

Illumina VariantStudio Data Analysis Software

A powerful analysis and reporting tool that provides biological insight into genomic variant data.

Highlights

· Easy-to-Use Application

Intuitive user interface enables easy data analysis and exploration, without requiring bioinformatics expertise

· Rapid and Complete Annotation

Software aggregates information from a broad range of sources into a single database for comprehensive annotation of genomic data

· Streamlined Filtering and Classification

Extensive filters and efficient categorization processes streamline assessment of biologically relevant variants

Customizable Reporting

Flexible report generation summarizes disease-relevant and biologically significant results in a structured format

Introduction

The extraction of phenotypically relevant information from genomic variant data often poses a challenge for genetic analysis. Variant annotation and filtering processes are critical for 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. Illumina VariantStudio software features rapid annotation capabilities, intuitive filtering processes, and flexible classification and

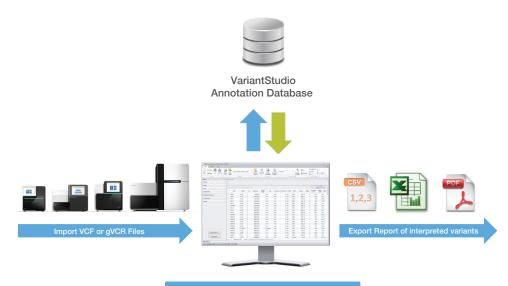
reporting functionalities to simplify data analysis and interpretation. This software application enables researchers to explore variant data, identify and classify disease-relevant variants, and report biologically significant findings (Figure 1).

Easy-to-Use Application

VariantStudio software features an intuitive user interface (Figure 2), enabling researchers to analyze sequencing data easily. Designed for scalable analysis, this tool is optimized for data sets ranging in size and complexity, from targeted sequencing to wholegenome data. With VariantStudio software, researchers can perform necessary analyses without informatics expertise.

Rapid and Complete Annotation

By aggregating information from multiple sources into a single, maintained database, the VariantStudio tool eliminates the need for manual assembly of variant information from disparate sources, streamlining the annotation process. This desktop client provides a convenient local environment to analyze and store variant data. The comprehensive database captures annotations at variant, gene, and transcript levels (Table 1). Variant Effect Predictor (VEP)¹ is a central resource for thorough annotation of transcript consequences. VEP also harnesses databases such as NCBI Reference Sequence Database (RefSeq)² and algorithms such as Polymorphism Phenotyping (PolyPhen)³ and SIFT.⁴ Information about known disease association can be obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC),⁵ ClinVar,⁶ and Online Mendelian Inheritance in



Illumina VariantStudio Desktop Client

Figure 1: Comprehensive Analysis and Interpretation of Data—The VariantStudio tool is a powerful software application for analyzing and interpreting variant data. This tool 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 actionable variants.

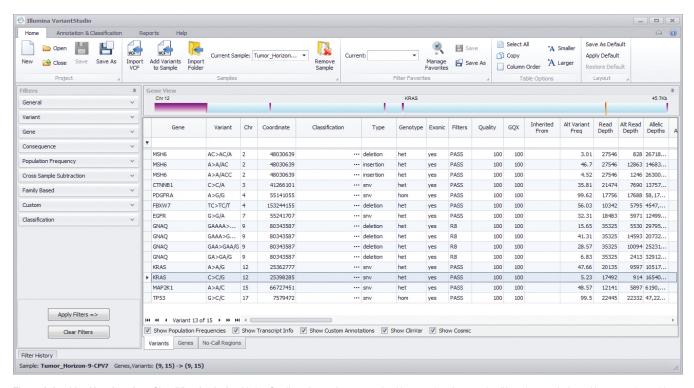


Figure 2: Intuitive User Interface Simplifies Analysis — VariantStudio software features an intuitive user interface to simplify variant analysis and interpretation, without requiring bioinformatics expertise.

Table 1: Annotation Categories

| Table 1. Annotation Categories | |
|---|---|
| Category | Description |
| Transcript consequence | Predicts transcript changes resulting from the variant of interest, allowing user 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 | Indicated whether variant has been previously associated with disease |
| VariantStudio software provides annotations at variant, transcrip | ot, and gene levels for comprehensive assessment of the biological impact of genetic variation. |

Man (OMIM),⁷ via the ClinVar database. Resources such as dbSNP,⁸ the Ensembl 1,000 Genomes Project,⁹ and Exome Variant Server¹⁰ provide information about the occurrence and frequencies of variants within a population. By delivering rapid and thorough annotation processes, VariantStudio software empowers researchers to identify biological significance in variant data.

Flexible Filtering Options

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 VariantStudio software, enabling the push-button creation of a broad sequencing portfolio. Using a cascade of filtering options, researchers can rapidly isolate the key variants that are consequential to the phenotype under study. For ease of use, VariantStudio software offers commonly applied filters, including variant quality, frequency, functional impact, and known disease association, which are easily accessible through the user interface (Figure 3). To maximize flexibility, the VariantStudio tool also enables users to filter variants based on any available associated information, including user-imported custom annotations. For further convenience, this application offers the option to construct workflows by saving concatenated filters, enabling researchers to standardize workflows and streamline analysis.

In addition to single-sample filtering, VariantStudio software enables multisample comparisons that accelerate identification of causative variants. Somatic mutations can be quickly identified in tumor-normal pairs by filtering for variants that are present in a tumor sample but absent in its normal counterpart. To support family-based analyses of inherited disease, this software application provides a collection of filters that identifies variants consistent with specified inheritance modes and patterns of disease progression. Using the VariantStudio tool, researchers can confidently isolate causative variants underlying inherited diseases and traits.

Streamlined Variant Classification

A critical component of translating genetic information into meaningful biological insight is determining the impact of identified variants within the context of observed phenotypes. The classification feature in VariantStudio software facilitates this process, enabling clinical researchers to combine their expertise with provided annotations to categorize variants. Classified variants and associated information entered by the user are saved to a local database. The classifications can then be applied easily to the same variants when they are observed in other samples (Figure 4). Researchers can enrich this database further by importing classified variants from external sources. By enabling automatic recording, tracking, and management of classified variants, VariantStudio software simplifies and accelerates data interpretation.

Customizable Reporting

Concise and effective reporting is critical for isolating actionable information from complex data. The VariantStudio application provides powerful and flexible report generation capabilities to facilitate variant reporting. Researchers can create and store multiple templates to optimize reporting for different disease areas. When a template

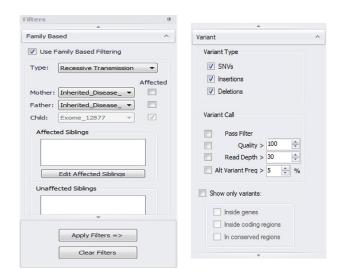


Figure 3: Comprehensive and Customizable Filters—Flexible filtering options enable users to filter variants based on inheritance mode, quality, frequency, functional impact, or custom annotations.

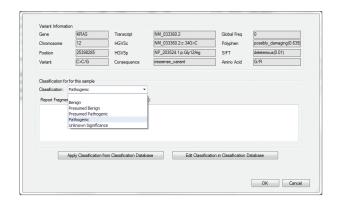


Figure 4: Information Storage Streamlines Interpretation—The VariantStudio tool stores classified variants and associated notes in a local database, making the information easily accessible for future analysis.

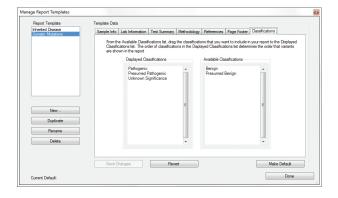


Figure 5: Templates Enable Customizable Reporting—Guided template generation and addition of sample-specific information enable customized reporting for different clinical research areas.

is applied to a sample, the user can populate the template with sample-specific information, such as detected variants in selected classification categories, along with their interpretation (Figure 5). 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 or rich-text formats for downstream use.

Software Availability

The VariantStudio standalone desktop tool is available with the purchase of TruSight® content sets¹¹ or with an Illumina supply agreement. Also, the VariantStudio software application can be accessed as a BaseSpace® App through BaseSpace Sequence Hub or BaseSpace Onsite Sequence Hub.¹² For more information, contact an Illumina representative or visit www.illumina.com/variantstudio.

Summary

VariantStudio software represents a powerful and flexible analysis method for deriving biological insight from genomic sequence information. The simple user interface delivers an intuitive framework for nonexpert users to annotate, filter, and classify variant data easily. Results can be exported in customizable reports tailored to meet the needs of various research areas. With thorough annotation and filtering capabilities, streamlined classification, and customizable reporting, VariantStudio software provides a robust method for enriching variant information with biological context.

References

- 1. uswest.ensembl.org/info/docs/variation/vep/index.html
- 2. www.ncbi.nlm.nih.gov/refseq
- 3. genetics.bwh.harvard.edu/pph2
- 4. sift.jcvi.org
- 5. cancer.sanger.ac.uk/cancergenome/projects/cosmic
- 6. www.ncbi.nlm.nih.gov/clinvar
- 7. www.ncbi.nlm.nih.gov/omim
- 8. www.ncbi.nlm.nih.gov/projects/SNP
- 9. www.1000genomes.org
- 10. evs.gs.washington.edu/EVS
- 11. www.illumina.com/trusight
- 12. www.illumina.com/basespace

Table 2: Specifications

| Feature | Specification |
|-----------------------------------|--|
| Broad range of annotation sources | Includes dbSNP, RefSeq, 1000 Genomes Project, VEP, Exome Variant Server, ClinVar, and COSMIC |
| 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 |

Table 3: Minimum System Requirements

| Parameter | Specification |
|------------------|---------------------------------|
| Operating System | Windows 7 |
| CPU | 64 bits |
| Memory | 2 GB or more (4 GB recommended) |
| Hard drive | 25 MB or larger |



