BaseSpace® Variant Interpreter

Rapidly extract and report biological insight from genomic data with a scalable solution.

Highlights

- **User-Friendly, Efficient Variant-to-Report Workflow**
  Seamless integration with BaseSpace Sequence Hub, and an intuitive user interface
- **Accelerated Variant Interpretation with BaseSpace Knowledge Network**
  Comprehensive compilation of genomic data from a broad range of sources
- **Increased Lab Operational Efficiency and Scalability**
  Integrated knowledge base containing genotype–phenotype associations
- **Supported Security and Lab Compliance**
  Security features include access control, audit trail, and test configuration

Introduction

The extraction of biologically relevant information from genomic variant data often poses a challenge for genetic analysis and interpretation. Variant annotation and filtering aid in placing data into biological context. These processes provide the functional information necessary to identify variants of interest, determine their biological impact, and identify potential correlations between genotype and phenotype.

BaseSpace Variant Interpreter is a cloud-based interpretation and reporting platform designed to decrease the time and effort required to extract biological insight from genomic data while maximizing operational efficiency (Table 1). It enables labs to implement standardized workflows, identify and interpret disease-relevant variants rapidly, and generate a report summarizing findings in a structured format.

BaseSpace Variant Interpreter provides direct access to the BaseSpace Knowledge Network, a system of knowledge bases containing genotype–phenotype associations. In the initial release of BaseSpace Knowledge Network, users have access to over 60,000 associations curated by the Illumina Biomedical Informatics team. This curated content leverages years of experience by a specialized team. Details on the content can be found in the BaseSpace Knowledge Network Tech Note.[1]

BaseSpace Variant Interpreter features rapid rich annotation capabilities (Table 1), an intuitive and highly customizable filtering system, and flexible reporting functionality, all within a software framework focused on data security and lab compliance (Figure 1, Table 2).

User-Friendly and Efficient Variant-to-Report Workflow

BaseSpace Variant Interpreter features an intuitive user interface, enabling clinical researchers to analyze DNA variants from sequencing data with ease. Integrated with BaseSpace Sequence Hub, BaseSpace Variant Interpreter can generate and store variant call format (VCF) and genomic VCF (gVCF) files. The seamless importation of data from BaseSpace Sequence Hub expedites downstream analysis and interpretation. With a flexible framework, users can quickly interpret and select variants of interest to include in customized reports tailored to the needs of each laboratory.

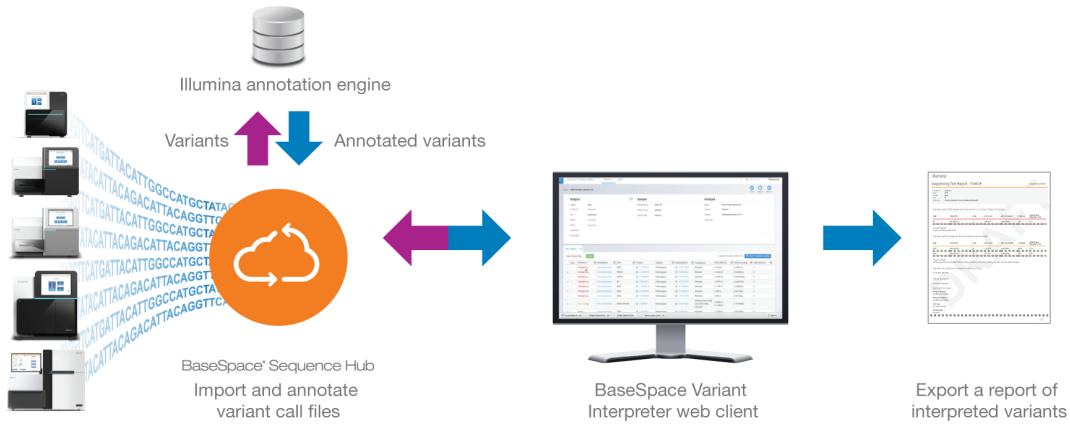

Figure 1: Intuitive User Interface Simplifies Analysis—BaseSpace Variant Interpreter is a powerful reporting solution for analyzing and interpreting variant data. This solution aggregates information from a collection of databases to streamline annotation. It also provides flexible filtering options for analyzing variant data and tools to enable classification and reporting of biologically relevant variants.

Table 1: Annotation Categories

Category	Description
Transcript Consequence	Predicts transcript changes resulting from the variant of interest, allowing users to segregate synonymous from different types of nonsynonymous changes
Functional Impact	Predicts whether amino acid substitutions affect protein function, indicating variants that are likely deleterious
Allele Frequency	Provides the frequency of a variant within a population
Conserved Sequence	Denotes sequence similarity if the variant occurs between species, providing phylogenetic information and evolutionary context
Disease Association	Indicates whether variant has been previously associated with disease

BaseSpace Variant Interpreter provides annotations at variant, transcript, and gene levels for comprehensive assessment of the biological impact of genetic variation.

Rapid, Rich, and Accurate Annotation

BaseSpace Variant Interpreter rapidly automates rich and accurate annotations from the Illumina annotation engine. By aggregating information from multiple sources into a single, maintained database, BaseSpace Variant Interpreter eliminates the need for manual assembly of variant information from disparate sources, streamlining the annotation process. Rich annotation is done at the gene, variant, and transcript levels. Annotations from both Ensembl[2] and NCBI Reference Sequence Database (RefSeq)[3] are supported and annotations can be viewed for canonical transcripts or for all transcripts. The Illumina annotation engine harnesses databases such as RefSeq with algorithms such as Polymorphism Phenotyping (PolyPhen)[4] and Scale-invariant Feature Transform (SIFT).[5] Information about known disease association can be obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC),[6] ClinVar,[7] and Online Mendelian Inheritance in Man (OMIM).[8] Resources such as the Single Nucleotide Polymorphism Database (dbSNP),[9] the Ensembl 1000 Genomes Project,[10] Exome Variant Server (EVS),[11] and Exome Aggregation Consortium (ExAC)[12] provide information about the occurrence and frequencies of variants within a population. By delivering rapid and thorough automated annotation processes, BaseSpace Variant Interpreter empowers researchers to identify biological significance in variant data.

Comprehensive, Flexible Filtering Tools

BaseSpace Variant Interpreter provides a comprehensive set of filters for isolating variants based on user-defined selections. Using a cascade of filtering options, researchers can rapidly isolate the key variants that are biologically relevant to the phenotype under study. For ease of use, BaseSpace Variant Interpreter offers commonly applied filters, including variant quality, population frequency, functional impact, modes of inheritance, and known disease association, which are easily accessible through the user interface (Figure 2). Filters can be saved and applied to future samples, enforcing consistent application of the validated workflow for each lab.

Streamlined Variant Interpretation

A critical component of translating genetic information into meaningful biological insight is determining the impact of identified variants within the context of observed phenotypes. BaseSpace Variant Interpreter applies a rule set for pathogenicity autoscoring. After autoscoring ranks and prioritizes variants, users must interpret each variant. The classification feature in BaseSpace Variant Interpreter facilitates this process, enabling clinical researchers to combine their expertise with the provided predictions and annotations to categorize variants. Germline classification is based on American College of Medical Genetics and Genomics (ACMG) guidelines and oncology classification is based on levels of associated evidence. By enabling automatic recording, tracking, and management of classified variants, BaseSpace Variant Interpreter simplifies and accelerates data interpretation.

The addition of BaseSpace Knowledge Network further streamlines variant interpretation by providing premium curated content directly associated with specific variants. The autoscoring feature uses this association content to provide additional data for generating predictions to categorize variants. The presence of curated content for a given variant is shown in the variant grid, and the association can be examined through the variant details view. The detailed variant view provides in-depth information on the genotype–phenotype association, including publication IDs, guidelines, and other forms of evidence. If users find this information relevant to their sample, the data can be copied and pasted into their own knowledge base, and added to a given report.

Table 2: Advantages of BaseSpace Variant Interpreter

Feature	Specification
Broad range of annotation sources	Includes dbSNP, COSMIC, ClinVar, 1000 Genomes Project, EVS, ExAC, PolyPhen, SIFT, and OMIM
Interactive filtering	Includes allele frequency, quality scores, read depth, variant type, functional impact, genotype, available annotation, and overlap with defined gene set
Flexible settings	Supports custom filters and saves selected filters and workflows for convenient reuse
Streamlined classification	Automatically stores variant annotations to facilitate future use
Customizable reporting	Enables generation of custom reports that can be optimized for different disease research areas

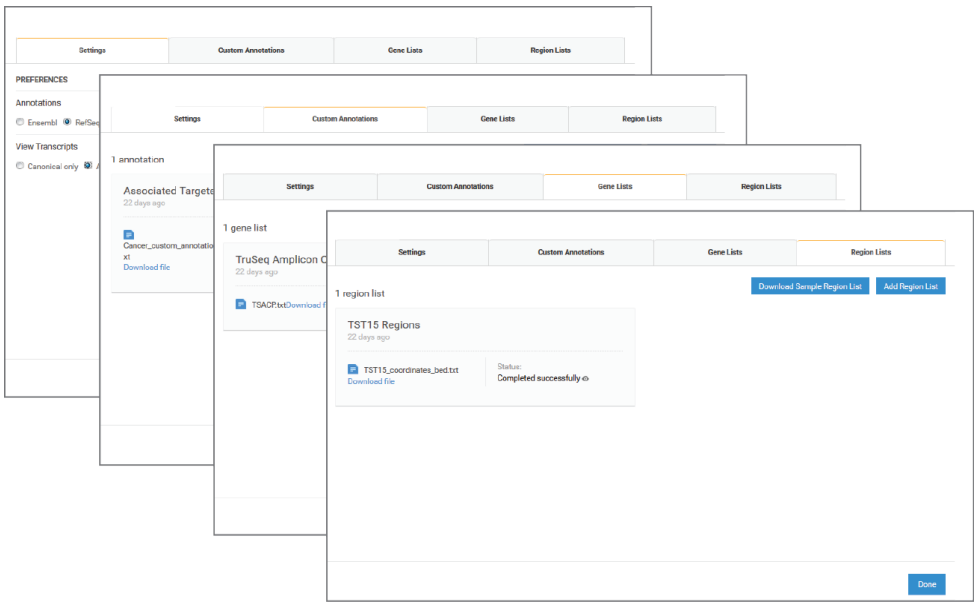

Figure 2: Settings and Custom Annotations—Filter variants are based on any available associated information, including user-imported custom annotations.

Increase Lab Operational Efficiency and Scalability

BaseSpace Variant Interpreter is built from the ground up with a software architecture capable of supporting high-throughput whole-genome sequencing. Labs can increase operational efficiency and scalability by using BaseSpace Variant Interpreter customization capabilities. This includes user workgroups, conducive to labs of any scale, with multiple users across multiple sites having the ability to build customized annotation files, gene lists, and region lists as well as the ability to customize filters, views, and workflows.

Customized Files

Exome and whole-genome sequencing often detect large numbers of variants per sample, requiring further analysis to identify the few variants associated with a given phenotype. Virtual subpanels from these broad sequencing assays can be generated within BaseSpace Variant Interpreter using a custom gene or region list enabling the push-button creation of a broad sequencing portfolio. To maximize flexibility, BaseSpace Variant Interpreter enables users to filter variants based on any available associated information, including user-imported custom annotations. Customized files can be saved, uploaded, and applied to future samples.

Customized Views and Workflows

For further convenience, BaseSpace Variant Interpreter allows a lab to construct workflows by saving concatenated filters, enabling researchers to standardize workflows and streamline analysis. In addition to single-sample filtering, BaseSpace Variant Interpreter enables multisample comparisons that accelerate identification of causative variants. BaseSpace Variant Interpreter provides a view of different analyses of the same data set with novel tab-based, side-by-side comparisons. This enables variants to be tested by different hypotheses easily. Using BaseSpace Variant Interpreter, researchers can confidently isolate disease-relevant variants.

Customizable Reporting

The ability to summarize interpretation of relevant variants from complex data is critical. BaseSpace Variant Interpreter provides powerful and flexible report generation capabilities to facilitate variant interpretation reporting. Users have the flexibility to select which variants they would like to include in their reporting. When selected, BaseSpace Variant Interpreter report building module enables researchers to create and store multiple customized templates. Sample-specific information may include detected variants in selected classification categories, along with their interpretation, references, and additional comments. This flexible system enables researchers to customize reports and accommodate requirements that are specific to different areas of disease research. Reports are then exported in PDF format for downstream use (Figure 3).

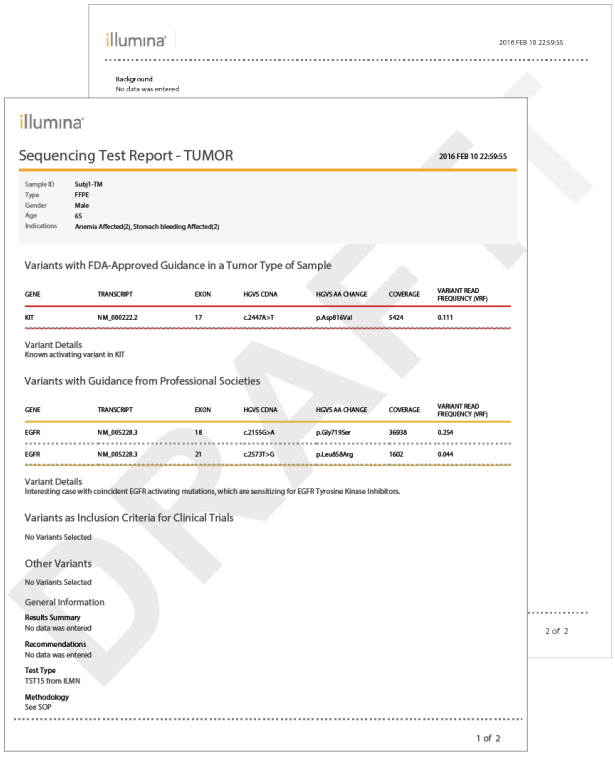

Figure 3: Templates Enable Customizable Reporting—Guided template generation and addition of sample-specific information enable customized reporting for different clinical research areas.

Supports Security and Lab Compliance

Labs are mandated to comply with many complicated regulations. BaseSpace Variant Interpreter is designed to support compliance with data security, version control, and traceability.

- Data Security—Each user of BaseSpace Variant Interpreter goes through authentication for access control.
- Version Control—The software and its internal components are version controlled and users must accept an upgrade to a newer version after it is released.
- Traceability—A trace log is created with every interpretation and report, including software versions that can be exported for auditing purposes.

BaseSpace Variant Interpreter uses Amazon Web Services that are ISO 27001 certified.

Summary

BaseSpace Variant Interpreter is a powerful analysis software platform that reduces time and effort in transforming genomic information into biological insight. The simple user interface delivers an intuitive framework for nonexpert users to annotate, filter, and interpret variant data easily. With flexibility, reliability and efficiency, BaseSpace Variant Interpreter provides a robust method for enriching variant information with biological context.

Learn More

BaseSpace Variant Interpreter is available for a limited amount of time as a free trial. For more information visit www.illumina.com/variantinterpreter.

References

1. Illumina (2016) BaseSpace Knowledge Network (www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/basespace-knowledge-network-technical-note-970-2016-023.pdf).
2. Ensembl. uswest.ensembl.org/index.html. Accessed September 8, 2017.
3. RefSeq: NCBI Reference Sequence Database. www.ncbi.nlm.nih.gov/refseq/. Accessed September 8, 2017.
4. PolyPhen-2 prediction of functional effects of human nsSNPs. genetics.bwh.harvard.edu/pph2/. Accessed September 8, 2017.
5. SIFT. J. Craig Venter Institute website. sift.jcvi.org/. Accessed September 8, 2017.
6. Catalogue of somatic mutations in cancer. cancer.sanger.ac.uk/cosmic. Accessed September 8, 2017.
7. ClinVar. www.ncbi.nlm.nih.gov/clinvar/. Accessed September 8, 2017.
8. OMIM. www.ncbi.nlm.nih.gov/omim. Accessed September 8, 2017.
9. dbSNP Short Genetic Variations. www.ncbi.nlm.nih.gov/projects/SNP. Accessed September 8, 2017.
10. IGSR: The International Genome Sample Resource. www.internationalgenome.org/. Accessed September 8, 2017.
11. NHLBI Exome Sequencing Project (ESP): Exome Variant Server evs.gs.washington.edu/EVS/. Accessed September 8, 2017.
12. Exome Aggregation Consortium: ExAC Browser. exac.broadinstitute.org/. Accessed September 8, 2017.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

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

illumina®