Exome and Genome Sequencing Data

Frequently Asked Questions

GeneDx will release exome sequence (ES) data for patients who have undergone XomeDx, XomeDxPlus, XomeDx*Xpress*, XomeDxPrenatal testing or XomeDxInsights, and will release genome sequence (GS) data for patients who have undergone GenomeSeqDx and Genome*Xpress* testing. Data is available as CRAM (compressed BAM) and VCF files with the signed consent of the patient and applicable family members. GeneDx has generated this sequence data using massively parallel (NextGen) sequencing on an Illumina platform. GeneDx has evaluated the data generated by the test for the purpose of a genetic diagnosis based on the reported clinical features. GeneDx does not provide interpretation or confirmation by an orthogonal test method for variants unrelated to the patient's clinical features, as indicated to GeneDx at the time of testing. Variants not included in the GeneDx report should be considered research results and should not be used for medical management without appropriate confirmation and interpretation by a qualified genetics provider.

Frequently Asked Questions for Data Requests

How is data provided to me?

GeneDx provides aligned sequence data as a CRAM file and VCF file. For exome or genome sequencing, data is delivered via email through a secure, HIPAA-compliant file sharing system. As an alternative for clients frequently requesting data, GeneDx can set up a sFTP account.

What is a CRAM file and what type of information does it contain?

- A CRAM file is a compressed BAM file. BAM files are the industry-standard format for storing aligned sequence data. Each CRAM file contains the sequenced data from a single sample as aligned to the human reference genome (hg19). Unaligned reads are also included. Reads have been trimmed to remove low-quality sequence at the ends of reads.
- CRAM files use reference-based compression to reduce file size. The files are aligned against and compressed using the UCSC hg19 human genome reference. The complete CRAM file specification is available here: https://samtools.github.io/hts-specs/CRAMv3.pdf

Can GeneDx provide BAM files?

No. Some software programs are directly compatible with CRAM files, such as IGV 2.4 (https://software.broadinstitute.org/software/igv/download) while others may require the conversion to BAM files. CRAM files can be converted to BAM files if necessary: http://www.htslib.org/workflow/#mapping_to_cram and http://www.ebi.ac.uk/ena/software/cram-usage#cram_to_bam.

Will GeneDx confirm any variant from my analysis of the data that I think is clinically relevant to the patient?

GeneDx can conduct targeted variant testing to confirm variants based on our standard fee schedule. Information on targeted testing can be found on our web site: http://www.genedx.com/tests/targeted-variant-testing

What is a VCF file and what type of information does it contain?

- VCF files are the industry-standard format for storing variant calls. Each VCF file contains the variants from a collection of samples, i.e. a family, with respect to the human reference genome (hg19). A variety of quality metrics are also included. VCF files are compatible with most variant annotation and interpretation software.
- The complete VCF file specification is available here: http://samtools.github.io/hts-specs/VCFv4.1.pdf

Does GeneDx provide FASTQ files or annotated data?

No. GeneDx does not provide or store FASTQ files because they are redundant with the CRAM file. All the read data in the FASTQ file, including quality scores, are also in the CRAM file. If a FASTQ file is needed, one can be re-extracted from a CRAM file using a wide variety of tools. Some methods are detailed here: http://seqanswers.com/forums/showthread.php?t=7061. GeneDx also does not provide annotated sequence data.

What software programs are available to assist me in annotation/manipulation/analysis of the sequence data?

A wide variety of open-source and commercial software is available for annotating and manipulating VCF files for ES or GS data analysis. Some popular open-source tools that are free for research use include:

- Variant Annotation, Analysis and Search Tool (http://www.yandell-lab.org/software/vaast.html)
- SnpEff (http://snpeff.sourceforge.net/)
- Ensembl Variant Effect Predictor (http://www.ensembl.org/info/docs/tools/vep/index.html)

T: (888) 729-1206 | **F:** (201) 421-2010 | zebras@genedx.com | GeneDx.com
207 Perry Parkway | Gaithersburg, MD 20877

200471 V5 08/21

Will GeneDx aid me in the evaluation of the data or provide technical support?

No. Data requests should be made by individuals who are familiar with data manipulation and analysis or who may be involved in independent research. GeneDx has already conducted a thorough analysis of the data in the context of the clinical information that was provided at the time testing was ordered. GeneDx does not provide interpretation of variants other than those determined to be related to the patient's clinical features, which were provided to GeneDx at the time of testing.

I think the data I received may be corrupt.

Before contacting GeneDx, get the MD5 checksum from the data file so that we can help to compare.

It is taking a long time to download my data. Is this normal?

- Exome and genome data are large data files (15-130 GB per individual file).
- Download speeds will depend on the requestor's internet speed.

Have the variants in the provided data been confirmed?

No. Aside from variants described in the GeneDx report, variants in the CRAM and/or VCF file are variants identified from exome or genome sequencing only, and have not been confirmed by an orthogonal test method.

Will GeneDx confirm any variant from my analysis of the data that I think is clinically relevant to the patient?

GeneDx can conduct targeted variant testing to confirm variants based on our standard fee schedule. Information on targeted testing can be found on our web site: http://www.genedx.com/tests/targeted-variant-testing

Does GeneDx release data for all family members submitted with the proband for exome or genome testing?

If relatives were submitted for exome or genome sequencing as part of the proband's testing, then GeneDx can release relative data if we receive a signed consent from the relative. GeneDx does not have data to release for those relatives' samples that were used for variant segregation analysis by targeted analysis only.

I/my child/my patient had XomeDxPlus testing (ES+mitochondrial genome sequencing and deletion testing). Can data from the mitochondrial component of this test be requested?

No. GeneDx is only able to provide ES data at this time.

Questions? Call us at 301-519-2100 or email clinicalgenomicsdata@genedx.com

What paperwork do I need to complete to request data?

Requests for CRAM and/or VCF files on the proband and/or other family members can be made by completing the appropriate consent form per person whose data is being requested. The consent form can be found on our web site: http://www.genedx.com/tests/resources#!#consent-forms

- If data is being requested for research purposes, GeneDx requires the Authorization for Use and Disclosure of Protected Health Information for Research to Third Party Researcher form be completed and signed for each individual whose data is being released to the researcher.
- It is preferred that ES or GS data be requested at the time testing is ordered to ensure data is sent within a timely manner.

What is the cost for the data?

- Data requests for exome or genome sequence data from XomeDx, XomeDxPlus, XomeDx*Xpress*, XomeDxPrenatal, XomeDxInsights, GenomeSeqDx , and Genome*Xpress* are provided at no charge. If data is being requested for research purposes, GeneDx requires the Authorization for Use and Disclosure of Protected Health Information for Research to Third Party Researcher form be completed and signed for each individual whose data is being released to the researcher.
- Data requests for the entire exome sequence data from a XomeDxSlice or Xpanded test is available for a fee; this requires a separate test request by the ordering provider.

How can I request a sFTP site for my institution?

- Contact GeneDx at clinicalgenomicsdata@genedx.com to request an account
- Have the public IP address for the machine you plan to use to access the data available

T: (888) 729-1206 | **F:** (201) 421-2010 | zebras@genedx.com | GeneDx.com
207 Perry Parkway | Gaithersburg, MD 20877