovaSeq: TruSeq PCR-Free 350 (Replicates of NA12878)
resequencing

NextSeq: TruSight Tumor 170 (FFPE and cell line samples)
resequencing

HiSeq 2500: SureCell WTA 3' - Mixed Species (4 replicates of HEK/3T3) - 1664 cells
rna-seq

NextSeq: SureCell WTA 3' - Mixed Species (4 replicates of HEK/3T3) - 1400 cells
rna-seq

NextSeq: SureCell WTA 3' - Mixed Species (4 replicates of HEK/3T3) - 1642 cells
rna-seq

NextSeq: SureCell RNA Single-Cell (4 replicates of HEK with 3T3 spike-in)
rna-seq

NextSeq: SureCell RNA Single-Cell (4 replicates of K562)
rna-seq

MiSeq: TruSight RNA Fusion Panel (8-plex) Human UHR and FFPE RNA
rna-seq

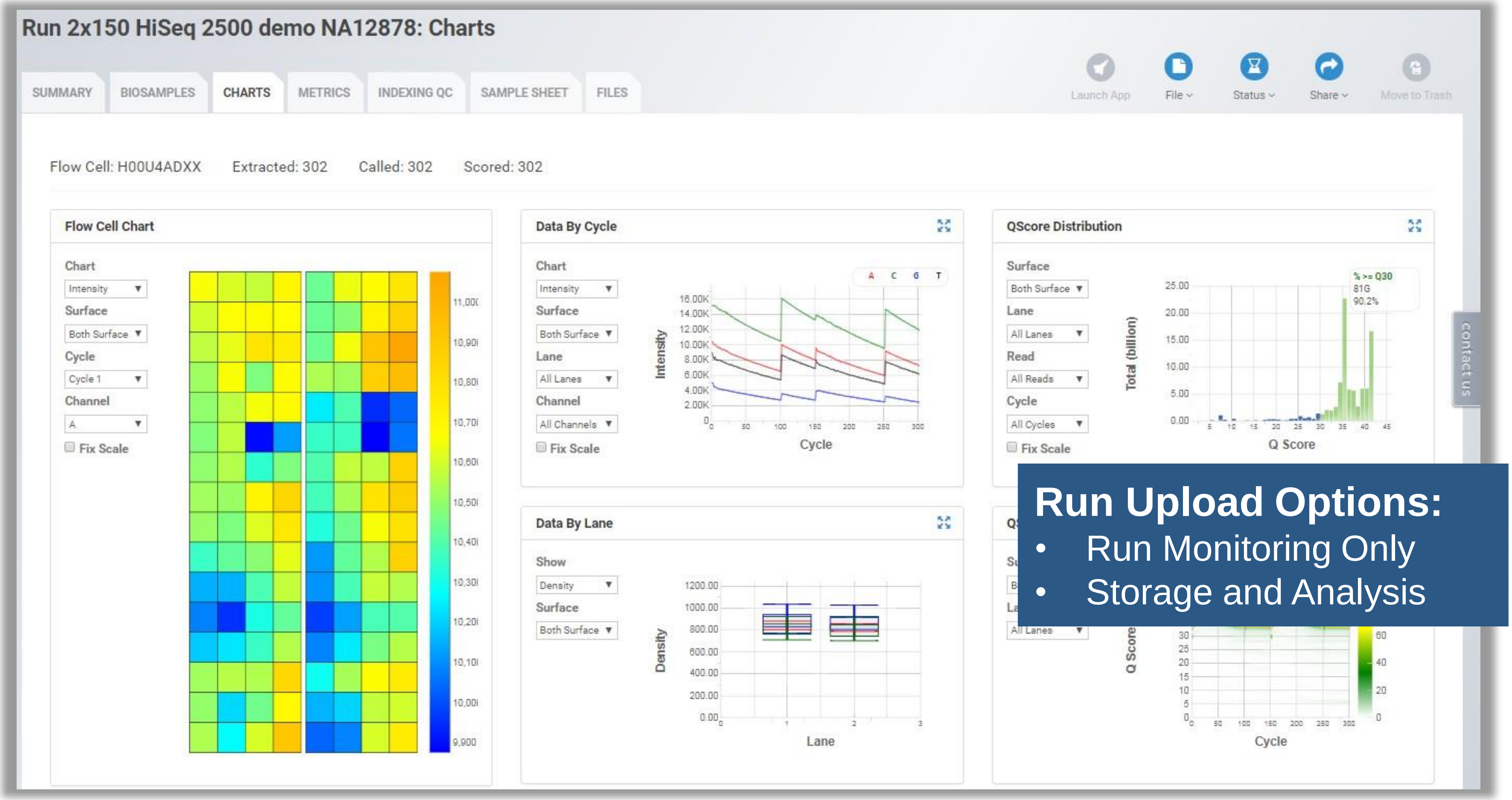
Search Public Data

Categories

Exome
Resequencing
Small RNA
Targeted Sequencing
De Novo Assembly
RNA-Seq
Gene Fusion Detection
ChIP-Seq
Methyl-Seq
Metagenomics
Tumor Normal
Variant Analysis
Differential Expression
Quality
Visualization

Remote Instrument Run Monitoring

Requires Web Connection



Run 2x150 HiSeq 2500 demo NA12878: Charts

Flow Cell: H00U4ADXX Extracted: 302 Called: 302 Scored: 302

Flow Cell Chart

Chart: Intensity
Surface: Both Surface
Cycle: Cycle 1
Channel: A
☐ Fix Scale

Data By Cycle

Chart: Intensity
Surface: Both Surface
Lane: All Lanes
Channel: All Channels
☐ Fix Scale

QScore Distribution

Surface: Both Surface
Lane: All Lanes
Read: All Reads
Cycle: All Cycles
☐ Fix Scale

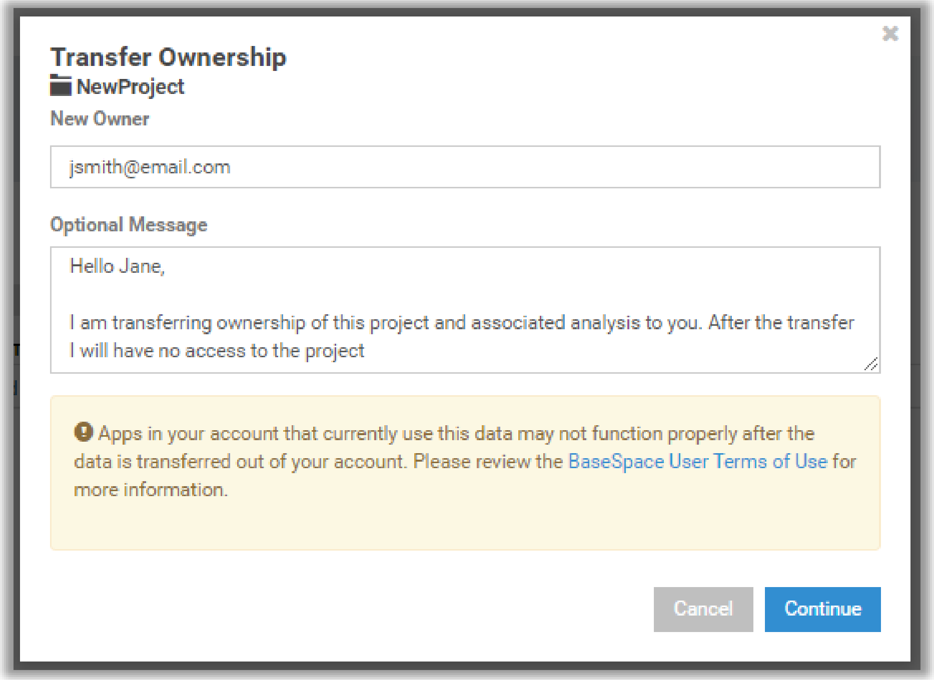
Data By Lane

Show: Density
Surface: Both Surface

Run Upload Options:

- Run Monitoring Only
- Storage and Analysis

Sharing and Collaboration



Transfer Ownership

NewProject

New Owner

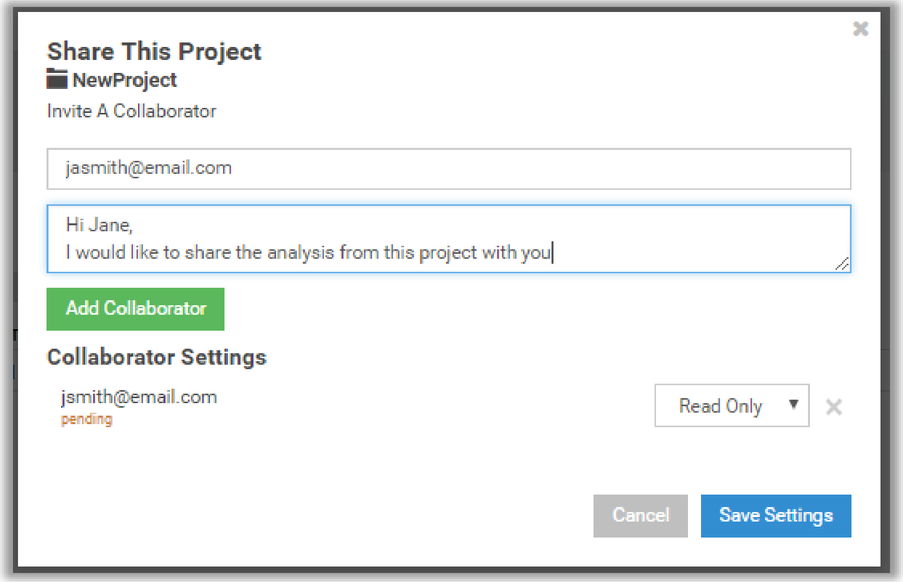
jsmith@email.com

Optional Message

Hello Jane,

I am transferring ownership of this project and associated analysis to you. After the transfer I will have no access to the project

Apps in your account that currently use this data may not function properly after the data is transferred out of your account. Please review the BaseSpace User Terms of Use for more information.



Share This Project

NewProject

Invite A Collaborator

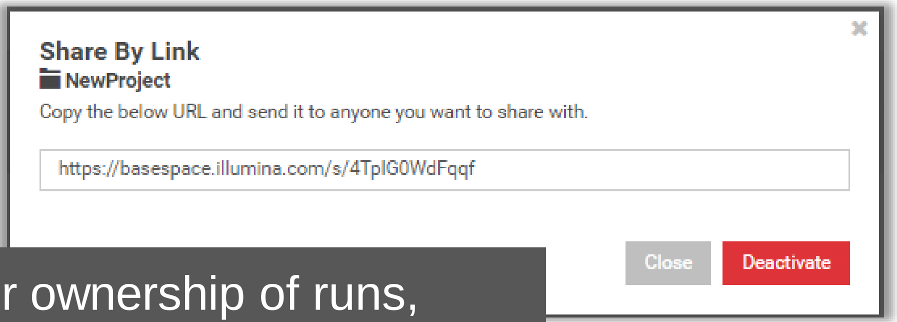
jasmith@email.com

Hi Jane,
I would like to share the analysis from this project with you

Collaborator Settings

jsmith@email.com
pending

Read Only



Share By Link

NewProject

Copy the below URL and send it to anyone you want to share with.

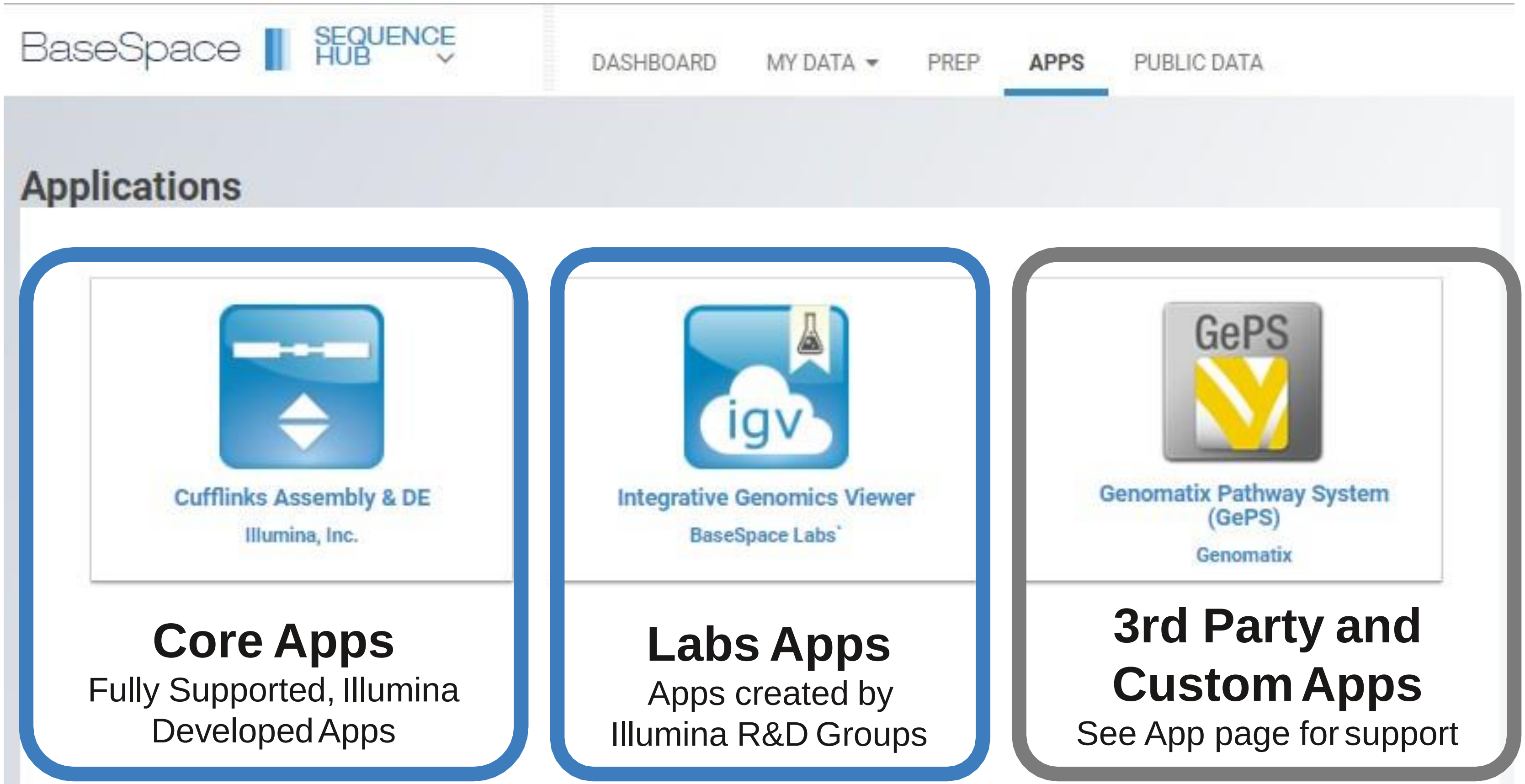
https://basespace.illumina.com/s/4TplG0WdFqqf

Easily share or transfer ownership of runs, samples, or projects with anyone in the world*

*Sequence Hub account with available storage required to accept transfer

Push-Button Analysis

Three types of pre-built Apps to analyze your data



Core Apps
Fully Supported, Illumina Developed Apps

Labs Apps
Apps created by Illumina R&D Groups

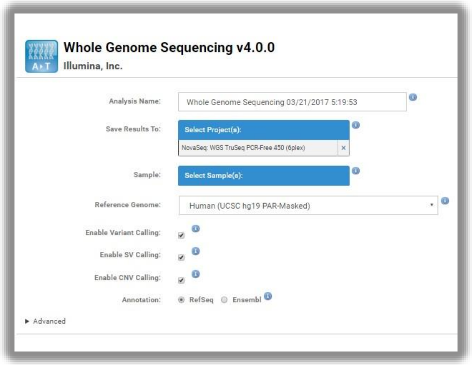
3rd Party and Custom Apps
See App page for support

Over 80 published Apps supporting all of Illumina library prep kits

3

Push-Button Analysis

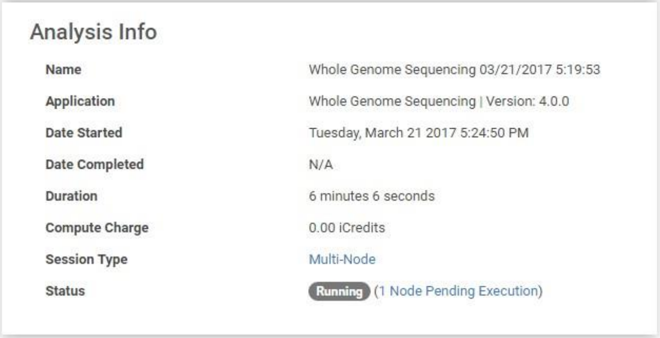
Common Workflow for All Illumina Core Apps



Whole Genome Sequencing v4.0.0
Illumina, Inc.

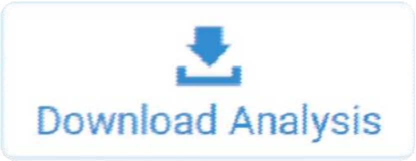
Analysis Name:	Whole Genome Sequencing 03/21/2017 5:19:53
Save Results To:	Select Project(s): NovaSeq: WGS TruSeq PCR-Free 450 (6plex)
Sample:	Select Sample(s):
Reference Genome:	Human (UCSC hg19 PAR-Masked)
Enable Variant Calling:	☑
Enable SV Calling:	☑
Enable CNV Calling:	☑
Annotation:	● RefSeq ○ Ensembl

▸ Advanced



Analysis Info

Name	Whole Genome Sequencing 03/21/2017 5:19:53
Application	Whole Genome Sequencing | Version: 4.0.0
Date Started	Tuesday, March 21 2017 5:24:50 PM
Date Completed	N/A
Duration	6 minutes 6 seconds
Compute Charge	0.00 iCredits
Session Type	Multi-Node
Status	Running (1 Node Pending Execution)

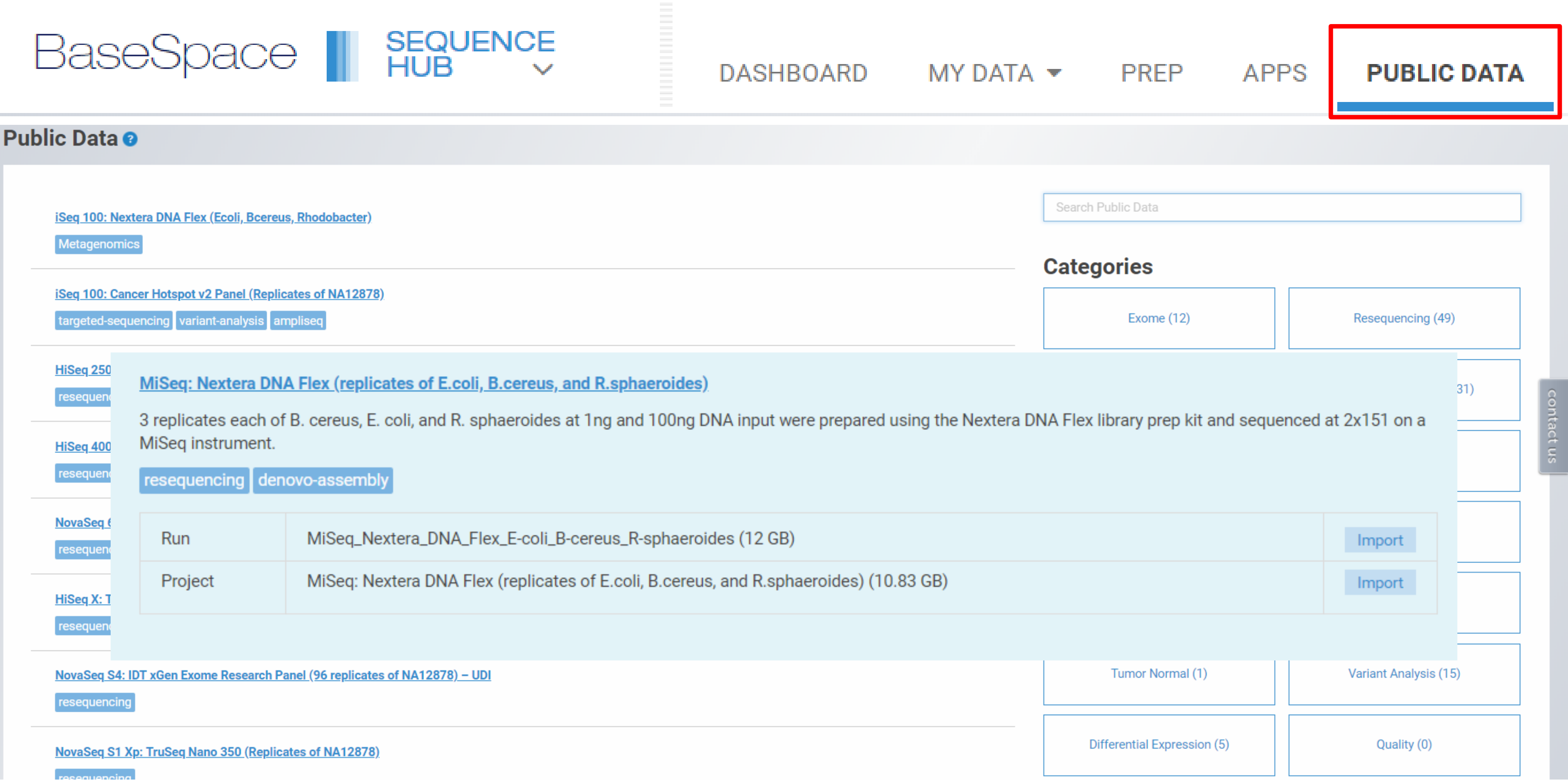


- Sequencing Summary Report
- BAM Alignment Map
- *.bai (Index for Alignment Map
- VCF Variant Call Files
- Log and Error Files
- Summary Reports

BaseSpace performs alignment, calls and annotates variants, summarizes results

illumina®

BaseSpace Public Data



Public Data

iSeq 100: Nextera DNA Flex (Ecoli, Bcereus, Rhodobacter)

Metagenomics

iSeq 100: Cancer Hotspot v2 Panel (Replicates of NA12878)

targeted-sequencing variant-analysis ampliseq

HiSeq 250

resequen

HiSeq 400

resequen

NovaSeq 6

resequen

HiSeq X: T

resequen

NovaSeq S4: IDT xGen Exome Research Panel (96 replicates of NA12878) – UDI

resequencing

NovaSeq S1 Xp: TruSeq Nano 350 (Replicates of NA12878)

resequencing

MiSeq: Nextera DNA Flex (replicates of E.coli, B.cereus, and R.sphaeroides)

3 replicates each of B. cereus, E. coli, and R. sphaeroides at 1ng and 100ng DNA input were prepared using the Nextera DNA Flex library prep kit and sequenced at 2x151 on a MiSeq instrument.

resequencing denovo-assembly

Run	MiSeq_Nextera_DNA_Flex_E-coli_B-cereus_R-sphaeroides (12 GB)	Import
Project	MiSeq: Nextera DNA Flex (replicates of E.coli, B.cereus, and R.sphaeroides) (10.83 GB)	Import

Search Public Data

Categories

Exome (12)

Resequencing (49)

31)

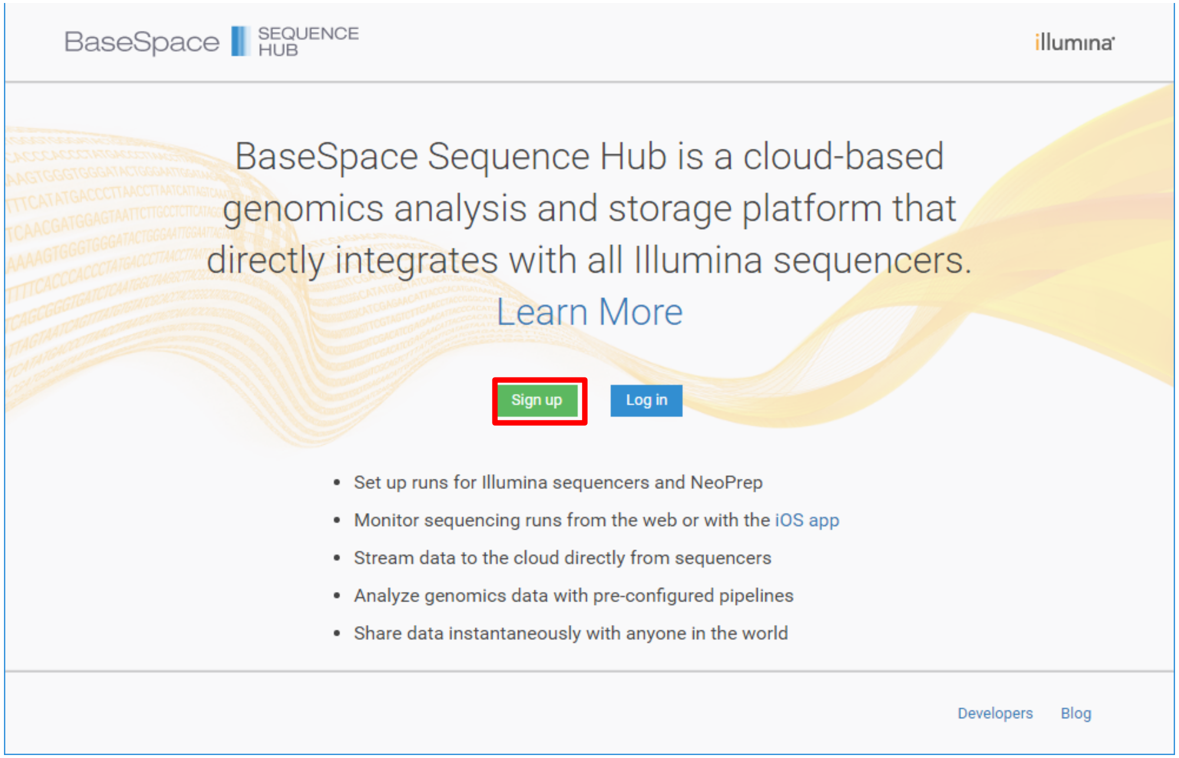
Tumor Normal (1)

Variant Analysis (15)

Differential Expression (5)

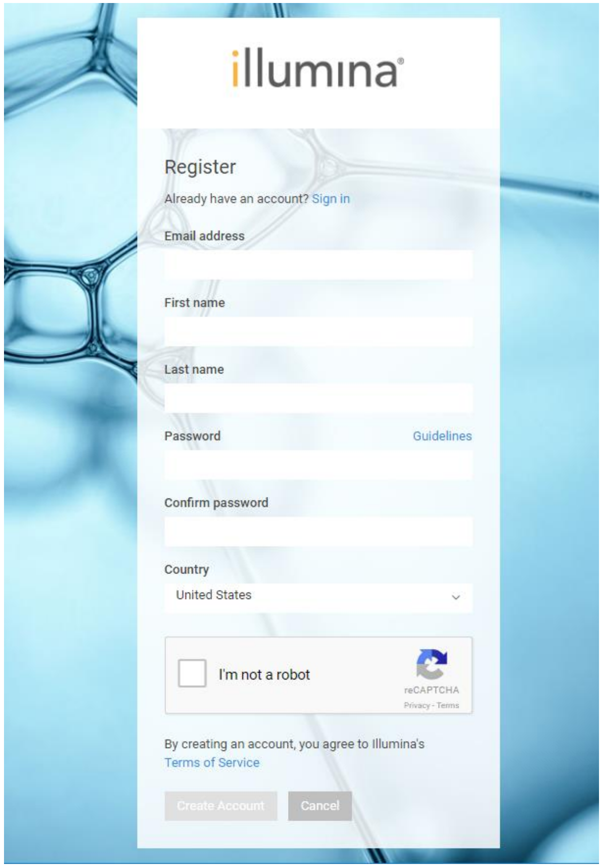
Quality (0)

Register a BaseSpace account



BaseSpace Sequence Hub is a cloud-based genomics analysis and storage platform that directly integrates with all Illumina sequencers. Learn More

- Set up runs for Illumina sequencers and NeoPrep
- Monitor sequencing runs from the web or with the iOS app
- Stream data to the cloud directly from sequencers
- Analyze genomics data with pre-configured pipelines
- Share data instantaneously with anyone in the world



Register

Already have an account? Sign in

Email address

First name

Last name

Password Guidelines

Confirm password

Country

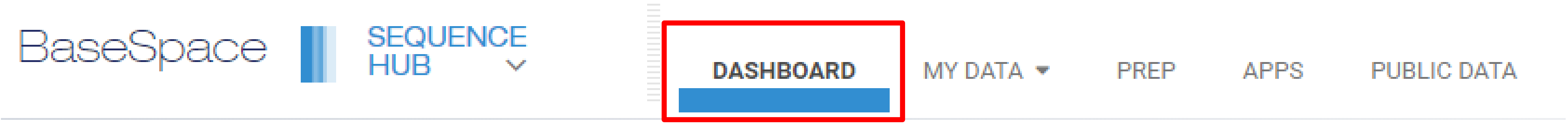
United States

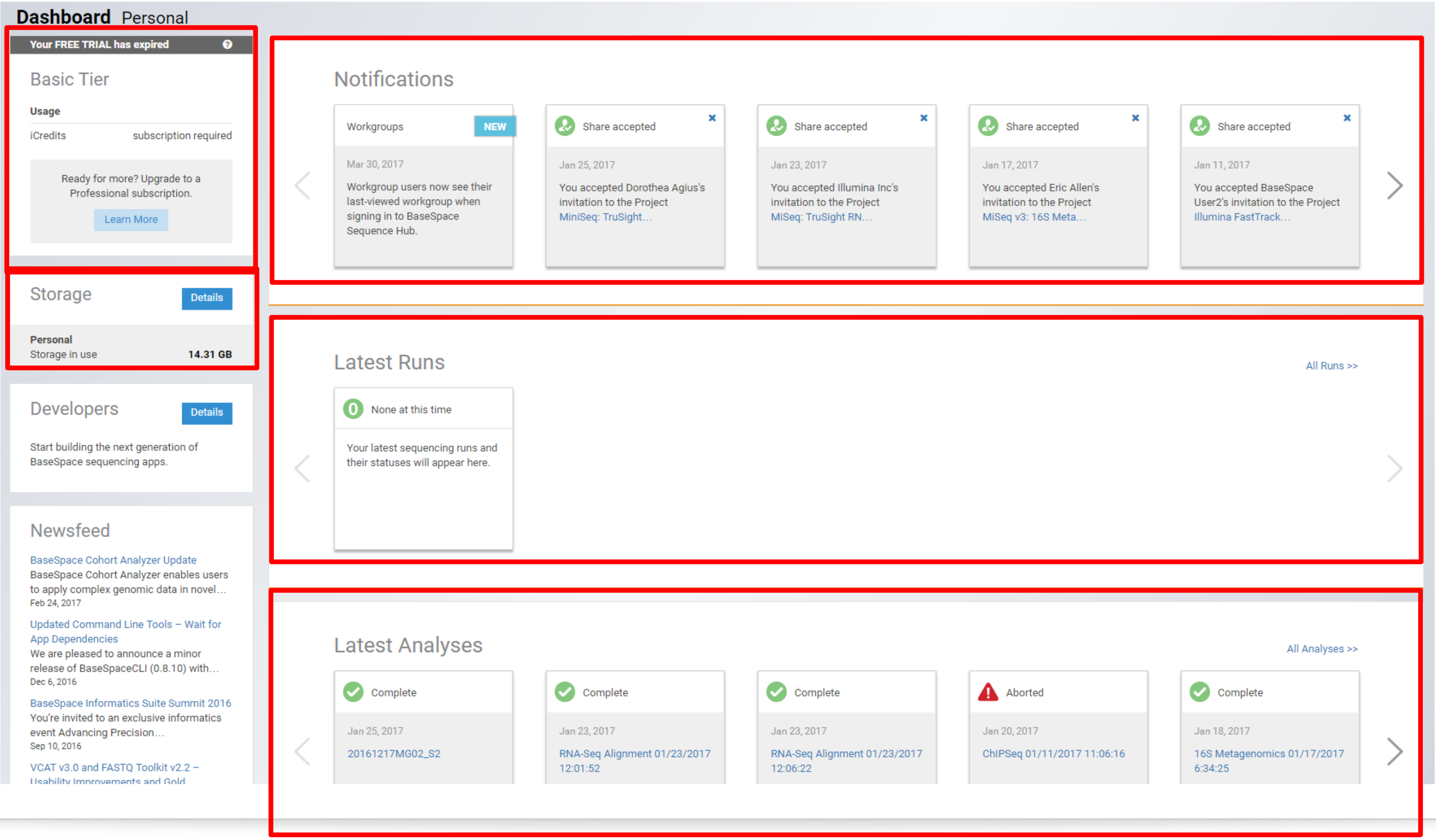
☐ I'm not a robot

By creating an account, you agree to Illumina's Terms of Service

basespace.illumina.com

BaseSpace Dashboard




Dashboard Personal

Your FREE TRIAL has expired

Basic Tier

Usage

iCredits	subscription required

Ready for more? Upgrade to a Professional subscription.

Learn More

Storage

Details

Personal

Storage in use	**14.31 GB**

Developers

Details

Start building the next generation of BaseSpace sequencing apps.

Newsfeed

BaseSpace Cohort Analyzer Update
BaseSpace Cohort Analyzer enables users to apply complex genomic data in novel…
Feb 24, 2017

Updated Command Line Tools – Wait for App Dependencies
We are pleased to announce a minor release of BaseSpaceCLI (0.8.10) with…
Dec 6, 2016

BaseSpace Informatics Suite Summit 2016
You're invited to an exclusive informatics event Advancing Precision…
Sep 10, 2016

VCAT v3.0 and FASTQ Toolkit v2.2 – Usability Improvements and Gold

Notifications

Workgroups NEW
Mar 30, 2017
Workgroup users now see their last-viewed workgroup when signing in to BaseSpace Sequence Hub.

Share accepted
Jan 25, 2017
You accepted Dorothea Agius's invitation to the Project MiniSeq: TruSight…

Share accepted
Jan 23, 2017
You accepted Illumina Inc's invitation to the Project MiSeq: TruSight RN…

Share accepted
Jan 17, 2017
You accepted Eric Allen's invitation to the Project MiSeq v3: 16S Meta…

Share accepted
Jan 11, 2017
You accepted BaseSpace User2's invitation to the Project Illumina FastTrack…

Latest Runs

All Runs >>

0 None at this time
Your latest sequencing runs and their statuses will appear here.

Latest Analyses

All Analyses >>

Complete
Jan 25, 2017
20161217MG02_S2

Complete
Jan 23, 2017
RNA-Seq Alignment 01/23/2017 12:01:52

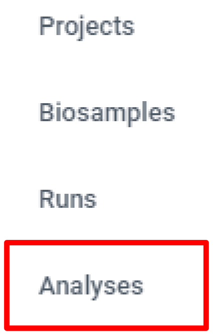
Complete
Jan 23, 2017
RNA-Seq Alignment 01/23/2017 12:06:22

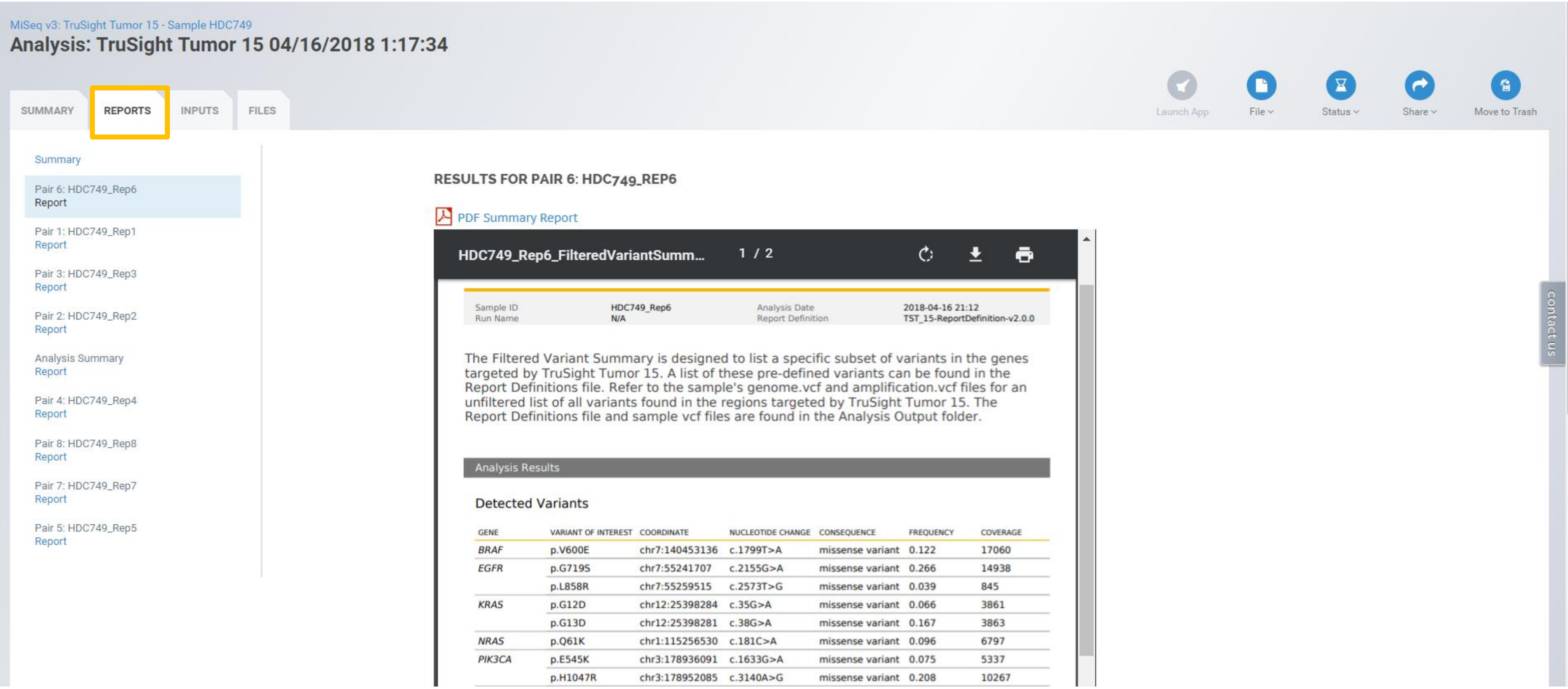
Aborted
Jan 20, 2017
ChIPSeq 01/11/2017 11:06:16

Complete
Jan 18, 2017
16S Metagenomics 01/17/2017 6:34:25

Report

Analysis Summary and result of each sample





Analysis: TruSight Tumor 15 04/16/2018 1:17:34

RESULTS FOR PAIR 6: HDC749_REP6

Sample ID	HDC749_Rep6	Analysis Date	2018-04-16 21:12
Run Name	N/A	Report Definition	TST_15-ReportDefinition-v2.0.0

The Filtered Variant Summary is designed to list a specific subset of variants in the genes targeted by TruSight Tumor 15. A list of these pre-defined variants can be found in the Report Definitions file. Refer to the sample's genome.vcf and amplification.vcf files for an unfiltered list of all variants found in the regions targeted by TruSight Tumor 15. The Report Definitions file and sample vcf files are found in the Analysis Output folder.

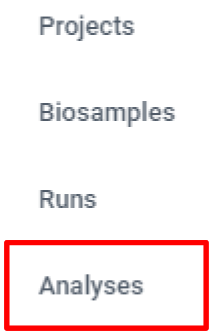
Analysis Results

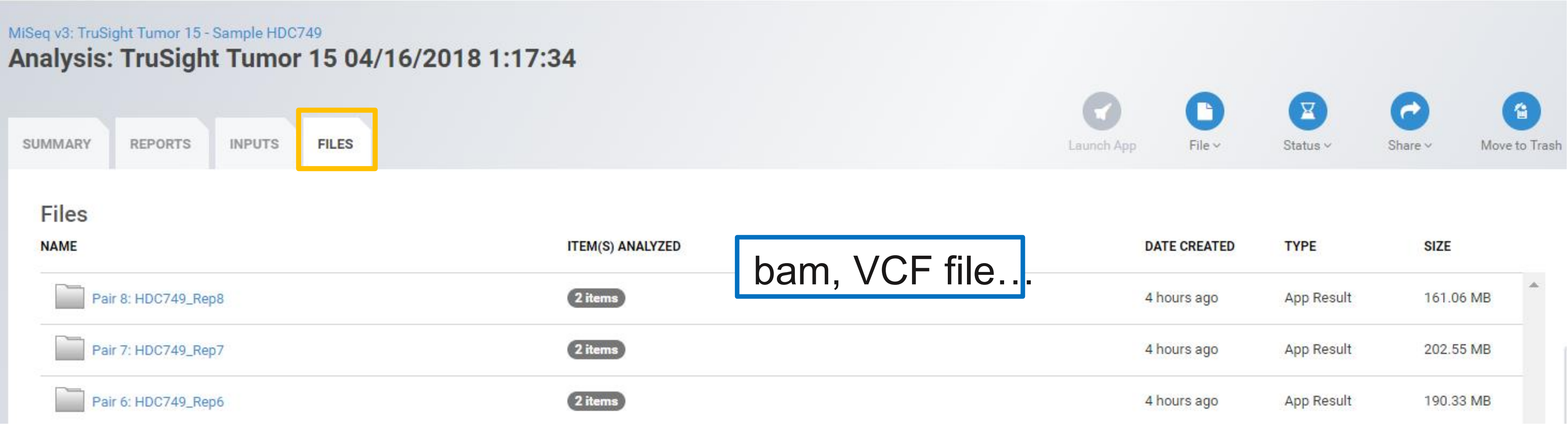
Detected Variants

GENE	VARIANT OF INTEREST	COORDINATE	NUCLEOTIDE CHANGE	CONSEQUENCE	FREQUENCY	COVERAGE
BRAF	p.V600E	chr7:140453136	c.1799T>A	missense variant	0.122	17060
EGFR	p.G719S	chr7:55241707	c.2155G>A	missense variant	0.266	14938
	p.L858R	chr7:55259515	c.2573T>G	missense variant	0.039	845
KRAS	p.G12D	chr12:25398284	c.35G>A	missense variant	0.066	3861
	p.G13D	chr12:25398281	c.38G>A	missense variant	0.167	3863
NRAS	p.Q61K	chr1:115256530	c.181C>A	missense variant	0.096	6797
PIK3CA	p.E545K	chr3:178936091	c.1633G>A	missense variant	0.075	5337
	p.H1047R	chr3:178952085	c.3140A>G	missense variant	0.208	10267

Output Files

Results organized into sample folders




MiSeq v3: TruSight Tumor 15 - Sample HDC749

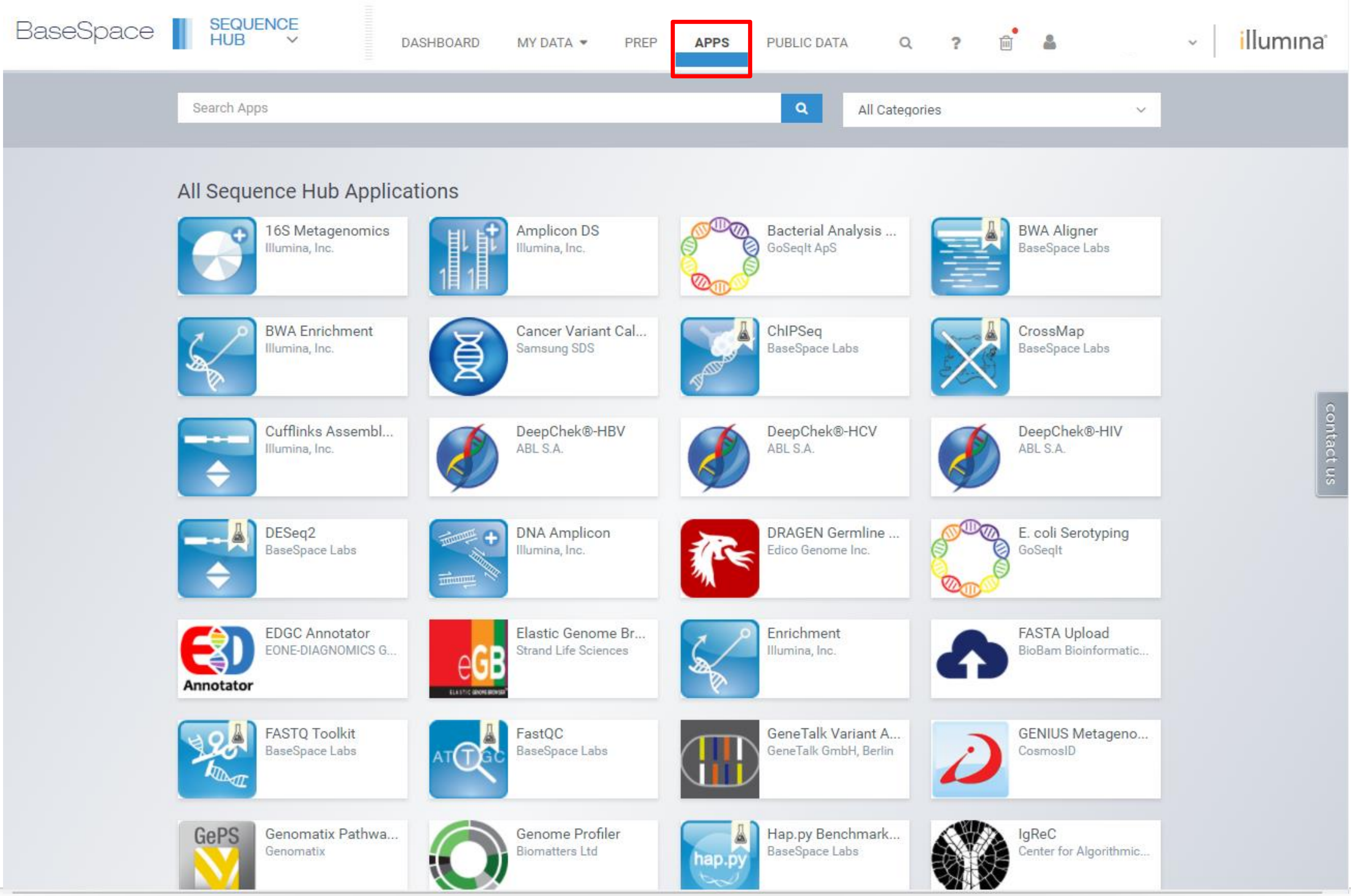
Analysis: TruSight Tumor 15 04/16/2018 1:17:34

Files

NAME	ITEM(S) ANALYZED	DATE CREATED	TYPE	SIZE
Pair 8: HDC749_Rep8	2 items	4 hours ago	App Result	161.06 MB
Pair 7: HDC749_Rep7	2 items	4 hours ago	App Result	202.55 MB
Pair 6: HDC749_Rep6	2 items	4 hours ago	App Result	190.33 MB

bam, VCF file…

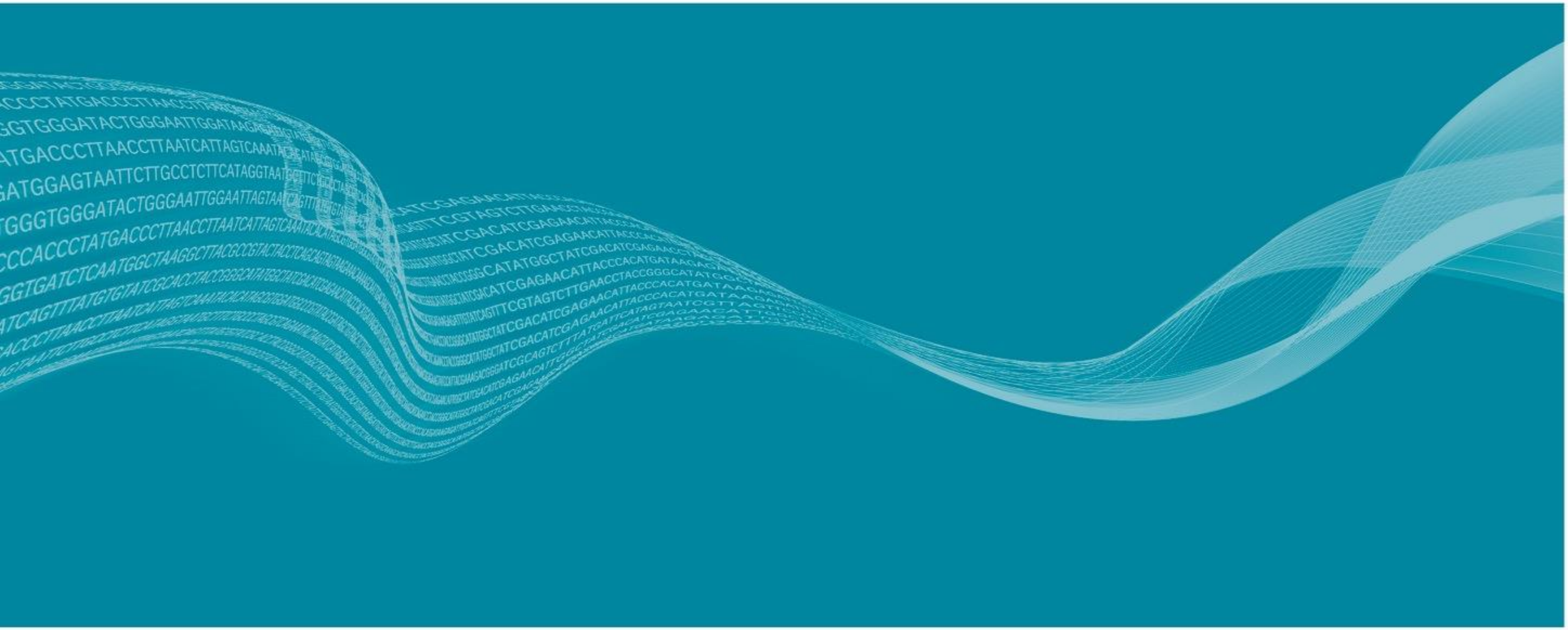
BaseSpace App



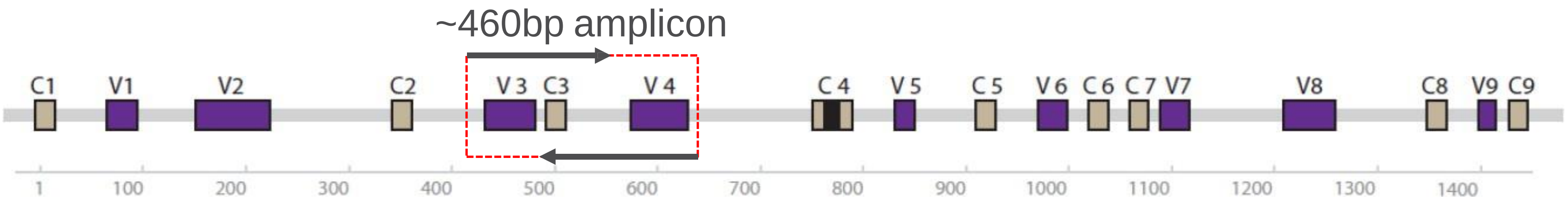
All Sequence Hub Applications

16S Metagenomics Illumina, Inc.	Amplicon DS Illumina, Inc.	Bacterial Analysis ... GoSeqIt ApS	BWA Aligner BaseSpace Labs
BWA Enrichment Illumina, Inc.	Cancer Variant Cal... Samsung SDS	ChIPSeq BaseSpace Labs	CrossMap BaseSpace Labs
Cufflinks Assembl... Illumina, Inc.	DeepChek®-HBV ABL S.A.	DeepChek®-HCV ABL S.A.	DeepChek®-HIV ABL S.A.
DESeq2 BaseSpace Labs	DNA Amplicon Illumina, Inc.	DRAGEN Germline ... Edico Genome Inc.	E. coli Serotyping GoSeqIt
EDGC Annotator EONE-DIAGNOMICS G...	Elastic Genome Br... Strand Life Sciences	Enrichment Illumina, Inc.	FASTA Upload BioBam Bioinformatic...
FASTQ Toolkit BaseSpace Labs	FastQC BaseSpace Labs	GeneTalk Variant A... GeneTalk GmbH, Berlin	GENIUS Metageno... CosmosID
Genomatix Pathwa... Genomatix	Genome Profiler Biomatters Ltd	Hap.py Benchmark... BaseSpace Labs	IgReC Center for Algorithmic...

16s Metagenomics Analysis



16S rRNA sequencing with Illumina protocols & software tools

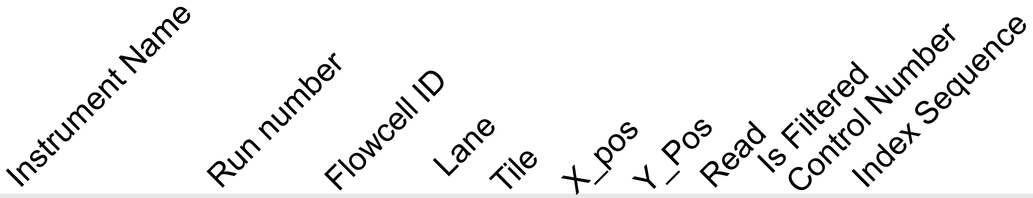


BaseSpace

16S Metagenomics App

- GreenGenes database to perform taxonomic classification
  - Illumina-curated version
- Bayesian classification method to assign taxonomies
  - Short sub-sequences (32 bases) are extracted from each read and compared to the database by the classifier
  - Classification down to genus/species-level

Demultiplexing **RAW** Files, FASTQ File



```
@HWI-BRUNOP20X:994:B809UWABXX:1:1101:13501:2240 1:N:0:CTTGTA
TGAAACCAGTGTTCTTAATTGGCATTTTACACACACACACACAGAATTTAAAAAAAAAATCAAAGGAAATCATTCTAAATGTACTATGATAGCATGTTAAA
+
=55>7;?::BDADDD@EE88DCD?DFFEFFECBE6666BB=B;<;<-34:;<CB51>=BBEE>EE?3D@??CB->:=:AA8DDDDDDBBE9;,=?:/89<E
```

Q-score = -10 log(error rate)

Phred Score
\+ 33

ASCII value

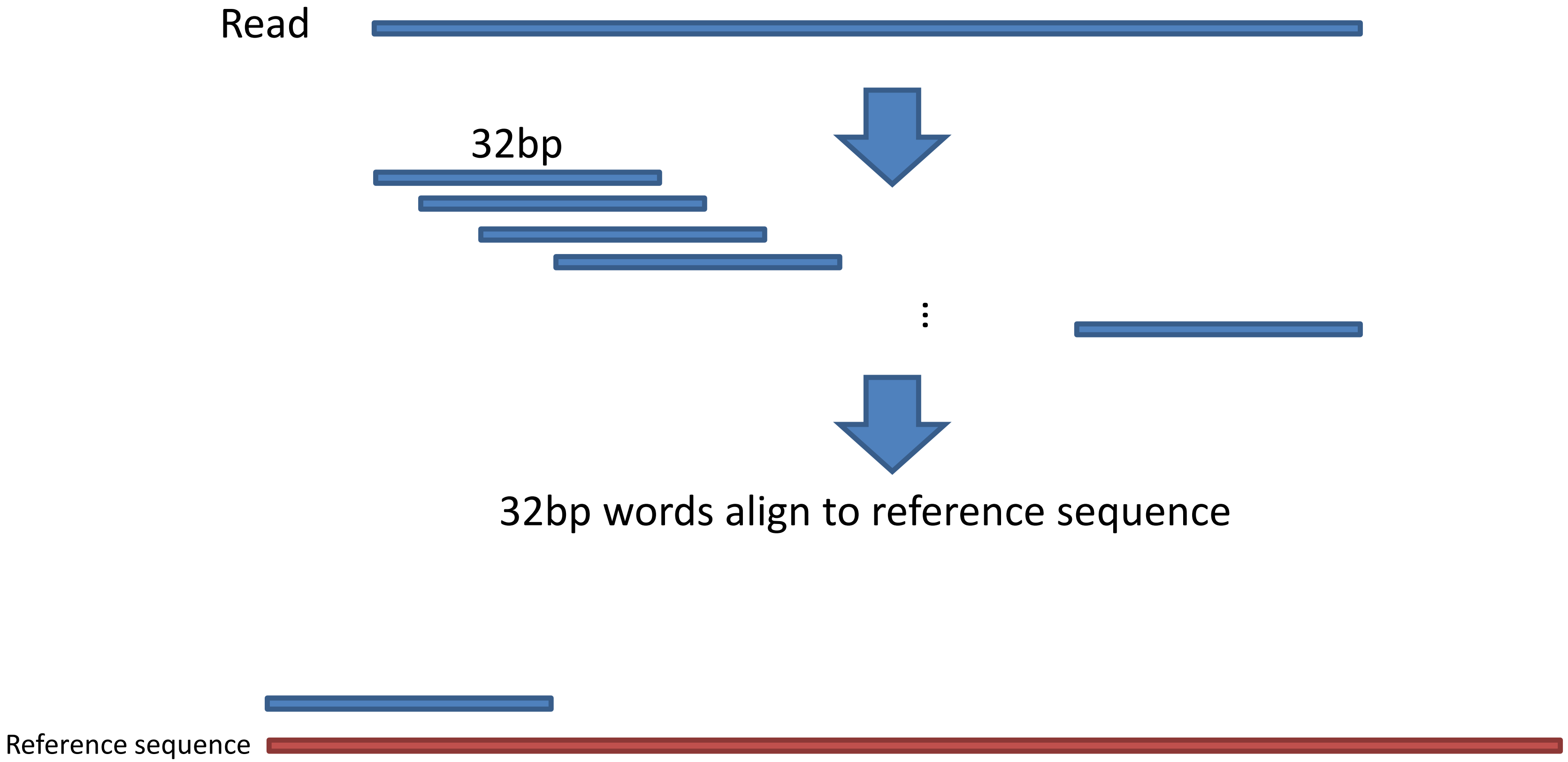
Q-score=Phred Score

Character	ASCII Value	Phred Score	Error probability
5	53	20	0.01
?	**63**	**30**	**0.001**
I	73	40	0.0001

Examples of ASCII in Sanger FastQ

Read Classification of 16s Metagenomics app

- The original RDP classifier algorithm used 8-base words due to implementation constraints. ClassifyReads uses more efficient data structures and is able to use 32-base words – giving each word more specificity for each species.


Read

32bp

32bp words align to reference sequence

Reference sequence

Appendix – Filters applied to GG by Illumina

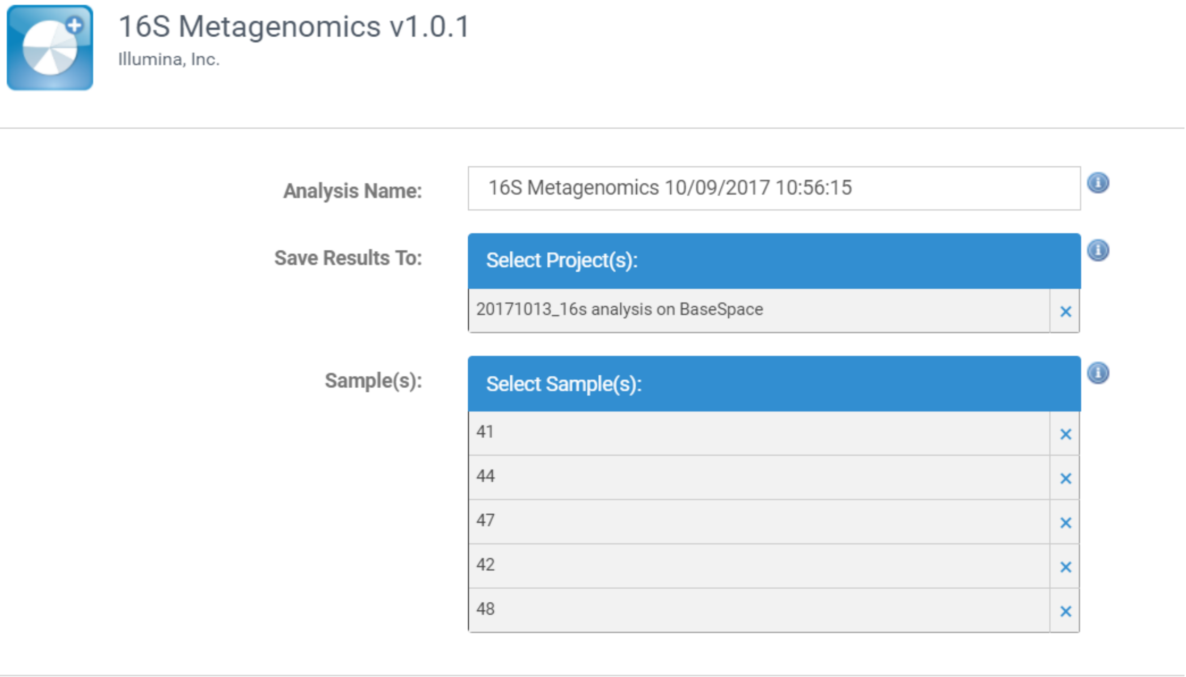
1. Filter all entries where the 16S sequence length was below 1250 bp
2. Filter all entries that had more than 50 wobble bases

(i.e. M, R, W, S, Y, K, V, H, D, B, N)

3. Filter all entries that were only partially classified

(no classification for genus or species)

16s Metgenomics app



16S Metagenomics v1.0.1

Illumina, Inc.

Analysis Name: 16S Metagenomics 10/09/2017 10:56:15

Save Results To: Select Project(s):

20171013_16s analysis on BaseSpace

Sample(s): Select Sample(s):

41

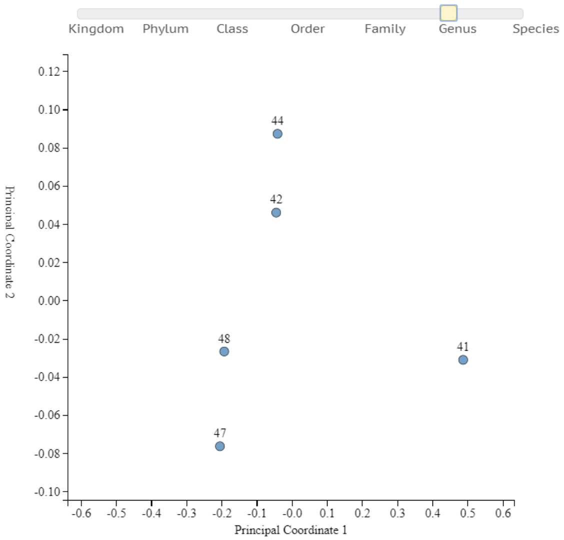
44

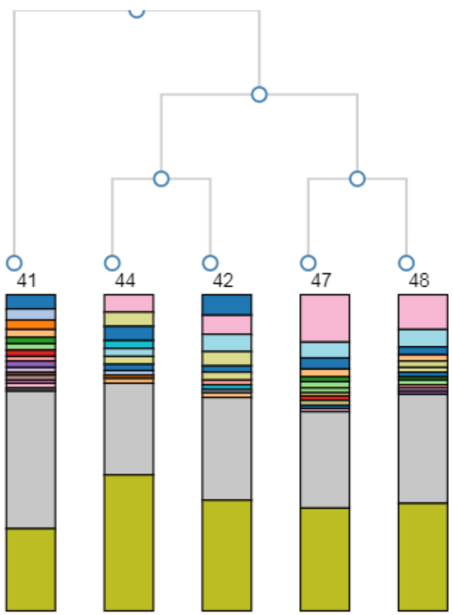
47

42

48

AGGREGATE RESULTS





AGGREGATE RESULTS

- PDF Aggregate Summary Report
- Kingdom-Level Aggregate Results
- Phylum-Level Aggregate Results
- Class-Level Aggregate Results
- Order-Level Aggregate Results
- Family-Level Aggregate Results
- Genus-Level Aggregate Results
- Species-Level Aggregate Results

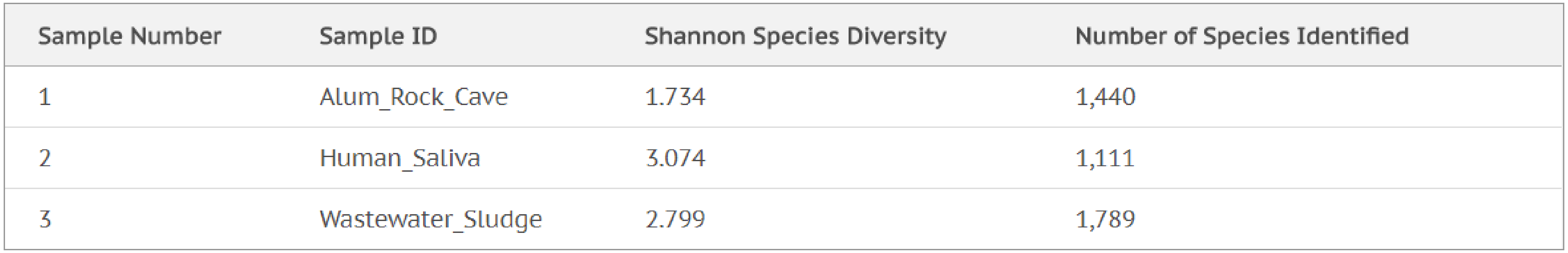
SAMPLE INFORMATION

Sample Number	Sample ID	Number Reads PF	% Reads PF Classified to Genus
1	41	41,862	74.16%
2	44	50,075	57.13%
3	47	110,434	67.41%
4	42	100,803	64.93%
5	48	67,342	65.76%

BaseSpace

16S Metagenomics App

- Shannon species diversity measures the entropy of Species-level classifications in the sample.



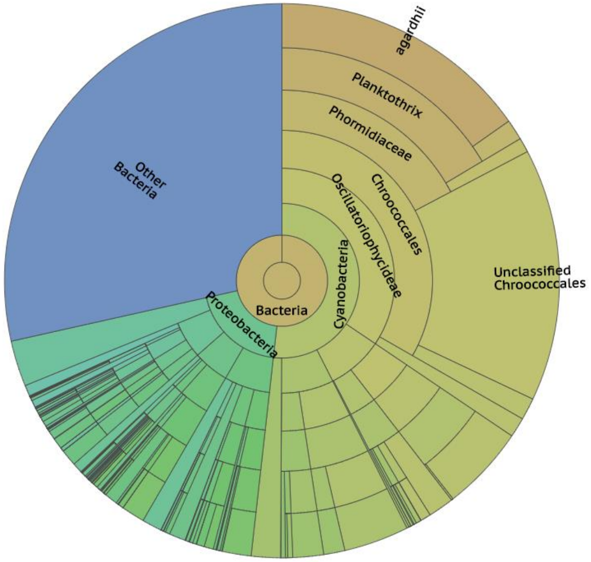
Sample Number	Sample ID	Shannon Species Diversity	Number of Species Identified
1	Alum_Rock_Cave	1.734	1,440
2	Human_Saliva	3.074	1,111
3	Wastewater_Sludge	2.799	1,789

http://en.wikipedia.org/wiki/Shannon-Wiener_index

RESULTS FOR SAMPLE

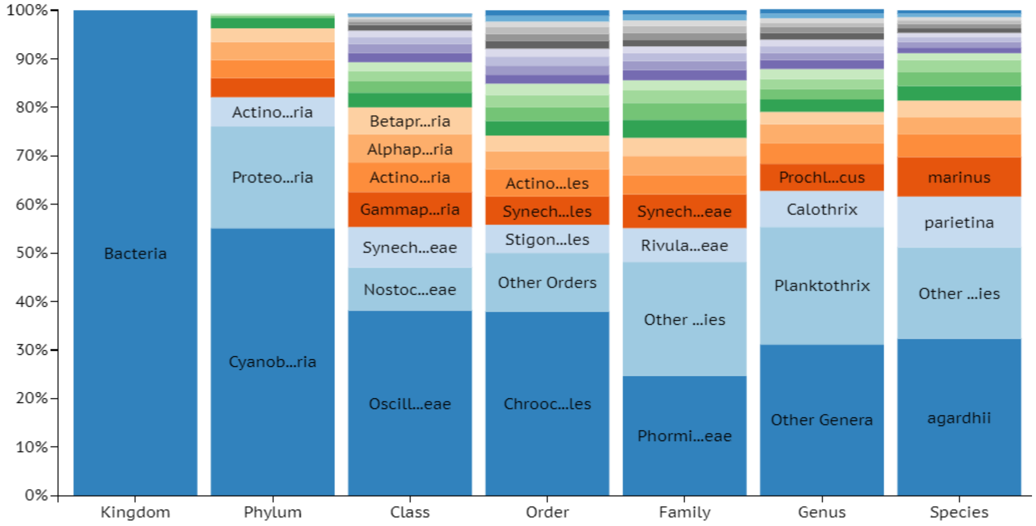
CLASSIFICATION STATISTICS

Taxonomic Level	Reads PF Classified to Taxonomic Level	% Reads PF Classified to Taxonomic Level
Kingdom	109,372	99.04%
Phylum	102,850	93.13%
Class	98,852	89.51%
Order	96,184	87.10%
Family	78,502	71.08%
Genus	74,448	67.41%
Species	51,579	46.71%



TOP 20 CLASSIFICATION RESULTS BY TAXONOMIC LEVEL

This column chart shows the relative abundance of the top 20 classification results within each taxonomic level. Mouse over any category to see its description and abundance.



Output Files

- Sample.report.pdf
  - Sample Information and Classification Information
- Sample.summary.csv
  - summarized counts of how reads were classified for each level in the sample
- Sample.txt.gz
  - The classification assigned to each read
- Class-level-Aggregate_count.csv
  - Per-level aggregate counts for all samples, each row represents a unique classification that occurred in one or more samples.
- MetagenomicsAggregateReport.pdf
  - Sample Classified Information and Species Diversity Results

BaseSpace

QIIME App

QIIME Preprocessing
QIIME Development Team

- **OTU picking, taxonomic assignment**
- **sl-out** folder: contain the summary of read quality file – split_library_log.txt
- **closed-ref** folder: contain the OTU table – otu_table.biom
- and tree file – 97_otus.tree
- If your reads will not be well represented by the Greengenes database, that might be underrepresented

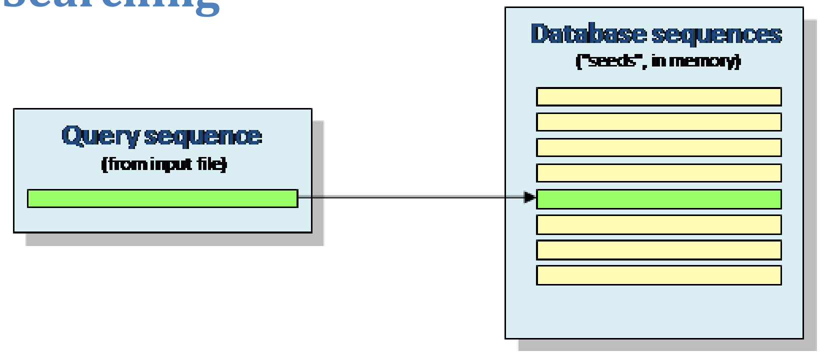
QIIME Visualizations
QIIME Development Team

- **diversity analyses and visualizations**
- Beta diversity plots
- Alpha rarefaction plots
- Taxonomic summaries

Read Classification of QIIME app

uclust_ref (Edgar, RC 2010), as uclust, but takes a reference database to use as seeds.
Taxonomy database: Greengeen v13.8

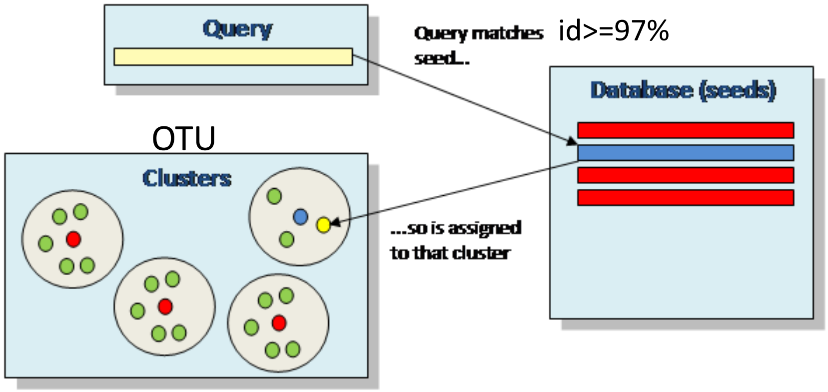
Searching



- ✓ searching a database stored in memory
- ✓ A query sequence matches a database sequence if the identity is high enough
- ✓ The minimum identity is set by the --id option, e.g. --id 0.97

ATCGGAC 5 matches, 1 mismatch (default allow 1 mismatch)
A–CAGAC **Identity= matches(5)/shorter length(6)=83.3%**

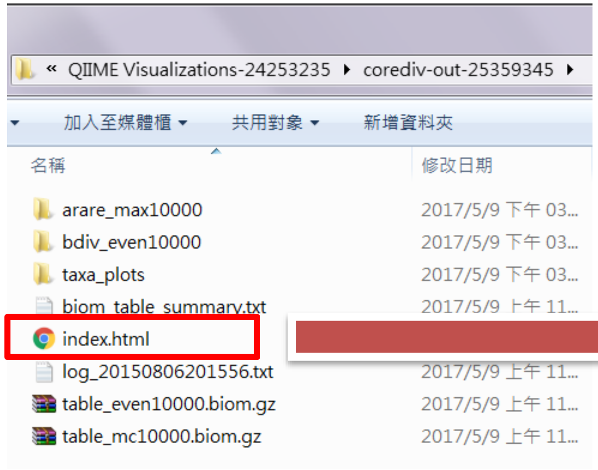
Clustering



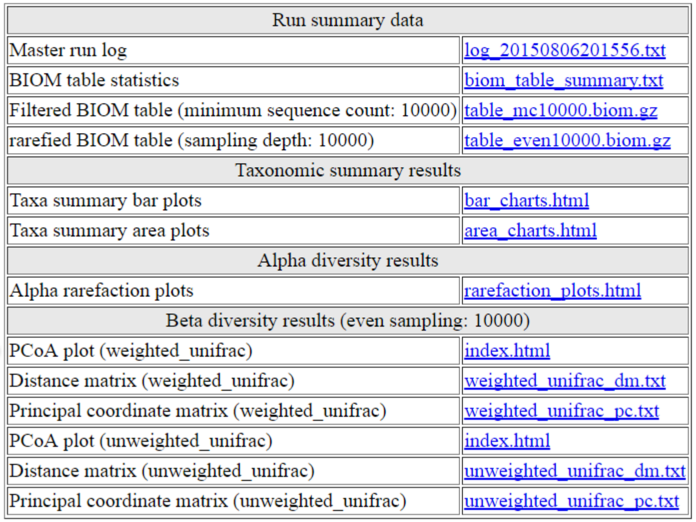
- ✓ Query sequences are processed in input order. If a match is found to a database sequence, then the query is assigned to that cluster

http://www.drive5.com/uclust/uclust_userguide_1_1_579.html#_Toc257997673

Output data form QIIME Visualizations



名稱	修改日期
arare_max10000	2017/5/9 下午 03…
bdiv_even10000	2017/5/9 下午 03…
taxa_plots	2017/5/9 下午 03…
biom_table_summary.txt	2017/5/9 上午 11…
index.html	
log_20150806201556.txt	2017/5/9 上午 11…
table_even10000.biom.gz	2017/5/9 上午 11…
table_mc10000.biom.gz	2017/5/9 上午 11…

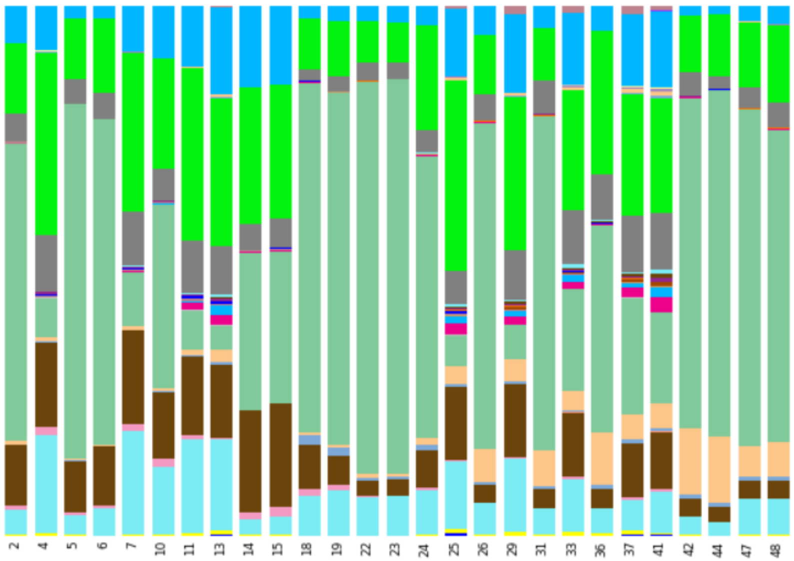


Run summary data	
Master run log	log_20150806201556.txt
BIOM table statistics	biom_table_summary.txt
Filtered BIOM table (minimum sequence count: 10000)	table_mc10000.biom.gz
rarefied BIOM table (sampling depth: 10000)	table_even10000.biom.gz
Taxonomic summary results	
Taxa summary bar plots	bar_charts.html
Taxa summary area plots	area_charts.html
Alpha diversity results	
Alpha rarefaction plots	rarefaction_plots.html
Beta diversity results (even sampling: 10000)	
PCoA plot (weighted_unifrac)	index.html
Distance matrix (weighted_unifrac)	weighted_unifrac_dm.txt
Principal coordinate matrix (weighted_unifrac)	weighted_unifrac_pc.txt
PCoA plot (unweighted_unifrac)	index.html
Distance matrix (unweighted_unifrac)	unweighted_unifrac_dm.txt
Principal coordinate matrix (unweighted_unifrac)	unweighted_unifrac_pc.txt

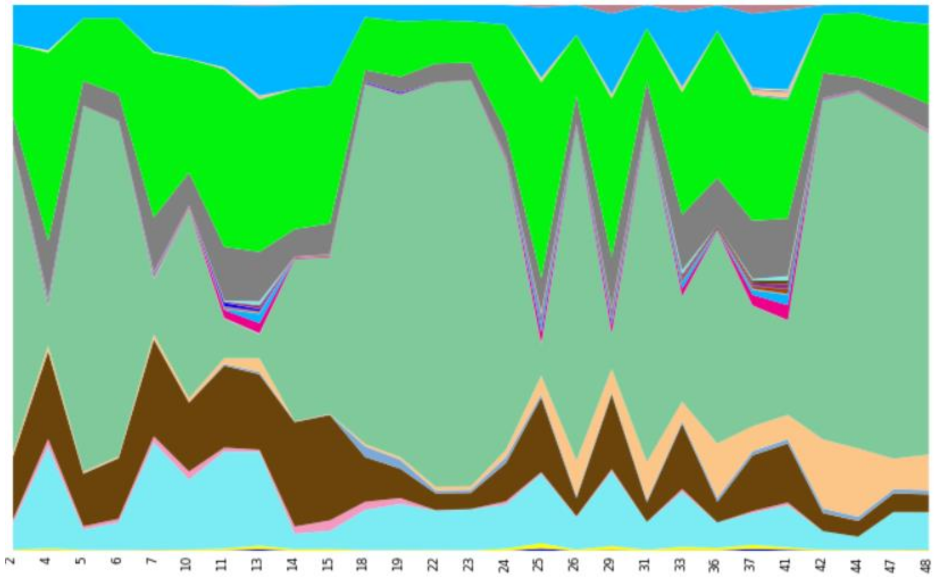
官方使用BaseSpace QIIME分析完的結果可於此處下載:
https://basespace.illumina.com/s/U8W5Qlq1SyfT

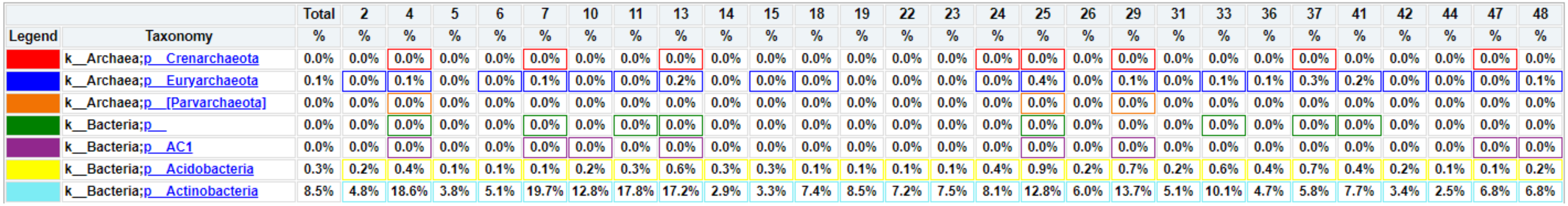
Output data form QIIME Visualizations

Taxa summary bar plots for each level



Taxa summary area plots for each level



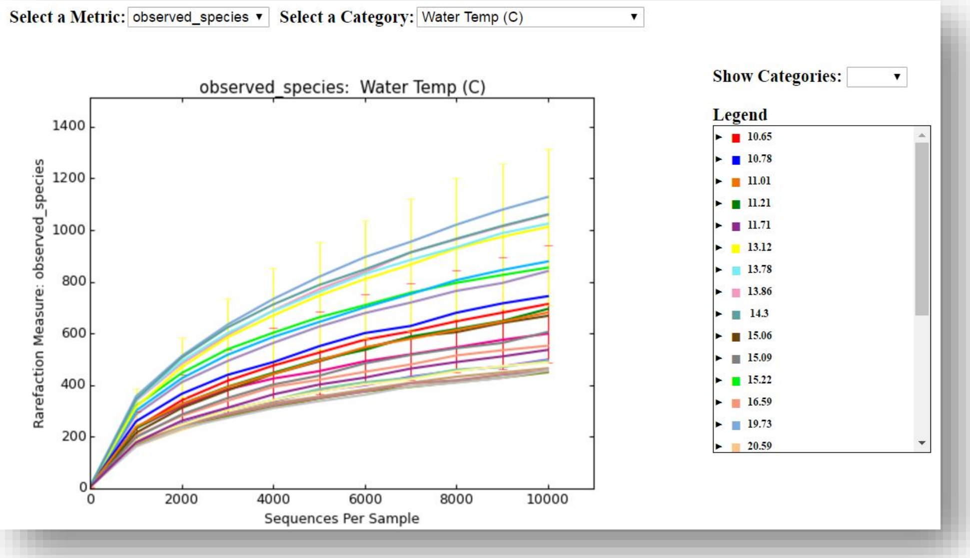


		Total	2	4	5	6	7	10	11	13	14	15	18	19	22	23	24	25	26	29	31	33	36	37	41	42	44	47	48
Legend	Taxonomy	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
🟥	k__Archaea;p__Crenarchaeota	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
🟦	k__Archaea;p__Euryarchaeota	0.1%	0.0%	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.1%	0.0%	0.1%	0.1%	0.3%	0.2%	0.0%	0.0%	0.0%	0.1%
🟧	k__Archaea;p__[Parvarchaeota]	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
🟩	k__Bacteria;p__	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
🟪	k__Bacteria;p__AC1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
🟨	k__Bacteria;p__Acidobacteria	0.3%	0.2%	0.4%	0.1%	0.1%	0.1%	0.2%	0.3%	0.6%	0.3%	0.3%	0.1%	0.1%	0.1%	0.1%	0.4%	0.9%	0.2%	0.7%	0.2%	0.6%	0.4%	0.7%	0.4%	0.2%	0.1%	0.1%	0.2%
🟦	k__Bacteria;p__Actinobacteria	8.5%	4.8%	18.6%	3.8%	5.1%	19.7%	12.8%	17.8%	17.2%	2.9%	3.3%	7.4%	8.5%	7.2%	7.5%	8.1%	12.8%	6.0%	13.7%	5.1%	10.1%	4.7%	5.8%	7.7%	3.4%	2.5%	6.8%	6.8%

Output data form QIIME Visualizations

Alpha diversity results

Alpha rarefaction plots



Select a Metric: observed_species

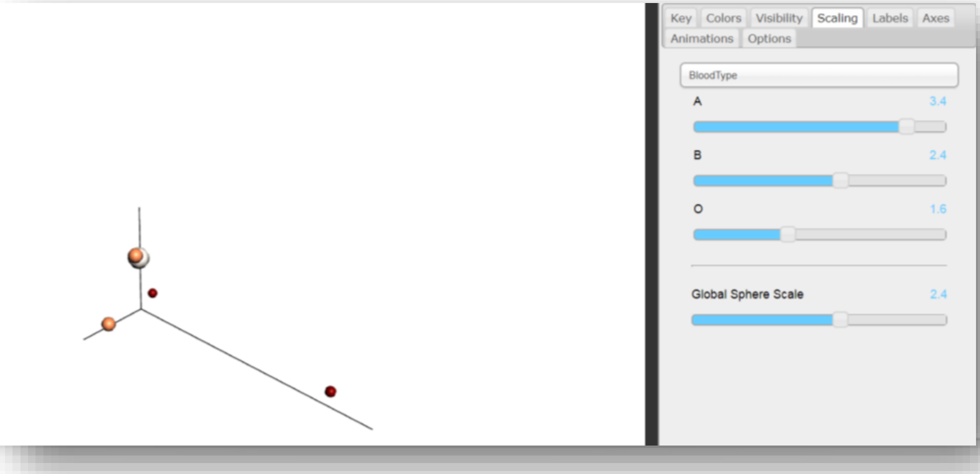
Select a Category: Water Temp (C)

Show Categories:

Beta diversity results

PCoA plot (weighted_unifrac)

Control panel



BloodType

A 3.4

B 2.4

O 1.6

Global Sphere Scale 2.4

Resource

- QIIME Preprocessing: https://basespace.illumina.com/s/Tk0L7PDydgQs
- QIIME Visualizations: https://basespace.illumina.com/s/IYcBNIzGBcW8
- Both: https://basespace.illumina.com/s/U8W5Qlq1SyfT
- QIIME Forum: http://forum.qiime.org/

DNA base sequencing

Library Prep Name	App Name
TruSeq Nano DNA TruSeq DNA PCR-Free	Whole Genome Sequencing Tumor Normal
TruSeq DNA Methylation TruSeq Methyl EPIC	Methylseq Methylkit
TruSeq ChIP	ChIPSeq

RNA base sequencing

Library Prep Name	App Name
TruSeq RNA TruSeq Stranded mRNA TruSeq Stranded Total RNA TruSeq RNA Access	RNA-seq Alignment Cufflinks Assembly & DE
TruSeq Targeted RNA Expression	TruSeq Targeted RNA
TruSeq Small RNA	Small RNA

Data preprocessing

- FASTQ Toolkit
- FastQC

Metagenomics

- 16s Metagenomics
- Prokka Genome Annotation
- Kraken Metagenomics

De novo assembly

- Velvet de novo Assembly

Visualization

- Integrative Genomics Viewer

Public database

- SRA Import
- SRA Submission

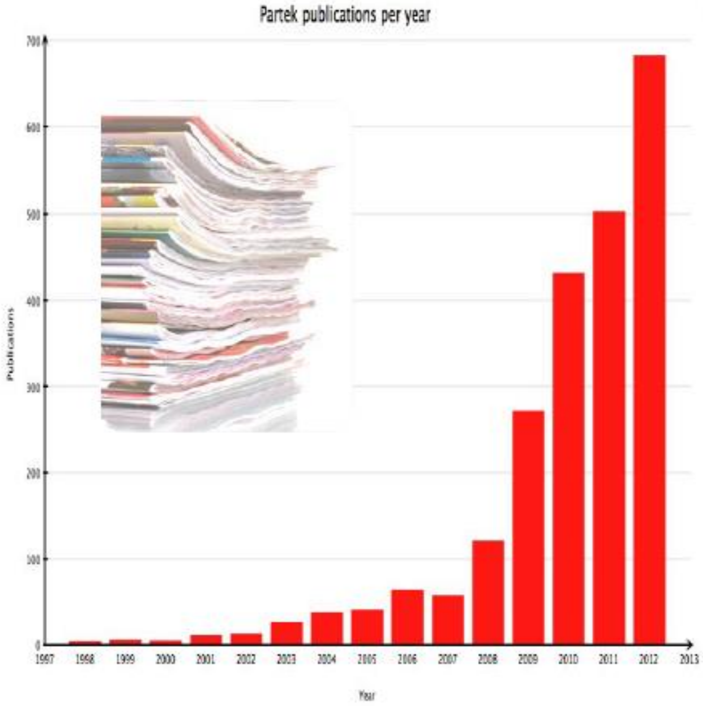
Partek Introduction

Jora Lin 林孟儒
Bioinformatic Specialist
均泰生物科技有限公司
techsupport@gtbiotech.com.tw

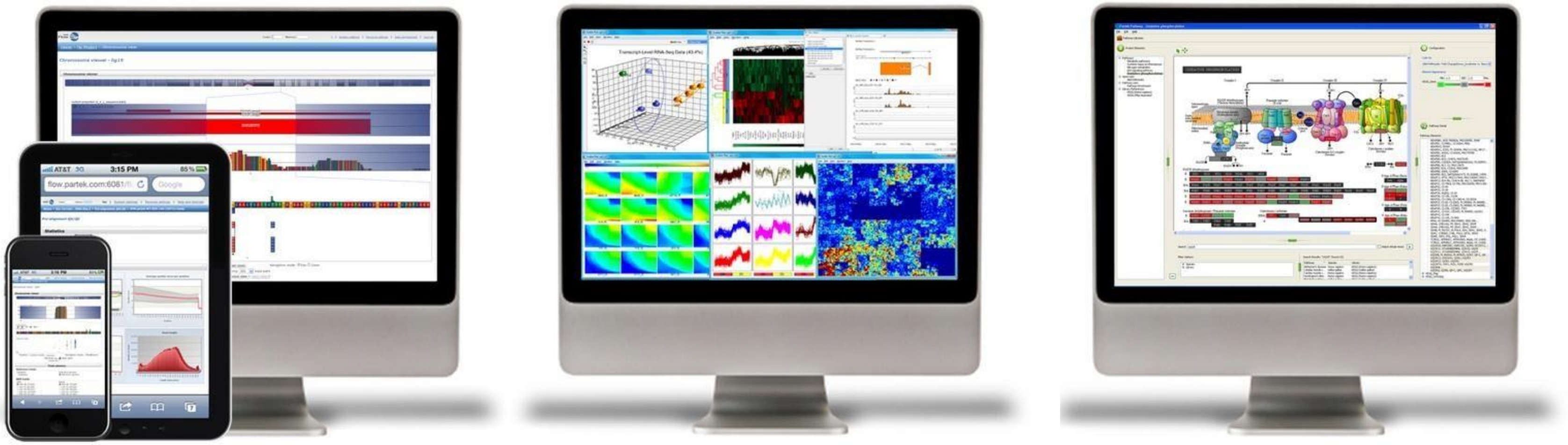
Partek®
turning data into discovery®

Who is Partek?

- Founded in 1993
- Building tools for statistics & visualization
- Focused on genomics
- Thousands of customers worldwide
- Worldwide, world-class customer support



Comprehensive Solution for NGS Analysis



Partek Flow™

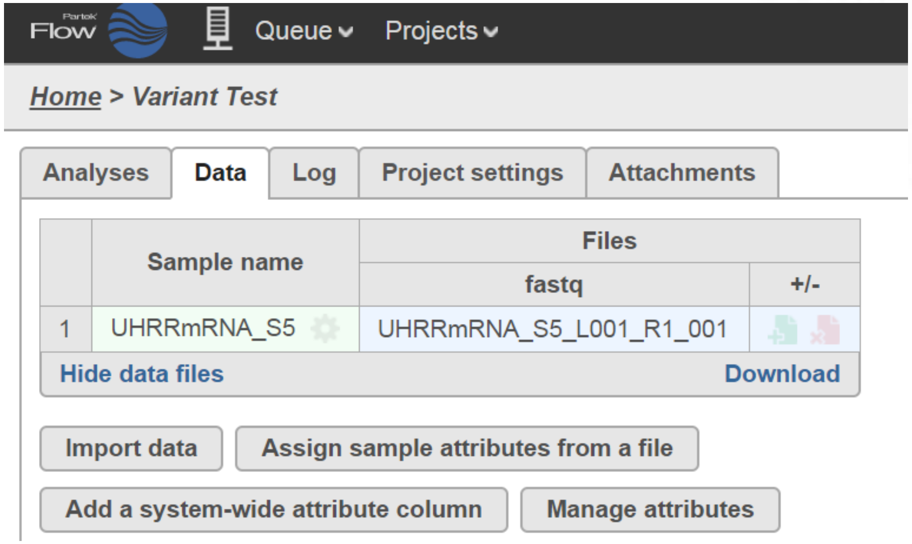
Partek Genomics Suite™

Partek Pathway™

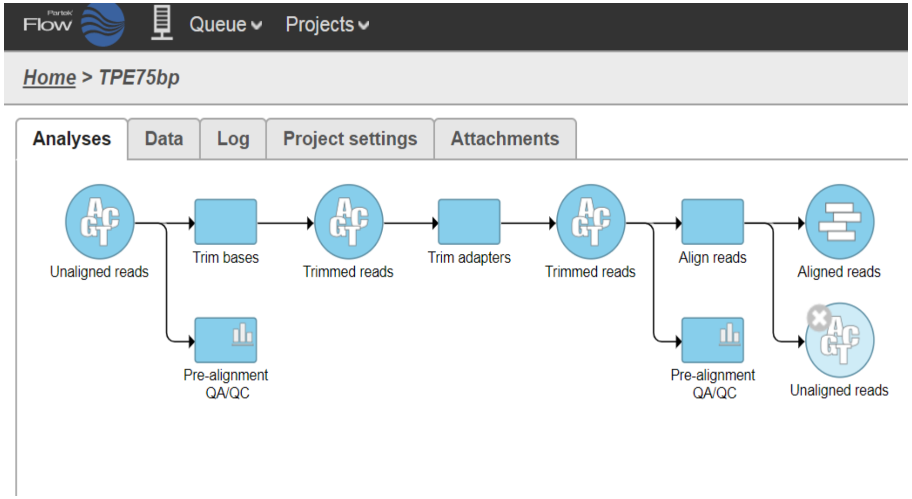
Raw data → QA/QC → Normalization → Statistics → Visualization → Genomics Integrated → Biological Interpretation

Partek® Flow™

- Flow Server
  - Core engine
- Alignment toolkits
  - Different choice of aligners
  - QA/QC on pre- and post-alignments
- Applications (NGS data)
  - RNA-seq, DNA-seq, smallRNA-seq, ChIP-seq, MeDIP-seq
- Customizable protocols
  - Collection of tasks
  - Protocol sharing
- Enable collaboration / project sharing
- Project/Data management
- Accessed from anywhere with internet connection

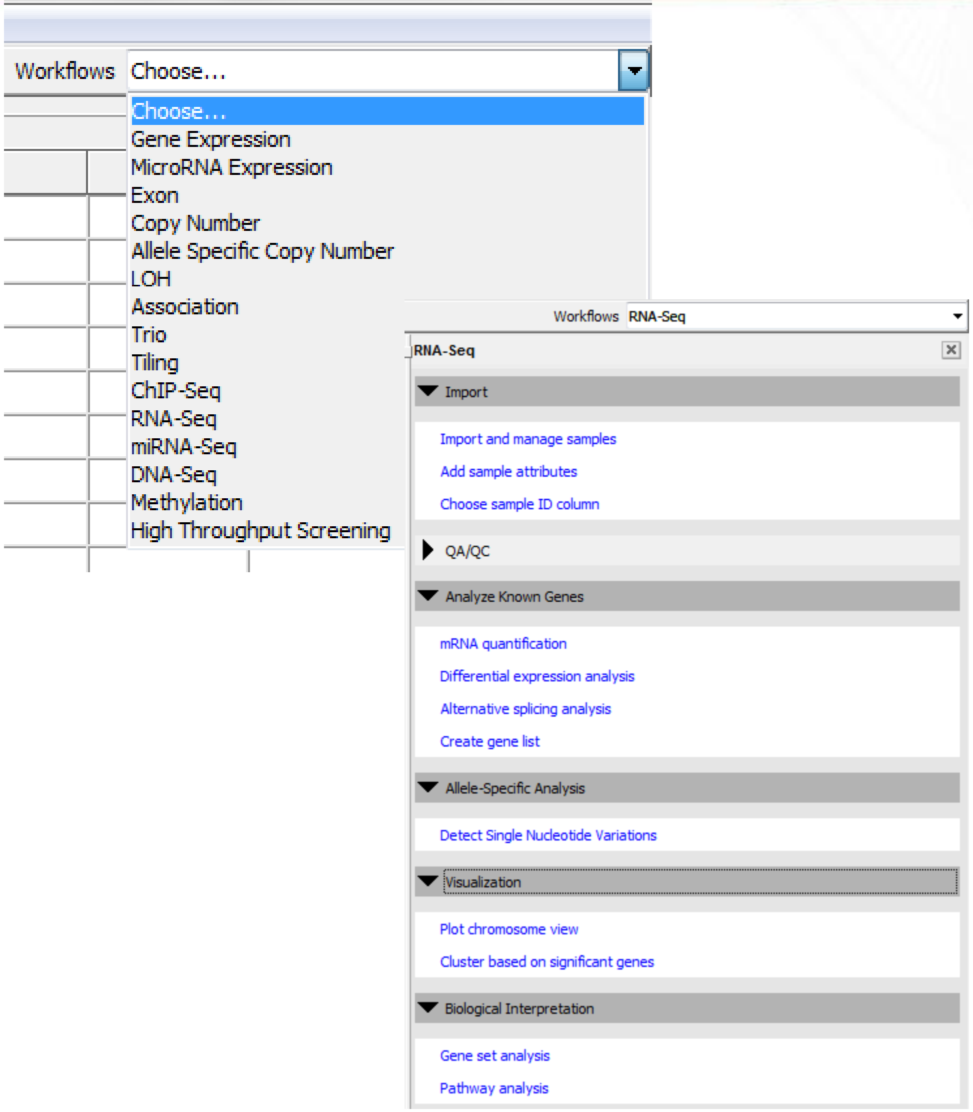


	Sample name	Files	
		fastq	+/-
1	UHRRmRNA_S5	UHRRmRNA_S5_L001_R1_001	



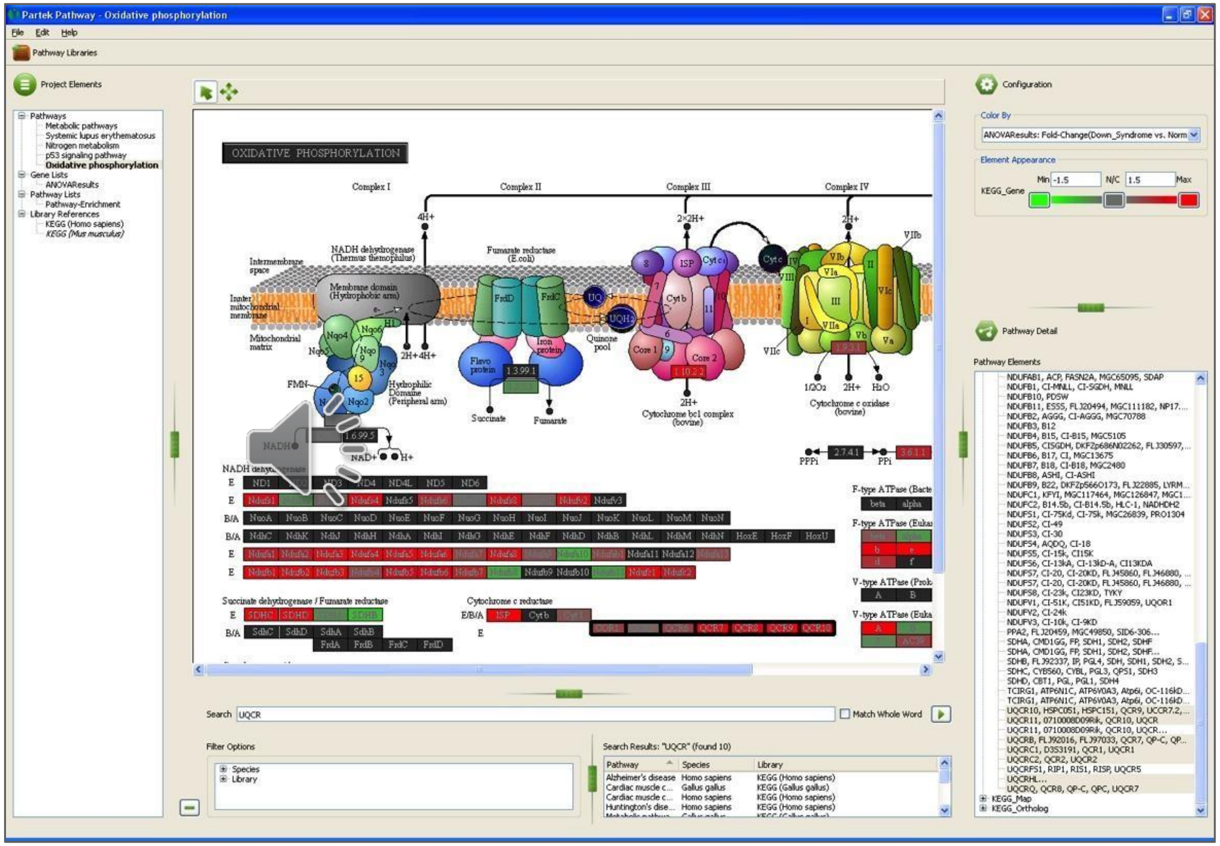
Partek® Genomics Suite™

- Built on statistical and visualization package
- Easy to use workflow by providing step-by-step guidance on how to analyze your data
- Support for integrative analysis between different datasets
- Provide integrative analysis with publicly available database.



Partek® Pathway™?

- Find Enriched Pathways
- Color-code genes based on stat results
- Integrated & interactive KEGG database for human, mouse, & rat as well as over 1,500 other organisms



Color By: ANOVAResults: Fold-Change(Down_Syndrome vs. Norm

Element Appearance: KEGG_Gene Min -1.5 N/C 1.5 Max

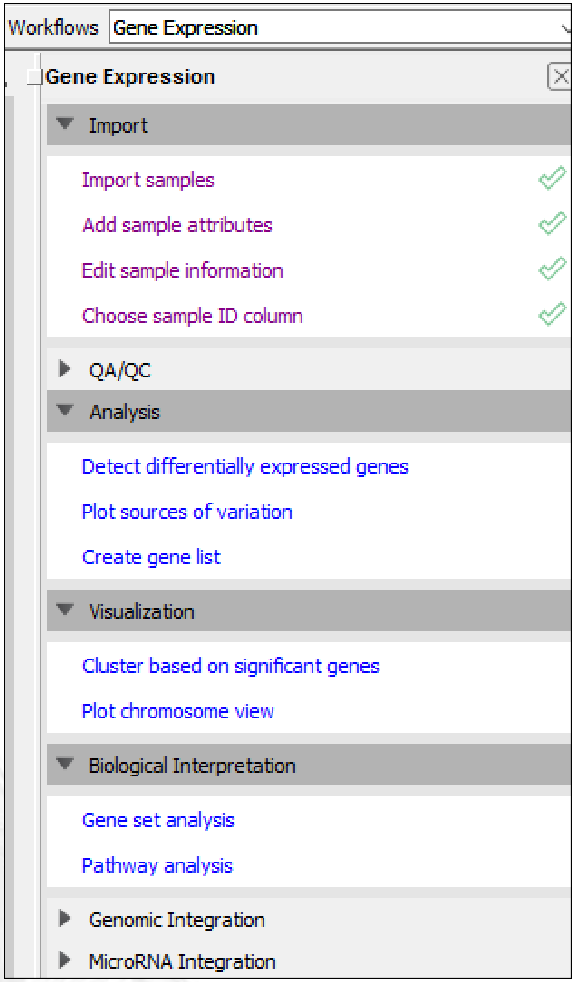
Search: UQCR ☐ Match Whole Word

Search Results: "UQCR" (found 10)

Pathway	Species	Library
Alzheimer's disease	Homo sapiens	KEGG (Homo sapiens)
Cardiac muscle c...	Gallus gallus	KEGG (Gallus gallus)
Cardiac muscle c...	Homo sapiens	KEGG (Homo sapiens)
Huntington's dise...	Homo sapiens	KEGG (Homo sapiens)

Partek Pathway
Extends Biological Interpretation

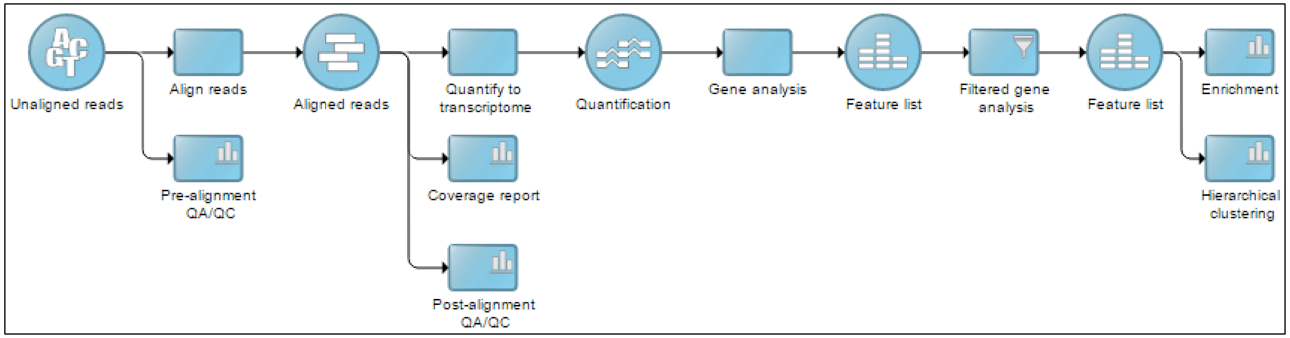
Dedicated, Intuitive Workflows for Ease of Use



Workflows: Gene Expression

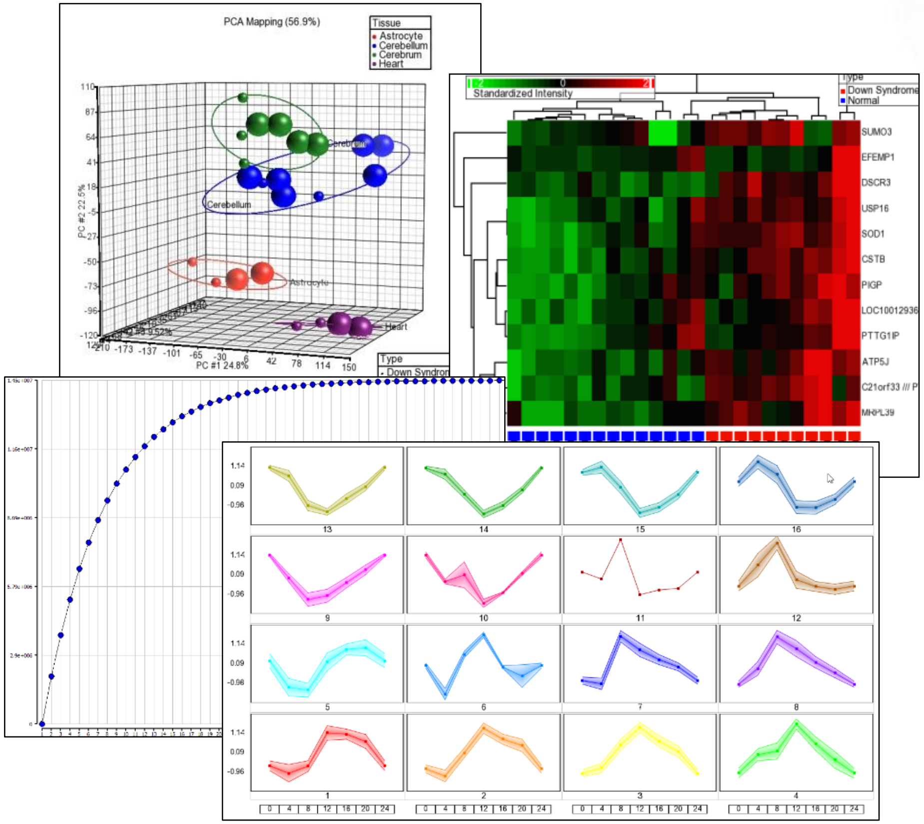
Gene Expression

- Import
  - Import samples ✓
  - Add sample attributes ✓
  - Edit sample information ✓
  - Choose sample ID column ✓
- QA/QC
- Analysis
  - Detect differentially expressed genes
  - Plot sources of variation
  - Create gene list
- Visualization
  - Cluster based on significant genes
  - Plot chromosome view
- Biological Interpretation
  - Gene set analysis
  - Pathway analysis
- Genomic Integration
- MicroRNA Integration



Exploratory Analysis

- High dimensional visualization
  - Principal components analysis (PCA)
  - Multidimensional scaling
- Clustering analysis
  - Hierarchical clustering
  - K-means clustering
  - Self organizing map
- Pattern searching
- Time series data analysis



Inferential Statistics

Parametric

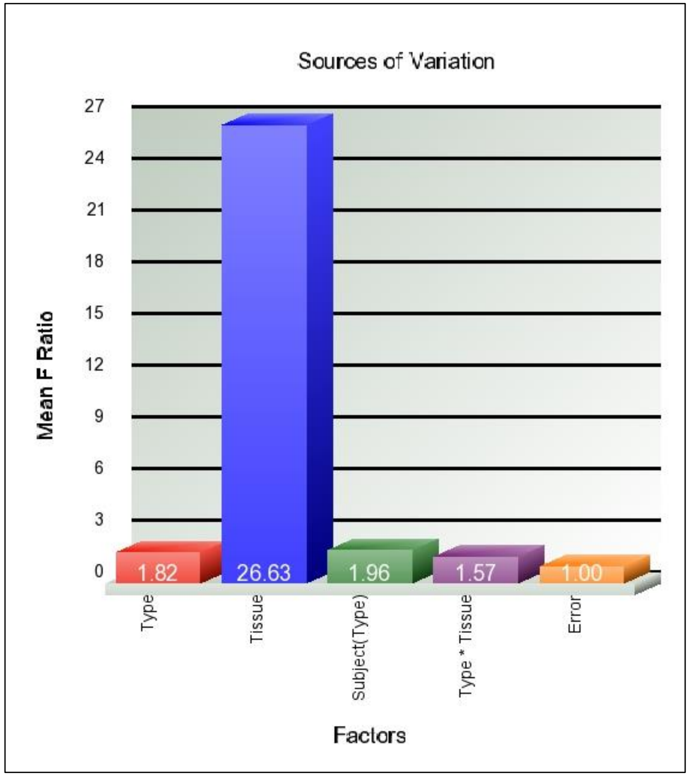
- t-Test
- z-Test
- ANOVA
- Welch’s ANOVA
- Pearson correlation

Non-Parametric

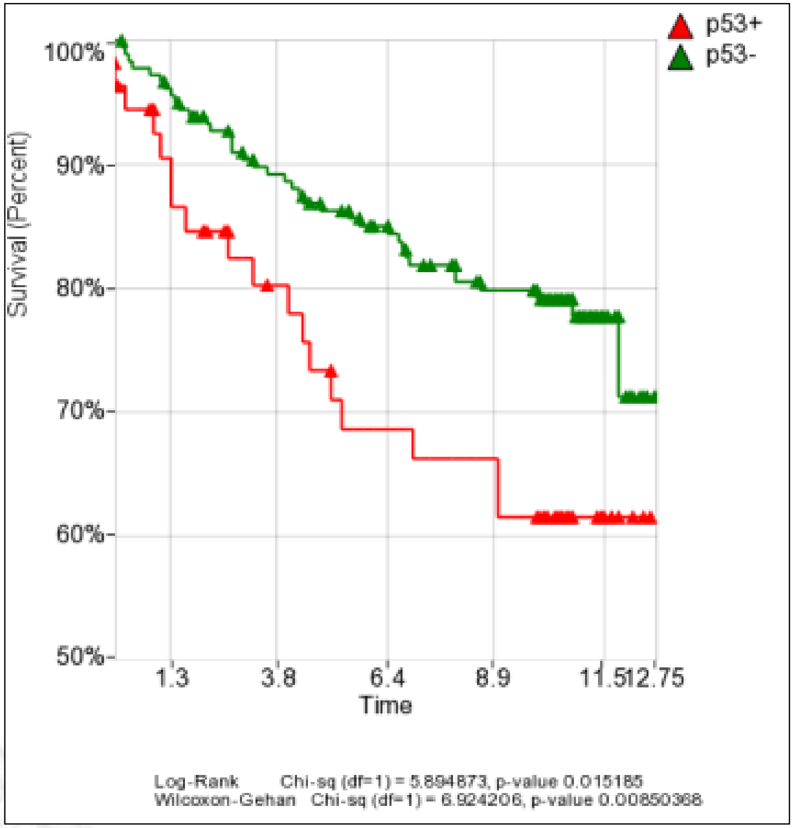
- Mann-Whitney
- Kruskal-Wallis
- Friedman
- Chi-square
- Rank correlation

Others

- Multiple test corrections
- Descriptive statistics



Survival Analysis



Log-Rank Chi-sq (df=1) = 5.894873, p-value 0.015185
Wilcoxon-Gehan Chi-sq (df=1) = 6.924206, p-value 0.00850368

Kaplan-Meier curve

Categorical variable correlation with outcome (e.g. p53 status)

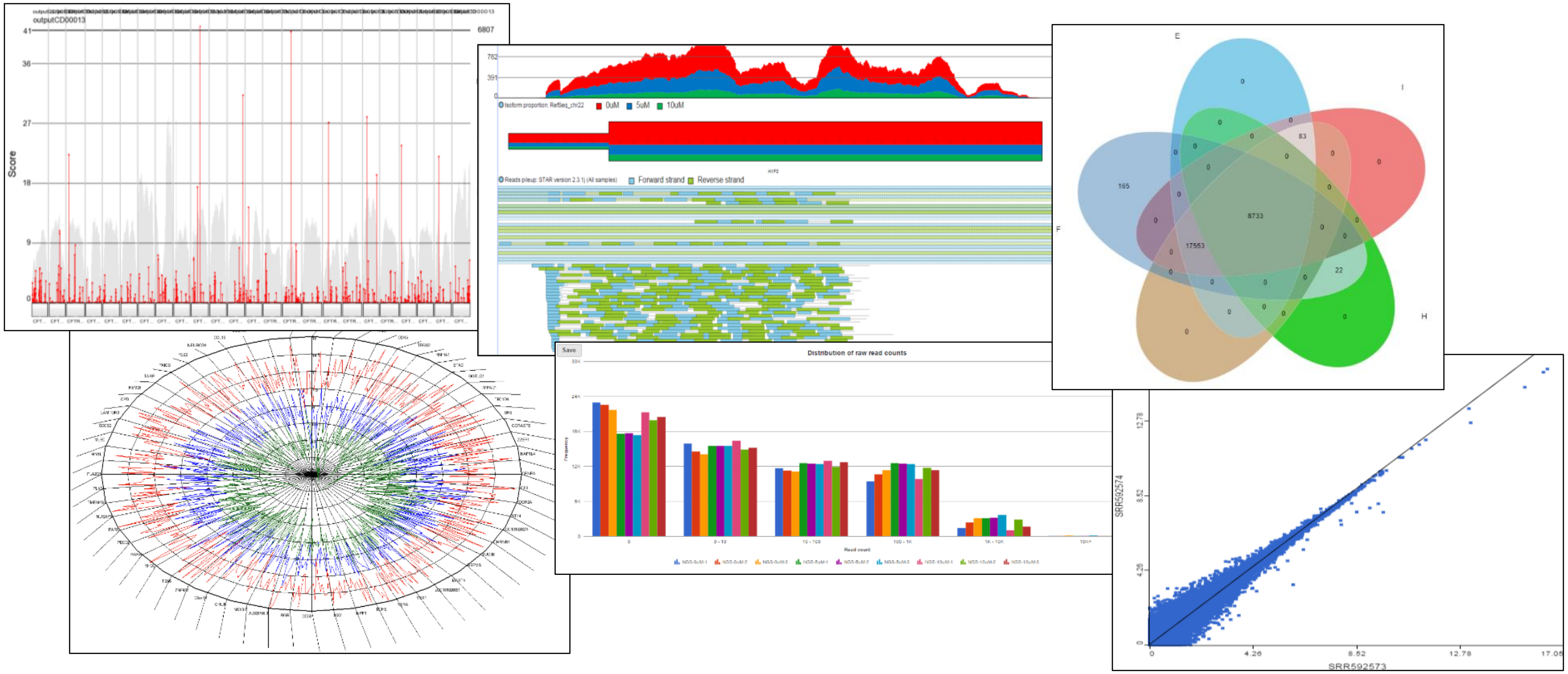
Cox regression

Continuous variable correlation with outcome (e.g. gene expression)

Calculate hazard ratio (HR)

Biomarker evaluation

Rich Visualizations



Intuitive for the Biologist. Powerful for the Bioinformaticist.

Summary

	BaseSpace	**Flow**	**Genomics Suite**
Install	illumina 網頁版雲端分析	Partek 網頁版雲端分析	Partek 單機版分析
可分析項目	Only NGS data 超過80種不同應用的分析流程(app)	Only NGS data DNA-seq, RNA-seq, ChIP-seq, Metagenomics	NGS, Microarray, qPCR Gene Expression, EPIC array, RNA-seq, ChIP-seq…etc.
特色	1. 使用簡單 2. 自動化分析流程 3. 不需調整分析參數	1. 使用簡單 2. 較多統計及繪圖功能 3. 可自建分析流程	1.較多統計及繪圖功能 2. 介面清楚易懂


THANK
YOU