BaseSpace® Apps for Genome-Scale Analysis

Push-button, cloud-based bioinformatics solutions for next-generation sequencing applications.

Highlights

- **Access industry-standard sequencing apps in BaseSpace Sequence Hub for the top biological experiments**
  Analyze genome-size data sets within your timeframe.
- **Unlimited scalability, unmatched usability**
  Server hardware won't limit experimental design; Parameters are optimized for biological application and Illumina data characteristics.
- **Analyze, share, and store securely**
  Designate who accesses your data, securely collaborate with confidence and ease.

Introduction

BaseSpace Apps include 4 of the most common next-generation sequencing (NGS) secondary analysis workflows within the BaseSpace Sequence Hub computing environment. Developed for life science researchers who need simple, comprehensive, and cost-effective analyses, these apps provide scalable, push-button bioinformatics solutions for Illumina sequencing data. Requiring only an internet connection, BaseSpace Apps simplify genome-scale data for 4 of the most commonly used biological applications, including whole-genome or exome sequencing (Figure 1), tumor-normal whole-genome sequence analysis, transcript assembly, and expression profiling. Researchers can now devote significantly less effort converting data to results, and more time focusing on the next cycle of research.

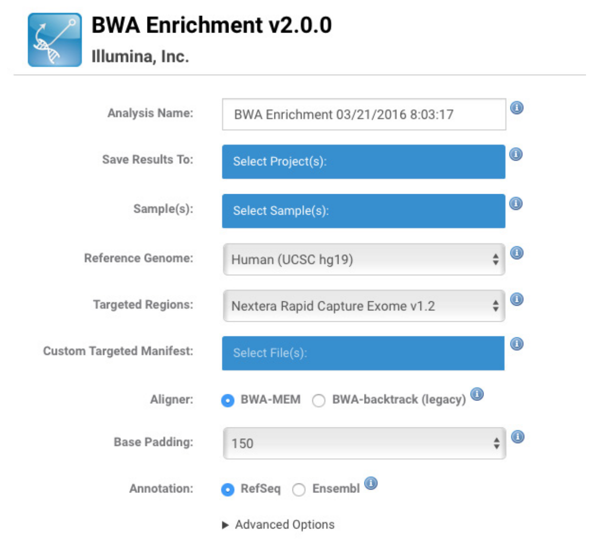
Tailor-Made Solutions for NGS Applications

Commonly, research progresses along a nonlinear path, and data from one experiment often open up additional avenues of exploration. For example, a whole-genome sequencing experiment may lead to targeted sequencing for larger sampling. Or somatic variants identified in a tumor-normal sample pair may reveal candidates for gene expression profiling using Tophat/Cufflinks. BaseSpace Apps were designed to complement the pace of research by making the most frequently used sequencing applications available within a single environment.

Easy Access to Genome-Scale Analysis

The suite of BaseSpace Apps enables genome-size analysis for any researcher working with Illumina sequencing data. With simple input parameters, getting started with secondary analysis is uncomplicated (Figure 1). Ready-to-publish output is a few clicks away, along with the ability to generate customized PDF reports that chart key stats, plots, and tables with your most relevant data. For example, use the Cufflinks app for an interactive RNA report to assess differential expression levels and drill down to individual genes or isoforms of interest (Figure 2).

Figure 1: Simple Input Parameters



BWA Enrichment v2.0.0
Illumina, Inc.

Analysis Name: BWA Enrichment 03/21/2016 8:03:17
Save Results To: Select Project(s):
Sample(s): Select Sample(s):
Reference Genome: Human (UCSC hg19)
Targeted Regions: Nextera Rapid Capture Exome v1.2
Custom Targeted Manifest: Select File(s):
Aligner: ● BWA-MEM ○ BWA-backtrack (legacy)
Base Padding: 150
Annotation: ● RefSeq ○ Ensembl

Set up a multisample exome analysis with just a few clicks.

Due to the virtually unlimited scalability of the cloud, the number of samples, or the amount of data per sample (read depth) can be scaled up to fine-tune the analysis according to your specific research needs. Server hardware limitations will no longer impede your experimental design. All the BaseSpace Apps allow aggregation of multisample reports, provide notification of job completion that can be received on mobile devices, and enable efficient file organization for collaboration and sharing.

Get More Out of Your Data

Data output in standardized file formats such as FASTQ for reads and VCF for variants, means you can reinterrogate data with different parameter settings, using different methods, or with varying amounts of data. With a wide variety of third-party tools available in BaseSpace Sequence Hub for downstream analysis, data from an experiment can be repurposed for a different analysis or added in to bigger studies. All BaseSpace Apps have demo data to familiarize you with the input/output and get you started quickly[1].

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACC
GTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGT

Table 1: BaseSpace App Details

BaseSpace App	Feature	Advantage
TopHat Alignment	RNA-Seq alignment using industry-standard tool, and cSNP calling using Isaac Variant Caller Works with all samples, either fresh-frozen or formalin-fixed, paraffin embedded (FFPE) Results seamlessly integrated into the Cufflinks Assembly and Differential Expression App	Single solution enables advanced experiments such as novel transcript assembly, variant calling, and fusion detection Unlock all your samples
Cufflinks Assembly and Differential Expression	Differential expression and gene fusion calling Works with all samples, either fresh-frozen or formalin-fxed, paraffin embedded (FFPE) Preformatted, interactive reports allow aggregation of multiple samples	Quickly assess expression levels and drill down on significantly expressed genes/ isoforms within the BaseSpace Sequence Hub
Whole-Genome Sequencing	Two analysis methods: BWA Alignment and GATK Variant Calling, or Isaac Alignment and Variant Calling Detects SNPs, indels, copy number variations, and structural variations Works with data from all standard whole-genome kits 12 common references supported	Flexibility in choosing either the most popular DNA aligner in the community or the fast and accurate Isaac Interrogate a wide variety of samples using a unified analysis method Study many model organisms
Exome Enrichment*	Two analysis methods: BWA Alignment and GATK Variant Calling for exomes, or Isaac Alignment and Variant Calling for exomes Provides standard enrichment metrics and Picard HS Metrics	Flexibility of using DNA aligner of choice Zero in on sample results and variants with a single click; quickly identify off-target rates and biological context of SNVs and indels across all your samples Integrate output into your preexisting workflows
Tumor-normal Analysis	Combined analysis of whole-genome tumor-normal data using Strelka[2] (recommended tumor-normal average target depth: 80×/40×) Detects SNPs, indels, copy number variations, and structural variations Works with data from all standard whole-genome kits	Accurate somatic calling method accounts for tumor impurity and heterogeneity Comprehensive assessment of the largest number of somatic events Interrogate a wide variety of samples using a unified analysis method

*For data generated with Nextera Rapid Capture Kits

Collaborate with Confidence

BaseSpace Apps users can instantly share data with colleagues across campus or across the globe. With the option to share data and results by anyone at any time, links can be easily created and emailed to anyone in the world, giving your collaborators instant access to your results and to the sequencing runs that created them. Equally suited for members of a lab or members of an international consortium, BaseSpace Sequence Hub is the central storage and analysis repository that keeps your data secure and accessible.

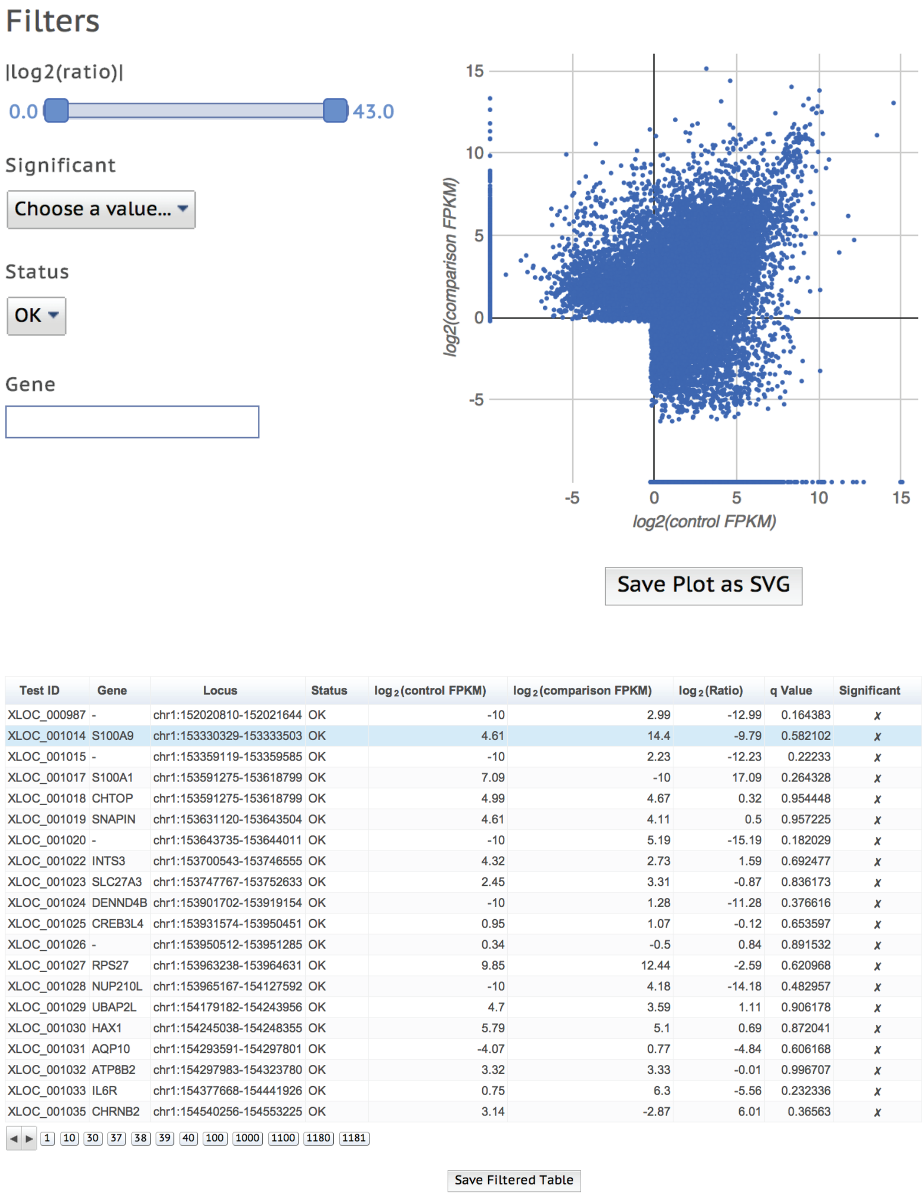
Get Started Today

Go to basespace.illumina.com to sign up for your free BaseSpace account and access the BaseSpace Apps.

References

1. https://basespace.illumina.com/datacentral
2. Saunders CT, Wong W, Swamy S, Becq J, Murray L, et al., (2012) Strelka: accurate somatic small variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28(14) 1811–1817.

Figure 2: Intuitive, Interactive Reports for Differential Expression



Test ID	Gene	Locus	Status	$\log_2$(control FPKM)	$\log_2$(comparison FPKM)	$\log_2$(Ratio)	q Value	Significant
XLOC_000987	-	chr1:152020810-152021644	OK	-10	2.99	-12.99	0.164383	✗
XLOC_001014	S100A9	chr1:153330329-153333503	OK	4.61	14.4	-9.79	0.582102	✗
XLOC_001015	-	chr1:153359119-153359585	OK	-10	2.23	-12.23	0.22233	✗
XLOC_001017	S100A1	chr1:153591275-153618799	OK	7.09	-10	17.09	0.264328	✗
XLOC_001018	CHTOP	chr1:153591275-153618799	OK	4.99	4.67	0.32	0.954448	✗
XLOC_001019	SNAPIN	chr1:153631120-153643504	OK	4.61	4.11	0.5	0.957225	✗
XLOC_001020	-	chr1:153643735-153644011	OK	-10	5.19	-15.19	0.182029	✗
XLOC_001022	INTS3	chr1:153700543-153746555	OK	4.32	2.73	1.59	0.692477	✗
XLOC_001023	SLC27A3	chr1:153747767-153752633	OK	2.45	3.31	-0.87	0.836173	✗
XLOC_001024	DENND4B	chr1:153901702-153919154	OK	-10	1.28	-11.28	0.376616	✗
XLOC_001025	CREB3L4	chr1:153931574-153950451	OK	0.95	1.07	-0.12	0.653597	✗
XLOC_001026	-	chr1:153950512-153951285	OK	0.34	-0.5	0.84	0.891532	✗
XLOC_001027	RPS27	chr1:153963238-153964631	OK	9.85	12.44	-2.59	0.620968	✗
XLOC_001028	NUP210L	chr1:153965167-154127592	OK	-10	4.18	-14.18	0.482957	✗
XLOC_001029	UBAP2L	chr1:154179182-154243956	OK	4.7	3.59	1.11	0.906178	✗
XLOC_001030	HAX1	chr1:154245038-154248355	OK	5.79	5.1	0.69	0.872041	✗
XLOC_001031	AQP10	chr1:154293591-154297801	OK	-4.07	0.77	-4.84	0.606168	✗
XLOC_001032	ATP8B2	chr1:154297983-154323780	OK	3.32	3.33	-0.01	0.996707	✗
XLOC_001033	IL6R	chr1:154377668-154441926	OK	0.75	6.3	-5.56	0.232336	✗
XLOC_001035	CHRNB2	chr1:154540256-154553225	OK	3.14	-2.87	6.01	0.36563	✗

The Cufflinks Report includes an interactive differential expression gene browser to view genes and transcripts that are differentially expressed with an interactive slider to control the fold-expression level and significance.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 970-2013-008 Current as of 21 March 2016