VariantStudio Data Security

Secure annotation, analysis, and storage of variant data.

Introduction

A critical component of translating genetic information into meaningful biological insight is determining the impact of identified variants within the context of observed phenotypes. VariantStudio enables this process by providing rapid annotation, intuitive filtering, and flexible classification and reporting functionalities. VariantStudio data analysis software has several measures in place to make sure that annotation and data analysis is performed securely, safeguarding data privacy.

VariantStudio is a Desktop Application

As a desktop application, VariantStudio is easily installed on a lab computer behind the institution firewall; data storage and analysis can be performed locally. Internet connection is only required for sample annotation, where a connection to the Illumina Annotation Service hosted on Amazon's Web Services (AWS) is required to retrieve annotations for each variant.

VariantStudio Data Analysis Workflow

VariantStudio software imports single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) reported in *.vcf v4.0, and later, file formats. After variants are imported, annotation of these variants can be initiated by sending a defined set of information for each variant to query the Illumina Annotation Service. Retrieved annotations are stored in the VariantStudio desktop client (Figure 1). After variants have been annotated, the added biological information can be used to filter for variants likely to be associated with the disease under study. These annotations are also used to classify variants of interest according to their biological impact. A report can then be generated in VariantStudio to summarize biologically significant and actionable results for a given sample.

AWS Security Standards

Amazon's own security processes and standards are available publicly for review[1,2]. AWS standards and accreditation include:

- **SOC 1/SSAE 16/ISAE 3402**

  The Service Organization Controls 1 (SOC 1) audit verifies that the AWS controls in place to protect customer data are designed properly and that the individual controls are operating effectively

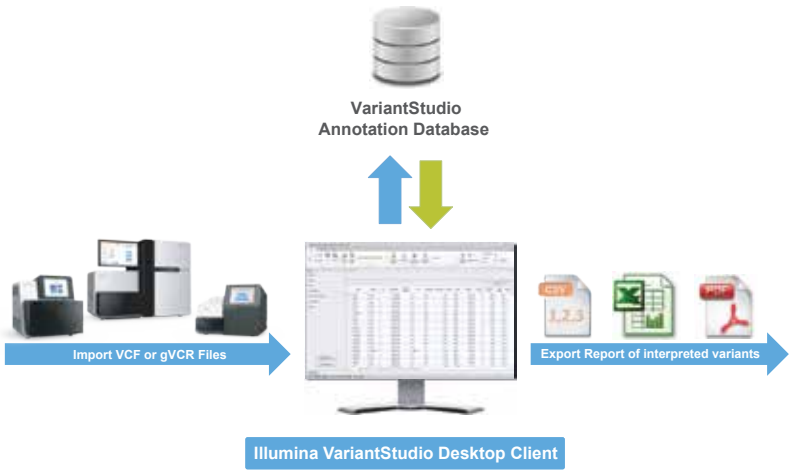
- **FISMA moderate**

  The Federal Information Security Management Act of 2002 (FISMA) recognizes the importance of information security and is an accreditation granted by the U.S. Federal Government to strengthen federal information system security. For reference, the NIH data centers are rated FISMA moderate

- **PCI DSS Level 1**

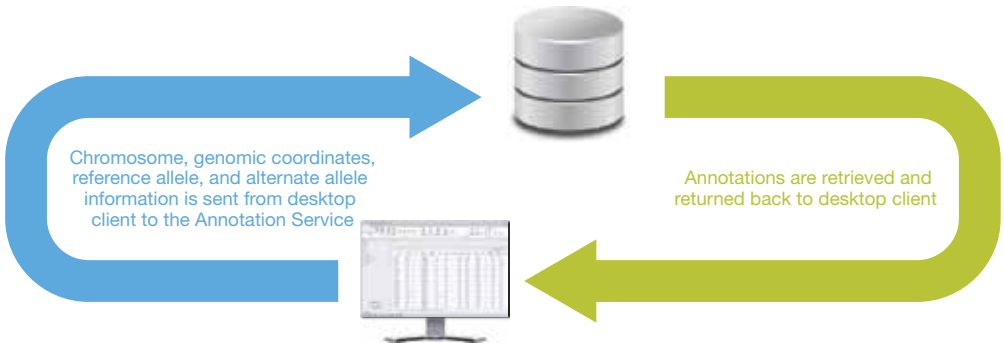
  The Payment Card Industry Data Security Standard (PCI DSS) increases electronic payment security. AWS is rated at Level 1, the highest level

Figure 1: Comprehensive Analysis and Interpretation of Variant 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. The Annotation Service is hosted on Amazon's Web Services (AWS).

Figure 2: Secure Transmission of Variant Information During Annotation



During the annotation process, limited information is sent from the desktop client to the Annotation Service (blue). Annotations are retrieved and returned back to desktop client (green).

- **ISO 27001**

  The International Organization for Standardization (ISO) sets widely recognized international security standards that specify security management best practices and comprehensive security controls

- **FIPS 140-2**

  The Federal Information Processing Standard (FIPS) Publication 140-2 is a U.S. government computer security standard that specifies the requirements for cryptography modules

Security of Data Transmission During Annotation

The variant annotation process in VariantStudio requires an internet connection to the Illumina Annotation Service hosted on Amazon's Web Services (AWS). Limited information is sent to the Annotation Service during variant annotation—the chromosome, genomic coordinates, reference allele, and alternate allele. No other information from the *.vcf file is needed for annotation purposes. After annotations are retrieved from the Annotation Service, all information is returned to the VariantStudio desktop client for storage and downstream analysis (Figure 2). VariantStudio communicates with the Annotation Service via an SSL/TLS secured channel; thus, all variant data are encrypted during transmission. Additionally, Illumina Annotation Service does not store or persist any variant data. The service log stores request time, BaseSpace™ ID, and the number of variants queried; no variant data are logged.

Learn More

To learn more about VariantStudio data analysis software, visit www.illumina.com/variantstudio.

References

1. Amazon Web Service security and compliance Information: aws.amazon.com/security/
2. Amazon Web Services: Overview of Security Processes (white paper, 2011): d36cz9buwru1tt.cloudfront.net/pdf/AWS_Security_Whitepaper.pdf

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. 0970-2014-002 Current as of 10 March 2014

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA