FAVOR: Functional Annotation of Variants Online Resource and Annotator for Variation across the Human Genome

Hufeng Zhou^{1,#, *}, Theodore Arapoglou^{1,#}, Xihao Li¹, Zilin Li¹, Xiuwen Zheng², Jill Moore³, Abhijith Asok⁴, Sushant Kumar^{5,6}, Elizabeth E. Blue^{7,8}, Steven Buyske⁹, Nancy Cox¹⁰, Adam Felsenfeld¹¹, Mark Gerstein^{12,13}, Eimear Kenny^{14,15,16}, Bingshan Li¹⁷, Tara Matise¹⁸, Anthony Philippakis¹⁹, Heidi Rehm^{20,21}, Heidi J. Sofia¹¹, Grace Snyder¹¹, NHGRI Genome Sequencing Program Variant Functional Annotation Working Group, Zhiping Weng³, Benjamin Neale^{20,22}, Shamil R. Sunyaev^{20,23}, and Xihong Lin^{1,20,24*}

¹ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

² Department of Biostatistics, University of Washington, Seattle, WA 98195, USA

³ Program in Bioinformatics and Integrative Biology, University of Massachusetts Chan Medical School,

Worcester, MA, USA

⁴ Microsoft Inc. Redmond, WA

⁵ Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

⁶ Princess Margaret Cancer Centre, University of Toronto, Toronto, ON, Canada

⁷ Division of Medical Genetics, University of Washington, Seattle, Washington, USA.

⁸ Brotman Baty Institute for Precision Medicine, Seattle, Washington, USA.

⁹ Department of Statistics, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

¹⁰ Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

¹¹National Human Genome Research Institute, Bethesda, DC, USA

¹² Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA

¹³ Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

¹⁴Department of Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY, USA

¹⁵Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

¹⁶Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

¹⁷Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN, USA

¹⁸ Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

¹⁹ Data Science Platform, Broad Institute of Harvard and MIT, Cambridge, MA, USA

²⁰ Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

²¹ Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

²² Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.28.505582; this version posted August 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

²³ Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

²⁴ Department of Statistics, Harvard University, Cambridge, MA, USA

[#] The two authors contributed equally.

* To whom correspondence should be addressed.

Tel: +1-617-432-2914; Fax: +1-617-432-5619

Email: hzhou@hsph.harvard.edu, xlin@hsph.harvard.edu

ABSTRACT

Large-scale whole genome sequencing (WGS) studies and biobanks are rapidly generating a multitude of coding and non-coding variants. They provide an unprecedented resource for illuminating the genetic basis of human diseases. Variant functional annotations play a critical role in WGS analysis, result interpretation, and prioritization of disease- or trait-associated causal variants. Existing functional annotation databases have limited scope to perform online queries or are unable to functionally annotate the genotype data of large WGS studies and biobanks for downstream analysis. We develop the Functional Annotation of Variants Online Resources (FAVOR) to meet these pressing needs. FAVOR provides a comprehensive online multi-faceted portal with summarization and visualization of all possible 9 billion single nucleotide variants (SNVs) across the genome, and allows for rapid variant-, gene-, and region-level online queries. It integrates variant functional information from multiple sources to describe the functional characteristics of variants and facilitates prioritizing plausible causal variants influencing human phenotypes. Furthermore, a scalable annotation tool, FAVORannotator, is provided for functionally annotating and efficiently storing the genotype and variant functional annotation data of a large-scale sequencing study in an annotated GDS file format to facilitate downstream analysis. FAVOR and FAVORannotator are available at *https://favor.genohub.org*.

INTRODUCTION

A rapidly increasing number of large-scale Whole Genome/Exome Sequencing (WGS/WES) studies and biobanks are being conducted. They provide rich opportunities for understanding the genetic bases of complex human diseases and traits. Examples of large WGS/WES studies and biobanks include the Trans-Omics Precision Medicine Program (TOPMed) of the National Heart, Lung and Blood Institute (NHLBI) (1), the Genome Sequencing Program (GSP) of the National Human Genome Research Institute (https://www.genome.gov/Funded-Programs-Projects/NHGRI-Genome-Sequencing-Program), UK biobank (2) and All of Us (3). These large WGS/WES studies and biobanks have sequenced hundreds of millions of coding and non-coding genetic variants across the human genome from hundreds of thousands of individuals, evaluating their relationship to diseases and traits.

Variant functional annotation represents a powerful and effective approach to leveraging functional information from many different bioinformatics sources to elucidate the multi-faceted functions of genetic variants for a wide range of analyses of WGS/WES studies and Genome-Wide Association Studies (GWAS) (4-14). A variety of functional annotations have been developed to measure multiple aspects of biological functionality of variants, including protein function (15-17), conservation (18,19), epigenetics (20,21), spatial genomics (22,23), network biology (24), mappability (25), local nucleotide diversity (26) and integrative composite annotations (4,27-29). These annotations have successfully prioritized plausible causal variants of underlying GWAS signals according to their functional impact in experimental studies following GWAS findings (5,30), localizing causal variants in fine-mapping studies (4,8), partitioning heritability in GWAS (6), predicting genetic risk (6,7,9), and improving rare variant (RV) analysis of WGS association studies (12-14,31). For example, large-scale WGS/WES studies (1,3,32,33) assess the associations between complex diseases/traits and coding and non-coding rare variants across the genome. The recently developed STAAR method incorporates multi-faceted variant functional annotations to boost the power of rare variant association tests in WGS/WES studies (12-14).

There is a pressing need to develop a comprehensive whole genome variant functional annotation database and browser for online queries to facilitate analysis and interpretation of GWAS and WGS/WES studies, as well as software that functionally annotates any GWAS and WGS/WES study for downstream statistical genetic analysis. Although there are several existing variant functional annotation databases, such as CADD (5,34), VEP (35), Annovar (36), WGSA (37), they have several limitations. First, these existing variant annotation databases have narrow scope. None of the databases provides overall and ancestry-specific allele frequencies from gnomAD (38) and TOPMed (1), and ClinVar information (39). Second, these resources have limited online query capabilities, and do not provide a user-friendly variant function annotation browser that summarizes and visualizes multi-faceted functional annotations of a single variant and multiple variants in a gene or a region. For example, WGSA does not provide a browser for querying variant functional annotations online. VEP provides a browser but only a few annotations. CADD allows for querying a single variant or variants in a region, but displays the annotation results in a large table that is difficult to navigate. Furthermore, most of these resources do not allow for gene- and region- level variant annotations. None of these databases provides a summary or visualization of query results. Third, there is a lack of scalable and easy-to-use tools that satisfy the need of functionally annotating large-scale WGS/WES studies. Existing functional annotation databases and tools are not scalable for functionally annotating a massive number of variants in large-scale WGS/WES studies. Moreover, few of the currently available functional annotation tools can provide organized output in a format that is both storage efficient and ready to be used in downstream statistical genetic analyses, such as fine-mapping (4,11), heritability (6), rare-variants association tests (13,14). There is a pressing community need to develop a convenient and comprehensive functional annotation tool that annotates any WGS study dataset at scale and generates a functionally annotated genotype file in an organized and compressed format, that can be readily integrated into the downstream analysis.

We developed Functional Annotation of Variants Online Resources (FAVOR), a comprehensive whole genome variant annotation database and a variant browser that provides hundreds of functional annotation scores for all possible 9 billion Single Nucleotide Variants (SNVs) and observed short insertions/deletions (indels) from a variety of biological functional aspects. FAVOR provides a fast, convenient, and user-friendly web interface that features online single variant and gene-/region-level variant queries. Search results are well-organized and visualized, according to their major functional categories. FAVOR distinguishes itself from the limitations of existing tools by providing functional annotation information that can be easily viewed through multiple functional category-based blocks and tables directly on its web interface. On top of that, FAVOR automatically generates dynamic summaries of search results by identifying important functional scores of the queried variant. These FAVOR unique features grant users immediate and intuitive insight into the search results while still maintaining users' access to the comprehensive display of multi-faceted functional scores. We have provided a comparison between FAVOR with the existing annotation databases (Supplementary Table 1).

We have also developed FAVORannotator, a tool that functionally annotates the genotype data of any WGS/WES study at scale using the FAVOR database (GRCh38 build) and stores the genotype data and their aligned functional annotation data in an annotated Genomic Data Structure (aGDS) file. The proposed aGDS data format extends the Genomics Data Structure (GDS) format (40), by storing the genotype data and the corresponding functional annotation data in a single file, making downstream integrative analysis of variants with their functional annotations more efficient and convenient. The GDS format

is highly storage-efficient, with a compression rate of a thousand times compared with the VCF format. FAVORannotator is scalable and computationally efficient for functionally annotating large-scale WGS/WES studies, for example, it completes the functional annotation of 1 billion variants in 38 CPU hours and storing those data in an aGDS file of size 488 GB.

FAVOR DATABASE

The FAVOR relational functional annotation database provides comprehensive multi-faceted variant functional annotations of all possible 9 billion SNVs in the whole genome by integrating data from multiple different sources, including CADD v1.5 (5,34), GENCODE v31(41), Annovar (36), WGSA (37), ClinVar (39), ENCODE (42), SnpEff (43), 1000 Genome (44), TOPMed Bravo Freeze 8 (1), gnomAD v3(38), and other individual studies (25,28,45-48). The preprocessing stage assigns the functional annotation values to each variant by using the variant as the primary key of the relational database. The FAVOR database is built using the PostgreSQL relational DBMS (Database Management System) for storing and retrieving variant annotation data and using a multi-table design that supports efficient integration of different types of scores. Specifically, it stores 160 functional annotation values for all possible 8,892,915,237 SNVs, and 79,997,898 observed indels from TOPMed Freeze 8 in 20 TB of space (Supplementary Table 2). These functional annotations are organized into 12 major types, including Variant Category, Allele Frequencies (AFs), ClinVar, Integrative Scores, Protein Functions, Conservation, Epigenetics, Chromatin States, Local Nucleotide Diversity, Mutation Density, Mappability and Proximity (Supplementary Table 3). The FAVOR database can be downloaded from the FAVOR website.

FAVOR ONLINE PORTAL

The online FAVOR portal facilitates fast and convenient online functional annotation query using an R shiny app (Figure 1). It allows users to search for a single variant (either in position format or rsID), multiple variants in a gene or genomic region (either in position format or gene name), or batches of tens of thousands of variants. The variant functional annotation results are displayed in tabular overviews in a summary tab (Figure 2), a full tables tab (Figure 3), and visualized using histograms (Figure 4). The FAVOR web interface is exceptionally nimble. Single Variant Search (both variant position and rsID) renders results on the webpage immediately, while Gene-based and Region-based Variant Search takes just a few seconds to display results, and Batch Annotation directly generates the annotation results for up to 10,000 variants allowing for a range of input file formats. This fast response speed is the product of its backend database indices and table design. The indices employ a diverse set of data structures, each tailored toward specific functionalities. The table design relies upon an original primary key (a combined string that consists of variant chromosome position and reference and alternative allele, e.g., 19-44908822-C-T) that efficiently relates the tables with regard to both computation and storage. This

implementation enables the fast query of 160 annotations for all 9 billion SNVs at the variant, gene and region levels.

Single Variant Search

For Single Variant Search, users can input a variant position (in hg38 build) or an rsID. The retrieved functional annotation results are displayed in three tabs: Summary, Full Table, Figures. The Summary Tab gives an overview of the biological functionality of a variant by providing the filtered annotations that flag the variant as plausibly functional, for example, Polyphen scores equal to 1 (probably_damag-ing), SIFT score equals to 0 (deleterious), or ClinVar Significance is Pathogenic, and the integrative scores greater than 10 on the PHRED scale (in the top 10% of the genome). By selecting and presenting the most informative functional annotation of a queried variant in the summary tab avoids overwhelming users with a large amount of information.

The Full Tables tab displays all functional annotation scores - organized into 17 blocks of annotation groups (Figure 4). These blocks are Basic, ClinVar, Variant Category, Overall Allele Frequencies (AFs), Ethnicity-Specific AF, Gender AF, Integrative Score, Protein Function, Conservation, Epigenetics, Transcription Factors, Chromatin States, Local Nucleotide Diversity, Mutation Density, Mapability, Proximity Table.

Different groups of functional annotation depict the variants from multiple functional perspectives. For example, ClinVar reports the relationships between genetic variants and phenotypes (39). FAVOR provides critical information from ClinVar, including Clinical Significance, Disease Name, Review Status, Disease Database ID, and Gene Reported related to the variants. Variant Category annotations provide the consequences of the genetic variants in the context of gene, categorical regulatory information, and the relative location of the variant with the closest gene (Supplementary Table 3).

FAVOR integrates multiple AFs of observed variants from multiple variant databases, including the overall AF, ethnicity-specific AF and gender-specific AF from 1000 Genome (44), TOPMed Bravo Freeze 8 (1) and gnomAD v3 (38). FAVOR provides multiple integrative scores for both coding and non-coding variants, including CADD v1.5 (5,34), LINSIGHT (49), FATHMM-XF (29), FunSeq (46), Aloft (28) and annotation Principal Components (aPCs). The aPCs summarize multiple aspects of variant function by calculating the first variant-specific PC from the individual functional annotation scores in a functional category (12). For example, aPC-conservation is the first PC of the eight individual standardized conservation scores. Furthermore, FAVOR displays category-specific individual functional annotations that represent multiple biological functionalities of each variant in a given functional category (Supplementary Table 3). For example, protein function scores describe various impact scores of the variant's damages to protein function. Conservation scores summarize the conservation functional annotation of the variants (both within and between species). Epigenetics scores summarize the signals of the open chromatin markers, close chromatin markers, and transcription markers. FAVOR also provides individual annotation scores of local nucleotide diversity, mutation density, mappability (e.g., using the unconverted genome Umap and the bisulfite-converted genome Bismap) (Supplementary Table 3). Data visualization is available using histograms in the Figures tab for a more intuitive look into functional importance (Figure. 4).

Region/Gene-based Search

For Region-/Gene-based search, users input either a gene name (official symbol), or region (starting and ending positions using the hg38 build). FAVOR will instantaneously output the functional annotation summary results of the variants in the gene or the region, as well as variant-specific annotations in a range of annotation categories. The fast display of the retrieved results of the Region/Gene-based Search is enabled through indexing and efficient multi-table database management.

The Region-/Gene-based Search summary tab provides the summary statistics of the variants in a region or a gene using several key summary tables and histograms, including Allele Frequency Distribution, GENCODE Category, ClinVar Clinical Significance, Functional Consequences and High Integrative Functional Scores (Figure 5). In the Region/Gene-based individual variant annotation table, 32 commonly used annotations (Supplementary Table 4) are displayed for each variant in the gene/region. The variants can be sorted by their values in any column. It also has a convenient search feature that allows users to filter the variants in the region/gene based on specified features and key words. For example, typing "pathogenic" in the search box above the displaying table provides only the pathogenic variants of the region/gene.

Batch Annotation

Batch Annotation provides functional annotations of a list of variants submitted by users in a file. It supports multiple file formats as input, including CSV, TSV, VCF, XLS, and RDS. Multiple formats and IDs of variants are also supported. For example, each row of a text file can specify a variant's chromosome, position, reference, and alternative allele value (e.g., 1-10253-CTA-C), or a variant's chromosome and position values (e.g., 1-10253), or an rsID (rs868413313). Users can upload the variants list using the above file formats on the FAVOR batch annotation page. Batch annotation files are currently limited to 10,000 variants in the interest of online wait time. It takes less than less than 1 minute to annotate 1,000

variants. The annotation results containing 160 annotations of the variants in the submitted variant list are available for download. FAVORannotator, discussed below, can be used to handle functional annotations of a larger number of variants, e.g., hundreds of millions of variants in a WGS/WES study.

ANNOTATED GENOMIC DATA STRUCTURE (aGDS)

Variant Call Format (VCF) (50) has been frequently used for storing variant call data of sequencing studies. However, VCF is text-based and thus inefficient with regard to storage, particularly for large-scale WGS data of hundreds of thousands to millions of subjects that have hundreds of millions to billions of variants. The recently developed Genomic Data Structure (GDS) format (40) provides a storage-efficient format to store WGS data. The GDS format has a compression rate of 1000 times compared to the VCF format. However, it does not incorporate variant functional annotations. We developed the annotated Genomic Data Structure (aGDS) format (Figure 6), that extends the GDS format by integrating both genotypes in a WGS study and variant functional annotations in a single file. There are three main advantages of the aGDS format. First, it provides fast query and simultaneous retrieval of genotype and matched functional annotation data defined by flexible filtering criteria. Second, it is convenient to integrate an aGDS file into functionally informed downstream analysis pipelines, such as STAARpipeline for rare variant association analysis (13). Third, it is also highly storage-efficient for genotype and functional annotation data. An aGDS file containing TOPMed Freeze 8 WGS data, including both genotype and functional annotations of 140,306 samples, only takes 487 GB, that is three orders of magnitude smaller compared to the VCF files (Supplementary Table 2).

FAVORANNOTATOR

FAVORannotator is an open-source tool that uses the FAVOR database to functionally annotate and efficiently store genotype and variant functional annotation data of a WGS/WES study in an aGDS file, and to facilitate downstream association analysis (Figure 7). FAVORannotator only requires genotype data or a variant list as input and automatically annotates the genotype data or the variant list, generating an aGDS file as an output. The former facilitates rare variant association analysis using individual level data, while the latter facilitates rare variant meta-analysis using summary statistics, by incorporating functional annotations, e.g., using STAAR (13).

Time and memory resources for annotating a large number of variants using FAVORannotator are very attractive, especially for large-scale WGS/WES datasets, such as TOPMed, GSP and UK Biobank. For example, FAVORannotator produces an annotated genotype file in the aGDS format for n=180,000 whole genome samples with 900 million variants of the TOPMed Freeze 10a WGS data in 38 hours, and for n=60545 whole genome samples of 450 million variants of GSP-CCDG Freeze 2 WGS data within 30

CPU hours. FAVORannotator has also been implemented as a workflow in the cloud-based platforms, including DNAnexus (UK Biobank), AnVIL (NHGRI) and BioData Catalyst (NHLBI) (Figure 8) (51). FAVORannotator's efficiency keeps cloud computing costs low, for example, costing ~\$25 to annotate the TOPMed Freeze 10a WGS data by chromosome in parallel, e.g., 3 CPU hours for chromosome 1.

DISCUSSION

FAVOR offers a comprehensive solution for the application of whole genome variant functional annotations, including an open access and downloadable database, a user-friendly browser, and a tool FAVORannotator, to annotate large-scale genetic data. The FAVOR database is a large relational data structure of multi-faceted functional annotations of all possible 8,812,917,339 SNVs and 79,997,898 observed indels in the human genome. It is built using a storage-efficient postgreSQL database with indexed and relational tables, that provide fast query speeds. The FAVOR web interface provides fast variant, gene, region level online multi-faceted functional annotations, as well as batch annotation. It emphasizes responsiveness while providing dynamic display and visualization features, and uses combined approaches, including visualizations, block organizations by categories, and convenient search and sorting functions, to provide a fast and convenient summary of the major functional impact of variants.

The FAVOR annotator software enables researchers to use the FAVOR database to efficiently functionally annotate a genetic study at scale, such as GWAS, WGS and WES studies, and build a highly compressed and well organized aGDS file, that includes both genotype data and their annotations and can be easily integrated into downstream analysis pipelines, Together, FAVOR and FAVORannotator provide a critical infrastructure for facilitating downstream analysis and interpretation of GWAS/WGS/WES studies. Although several compression methods are available for storing WGS data, such as gzip (vcf.gz), Bgzip or BCF(52), they are subject to two major limitations. First, they are not efficient for storing large-scale WGS data. Second, they are difficult to read while compressed. For instance, although the BCF format is more storage-efficient than the VCF format, the compression rate is 100 times. In contrast, the GDS format has a compression rate of 1000 times. Furthermore, both VCF and BCF formats only store genotype data and do not store variant annotations. The aGDS format resolves both limitations successfully.

In summary, FAVOR and FAVORannotator provide an intuitive and indispensable infrastructure for facilitating downstream analysis and result interpretation of large-scale WES/WGS studies. FAVOR currently provides non-tissue specific epigenetic functional annotations for non-coding variants. It is of future interest to integrate tissue and cell-type specific epigenetic functional annotations in FAVOR. As functional annotations continue to grow in depth and breadth, we will continue to improve and expand FAVOR by integrating more and state-of-art annotations and supporting more analytical scenarios.

FUNDING

This work was supported by grant nos. R35-CA197449, P01-CA134294, U19-CA203654 and R01-HL113338 (to X. Lin), U01-HG012064 (to Z. Weng and X. Lin), U01-HG009088 (to X. Lin, S.R.S. and B.M.N.).

AVAILABILITY

FAVORannotator is an open-source annotation tool available in the GitHub repository (https://github.com/zhouhufeng/FAVORannotator)

The FAVOR essential database (containing 20 essential functional annotation scores) for all possible SNVs (8,812,917,339) and observed Indels (79,997,898) in Build GRCh38/hg38 is hosted on Harvard Dataverse (https://doi.org/10.7910/DVN/1VGTJI).

The FAVOR full database (containing 160 essential functional annotation scores) for all possible SNVs (8,812,917,339) and observed Indels (79,997,898) in Build GRCh38/hg38 is hosted on Harvard Dataverse (https://doi.org/10.7910/DVN/KFUBKG).

CONFLICT OF INTEREST

B.M.N. is on the Scientific Advisory Board of Deep Genomics, a consultant for Camp4 Therapeutics, Takeda Pharmaceutical and Biogen. S.R.S. is consultant to NGM Biopharmaceuticals and Inari agriculture. He is also on Scientific Advisory Board of Veritas Genetics. G.R.A. is an employee of Regeneron Pharmaceuticals and owns stock and stock options for Regeneron Pharmaceuticals. X. L. is a consultant of AbbVie Pharmaceuticals. Z. W. co-founded and serves as a scientific advisor for Rgenta Inc. A.P. is a Venture Partner at GV, a subsidiary of Alphabet corporation. He has received funding from Verily, MSFT, Intel, IBM, Bayer, Novartis, Pfizer, Biogen, Abbvie.

Reference

- Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M. *et al.* (2021) Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, **590**, 290-299.
- Halldorsson, B.V., Eggertsson, H.P., Moore, K.H.S., Hauswedell, H., Eiriksson, O., Ulfarsson, M.O., Palsson, G., Hardarson, M.T., Oddsson, A., Jensson, B.O. *et al.* (2022) The sequences of 150,119 genomes in the UK Biobank. *Nature*.
- 3. Investigators, A.o.U.R.P. (2019) The "All of Us" research program. *New England Journal of Medicine*, **381**, 668-676.
- 4. Kichaev, G., Yang, W.Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P. and Pasaniuc, B. (2014) Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet*, **10**, e1004722.

- 5. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. and Shendure, J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, **46**, 310-315.
- 6. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K. *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*, **47**, 1228-1235.
- 7. Hu, Y., Lu, Q., Powles, R., Yao, X., Yang, C., Fang, F., Xu, X. and Zhao, H. (2017) Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput Biol*, **13**, e1005589.
- Kichaev, G., Roytman, M., Johnson, R., Eskin, E., Lindstrom, S., Kraft, P. and Pasaniuc, B. (2017) Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics*, 33, 248-255.
- 9. Morrison, A.C., Huang, Z., Yu, B., Metcalf, G., Liu, X., Ballantyne, C., Coresh, J., Yu, F., Muzny, D., Feofanova, E. *et al.* (2017) Practical Approaches for Whole-Genome Sequence Analysis of Heart- and Blood-Related Traits. *Am J Hum Genet*, **100**, 205-215.
- 10. Lee, P.H., Lee, C., Li, X., Wee, B., Dwivedi, T. and Daly, M. (2018) Principles and methods of in-silico prioritization of non-coding regulatory variants. *Hum Genet*, **137**, 15-30.
- 11. Schaid, D.J., Chen, W. and Larson, N.B. (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*, **19**, 491-504.
- 12. Li, X., Li, Z., Zhou, H., Gaynor, S.M., Liu, Y., Chen, H., Sun, R., Dey, R., Arnett, D.K., Aslibekyan, S. *et al.* (2020) Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat Genet*, **52**, 969-983.
- Li, Z., Li, X., Zhou, H., Gaynor, S.M., Selvaraj, M.S., Arapoglou, T., Quick, C., Liu, Y., Chen, H., Sun, R. *et al.* (2021) A framework for detecting noncoding rare variant associations of large-scale whole-genome sequencing studies. *bioRxiv*, 2021.2011.2005.467531.
- 14. Li, X., Yung, G., Zhou, H., Sun, R., Li, Z., Hou, K., Zhang, M.J., Liu, Y., Arapoglou, T., Wang, C. *et al.* (2022) A multi-dimensional integrative scoring framework for predicting functional variants in the human genome. *Am J Hum Genet*, **109**, 446-456.
- 15. Ng, P.C. and Henikoff, S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, **31**, 3812-3814.
- 16. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat Methods*, **7**, 248-249.
- Adzhubei, I., Jordan, D.M. and Sunyaev, S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*, Chapter 7, Unit7 20.
- 18. Cooper, G.M., Goode, D.L., Ng, S.B., Sidow, A., Bamshad, M.J., Shendure, J. and Nickerson, D.A. (2010) Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods*, **7**, 250-251.
- Goode, D.L., Cooper, G.M., Schmutz, J., Dickson, M., Gonzales, E., Tsai, M., Karra, K., Davydov, E., Batzoglou, S., Myers, R.M. *et al.* (2010) Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. *Genome Res*, 20, 301-310.
- 20. Skipper, M., Eccleston, A., Gray, N., Heemels, T., Le Bot, N., Marte, B. and Weiss, U. (2015) Presenting the epigenome roadmap. *Nature*, **518**, 313.

- Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K. *et al.* (2018) The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*, 46, D794-D801.
- 22. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665-1680.
- 23. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B. *et al.* (2015) CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, **163**, 1611-1627.
- 24. Yu, H., Kim, P.M., Sprecher, E., Trifonov, V. and Gerstein, M. (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol*, **3**, e59.
- 25. Karimzadeh, M., Ernst, C., Kundaje, A. and Hoffman, M.M. (2018) Umap and Bismap: quantifying genome and methylome mappability. *Nucleic Acids Res*, **46**, e120.
- Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A. *et al.* (2017) Linkage disequilibriumdependent architecture of human complex traits shows action of negative selection. *Nat Genet*, 49, 1421-1427.
- 27. Coordinators, N.R. (2014) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, **42**, D7-17.
- 28. Balasubramanian, S., Fu, Y., Pawashe, M., McGillivray, P., Jin, M., Liu, J., Karczewski, K.J., MacArthur, D.G. and Gerstein, M. (2017) Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nat Commun*, **8**, 382.
- 29. Rogers, M.F., Shihab, H.A., Mort, M., Cooper, D.N., Gaunt, T.R. and Campbell, C. (2018) FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*, **34**, 511-513.
- Ionita-Laza, I., McCallum, K., Xu, B. and Buxbaum, J.D. (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat Genet*, 48, 214-220.
- 31. Quick, C., Wen, X., Abecasis, G., Boehnke, M. and Kang, H.M. (2020) Integrating comprehensive functional annotations to boost power and accuracy in gene-based association analysis. *PLoS Genet*, **16**, e1009060.
- 32. Ewan Birney, R.C., Andy Clark, Daniele Fallin, Jonathan Haines, Monica Justice, Rod McInnes, Len Pennacchio. NHGRI Genome Sequencing Program, pp. <u>https://www.genome.gov/Funded-Programs-Projects/NHGRI-Genome-Sequencing-Program</u>.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M. *et al.* (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, 12, e1001779.
- 34. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. and Kircher, M. (2019) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*, **47**, D886-D894.
- 35. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F. (2016) The Ensembl Variant Effect Predictor. *Genome Biol*, **17**, 122.
- 36. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, **38**, e164.

- 37. Liu, X., White, S., Peng, B., Johnson, A.D., Brody, J.A., Li, A.H., Huang, Z., Carroll, A., Wei, P., Gibbs, R. *et al.* (2016) WGSA: an annotation pipeline for human genome sequencing studies. *J Med Genet*, **53**, 111-112.
- 38. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434-443.
- 39. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. and Maglott, D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*, **42**, D980-985.
- 40. Zheng, X., Gogarten, S.M., Lawrence, M., Stilp, A., Conomos, M.P., Weir, B.S., Laurie, C. and Levine, D. (2017) SeqArray-a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics*, **33**, 2251-2257.
- 41. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S. *et al.* (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*, **22**, 1760-1774.
- 42. Consortium, E.P., Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A. *et al.* (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, **583**, 699-710.
- 43. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*, **6**, 80-92.
- 44. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68-74.
- 45. Consortium, F., the, R.P., Clst, Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462-470.
- 46. Fu, Y., Liu, Z., Lou, S., Bedford, J., Mu, X.J., Yip, K.Y., Khurana, E. and Gerstein, M. (2014) FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol*, **15**, 480.
- 47. Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H. and Kasukawa, T. (2017) FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Sci Data*, **4**, 170107.
- 48. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M. *et al.* (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*, **2017**.
- 49. Huang, Y.F., Gulko, B. and Siepel, A. (2017) Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat Genet*, **49**, 618-624.
- 50. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.
- 51. Schatz, M.C., Philippakis, A.A., Afgan, E., Banks, E., Carey, V.J., Carroll, R.J., Culotti, A., Ellrott, K., Goecks, J., Grossman, R.L. *et al.* (2022) Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space. *Cell Genom*, **2**.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.28.505582; this version posted August 29, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

52. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. *et al.* (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, **10**.

Figures

Figure 1. FAVOR web interface: This online portal provides a convenient web interface allowing for variant, gene, and region-level variant annotation queries. The home page displays the supported query methods, and examples of expected input.

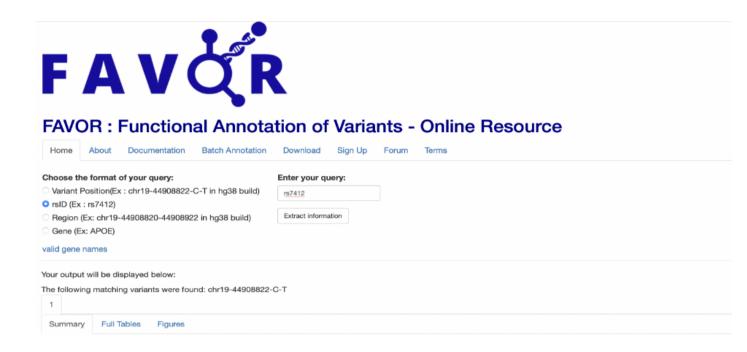


Figure 2. Single Variant Query Summary page: The Single Variant Query Summary page shows a dynamic overview of the filtered annotations with evidence for plausible functional consequences. For example, the annotations are displayed if Polyphen scores equal to 1 (probably_damaging), SIFT score equals to 0 (deleterious), ClinVar Significance is Pathogenic, and the integrative scores that are greater than 10 on the PHRED scale.

Basic	
Variant	chr19- 44908822- C-T
rsID	rs7412
TOPMed QC Status	PASS
TOPMed Bravo AF	0.0821388
GNOMAD Total AF	0.0788183
ALL 1000G AF	0.07508

Integrative

Score	PHRED	Percentile
aPC-Protein- Function	39.81	0.01
aPC- Conservation	15.81	2.62
aPC- Epigenetics	10.85	8.22
aPC- Transcription- Factor	13.17	4.82
aPC- Mappability	14.58	3.48
aPC- Proximity-To- TSS-TES	19.00	1.26
CADD PHRED	26.30	0.23

Epigenetics

Open	aPC- Epigenetics	10.85
Open	DNase	0.47
Open	H3K4me2	3.80
Open	H3K4me3	6.45
Open	H3K9ac	7.14
Open	H4K20me1	4.91
Open	H2AFZ	7.76
Closed	H3K27me3	4.98
Transcription	totalRNA	16.31

Gencode APOE Gencode APOE Comprehensive exonic Info exonic Gencode exonic Category SNV Category Category CAGE YES

Protein

Promoter

aPC-Protein- Function	39.81
PolyPhenCat	probably_damaging
PolyPhenVal	1
Polyphen2 HDIV	1.0
Polyphen2 HVAR	1.0
Grantham	180
MutationTaster	0.930
SIFTcat	deleterious
SIFTval	0

ClinVar Clinical drug_response Significance

orginitearies	
Clinical Significance (genotype includes)	441262:Pathogenic, 441265:Pathogenic, 441266:Pathogenic, 666796:Uncertain_significance
Disease Name	Familial_type_3_hyperlipoproteinemia, Warfarin_response, atorvastatin_responseEfficacy, not_provided
Disease Name (included variant)	Apolipoproteinemia_E1, Familial_type_3_hyperlipoproteinemia, not_specified

Conservation

aPC- Conservation	15.81
mamPhCons	1.00
priPhyloP	0.42
GerpN	13.40

Figure 3. Single Variant Query Functional Annotation Tabulation: The Full Tables tab in the FAVOR Single Variant Online Query organizes functional annotation results in blocks defined by annotation types .

Basic		ClinVar		Variant Category		
Overall AF Eth		nnicity AF	AF Gender AF I		Gender AF Female	A Constant of the second secon
Integrative Sco	re	Protein	Protein Function Conservation		LILLI LILILLI LILI	
Epigenetics		Transcripti	ion Factors	Cł	romatin States	
Local Nucleotide Diversity			Mappabi	lity	Proximity Table	a and a definition of the second seco

Figure 4. Single Variant Query Functional Annotation Visualization: The Figures tab in the FAVOR Single Variant Online Query displays a visualization of the functional annotation results of a queried variant in histogram.

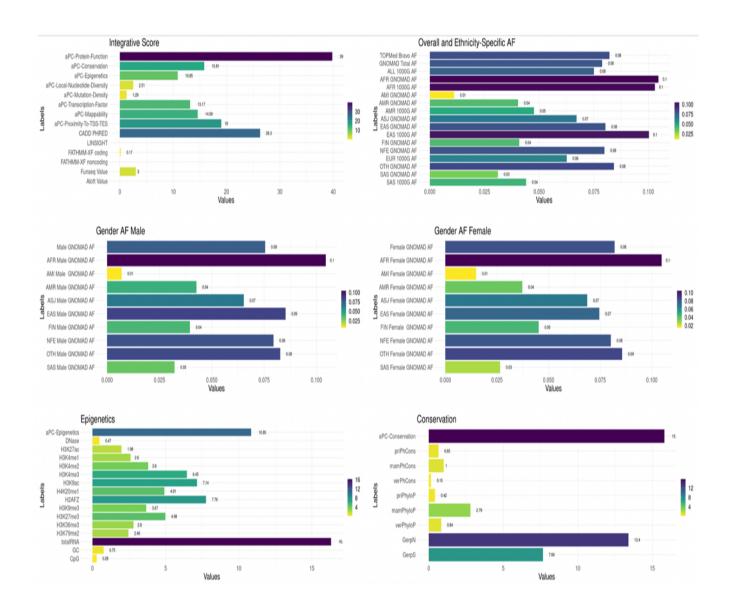


Figure 5. Region-/Gene-based Query Summary tab: The Summary tab of the Region-/Gene-based Query shows the multi-faceted functional annotation summary statistics of the variants in a gene or a region.

Allele Count Distribution Variant Cate Total

Allele Count Distribution

u 1475 1.0000 mmon (MAF > 0.01) 451 0.0006 Progenori (0.001 ~ MAF < 0.01) 12428 0.4447 < 0.001 12428 0.4459 piona 6060 0.4459 piona 6060 0.4459 piona 724 0.0015 atta Clasgoria Courta Proportions mar Clasgoria Courta Proportions 106 Conservation 106 0.0124 Courta Multicol Conservation 106 0.0124 Courta Multicol Conservation 106 0.0451 Conservation 106 0.0451 Conser	al 14775 1.0000 mmon (MAF > 0.01) 451 0.0036 mediamery (DOD) <= MAF < 0.01) 1940 0.1255 regamery (DOD) <= MAF < 0.01) 1426 0.4459 ubit constraints 2258 0.1528 const Category 1122 0.0016 rine (DAF < 0.01) 12426 0.4459 ubit const 749 0.0016 rine (Datagories Courts Propertions on a 13358 0.9941 rine (Datagories Courts Propertions on a 9 0.00000 0.0000 rine (Datagories Courts Propertions on a 9 0.00000 0.0000 rine (Datagories Courts Propertions on a 9 0.00000 0.0000 ring (Responts 0 0 0.0000 0.00000 rine (Datagories Courts Propertions 0 0.0000 0.00000 ring (Responts 0 120 0.00001 0.00000 ring (Responts 0 120 0.00001 0.00000 ring (Responts 0 120 0.00001 0.00001 ring (Responts 0 120 0.00001 0.00001<	llele Count Distribution		
mmm (MAF >= 0.01) 451 0.0305 F Frequency (0.001 == MAF < 0.011	all 14775 1.0000 wirequexy D.0001 <= MJK < 0.001	ariant Categories	(Counts Proporti
Immem MAT O.0.005 Pregnancy Outor Outor Science Outor </td <td>mmmin MMP = 001) 45 0.000 mmmin MMP = 001) 125 0.000 me MAP < 0.0001)</td> 12426 0.410 genome 0.005 0.4100 genome 0.005 0.4100 genome 0.0001 12426 0.4100 genome 0.0001 12426 0.4100 genome 0.0001 10.1528 0.0001 main 2122 0.0003 0.0000 0.0000 main 212 0.00145 0.0000 0.0000 main 212 0.0003 0.0000 0.0000 main 214 0.0145 0.0000 0.0000 main 214 0.0158 0.0000 0.0000 mign 0.0002 0.0000 0.0000 0.0000 mign 0.0015 0.0001 0.0000 0.0000 mign 0.0016 0.00001 0.0000 0.0000 0.0000 mign 0.0017 0.00001 0.0001 0.0001 0.0001 0.0001 ming 0.00012 Thing 0.0012 0.0001	mmmin MMP = 001) 45 0.000 mmmin MMP = 001) 125 0.000 me MAP < 0.0001)			
Initia 00.00 0.1400 et al. 225 0.1528 de Categories Counts Proportions i 744 0.0510 i 7343 0.0215 i 0.1538 0.0041 etaram 122 0.0008 etaram 122 0.0008 etaram 122 0.0008 etaram 223 0.0145 genes 7 0.0009 pelogonic 5 0.0400 Bengonia 665 0.0400 Bengonia 666 0.0234 ym 223 0.0001 rotargeories 102 0.0000 rotargeories 102 0.0001 rotargeories 102 0.0001 rotargeories 102 0.0001 rotargeories 102 0.0001 rotargeories 103 0.0002 rotargeories 103 0.0002 rotargeories 103 0.0002 rotargeories 103 0.0001 rotargeories 103 0.0002 rotargeories 103 0.0002 rotargeories 104 0.0012 rotargeo	$ \frac{\log a}{\log p} = \frac{\log a}{\log p} =$	non (MAF >= 0.01)		451 0.0
janita 0000 0.4400 code 0.4400 0.1528 code 7.24 0.0510 sind Categories Counts Propertions reid 7.24 0.0510 sind Sageories Counts Propertions reid 0.0145 Counts sind Categories Counts Propertions reid 0.0145 Counts sind Categories Counts Propertions reid Categories Counts Properinon reid Categories	approxima 2258 0.1528 code 0.4486 0.4486 code 0.4486 0.1528 code 0.4166 0.1528 code 0.4167 0.1528 code 0.0016 0.1528 code 0.0016 0.0016 atic Clasports 0.0016 0.0000 code 0.0000 0.0000 code 0.0000 0.0000 rise Clasports 0.0001 0.0001 rise Clasports <t< td=""><td>v Frequency (0.0001 <</td><td>MAF < 0.01)</td><td>1898 0.1</td></t<>	v Frequency (0.0001 <	MAF < 0.01)	1898 0.1
initian 0x1000 0x1000 0x1000 code 0.1528 0.1528 0 code Code 0x1000 0	appendix appendix bit mode bit mode concertance 2258 0.1528 Concertance concertance Concertance Concertance Concertance concertance 0.0000 Concertance Concertance concertance 0.0000 Concertance Concertance concertance 0.0000 Concertance Concertance concertance Concertance Concertance Concertance Concertance concertance Concertance Concertance Concertance Concertance concertance Concertance Concertance Concertance Concertance concertance Concertance Concertance Concertance Concertance conconcertance Concertance			
coda Coda Proportions 1 3385 Coda	coda Catagory ama Catagoria Properties rain Catagoria 744 0.0015 rain Catagoria 0.0005 vex.rumani 122 0.0005 vex.rumani 124 0.0145 genegoria 67 0.0005 rain Catagoria 0 0.0005 regresoria 67 0.0005 regresoria 63 0.0005 conservation 103 0.0005 conservation 105 0.0005 conse	-		
San Clargory Courts Properties ric 1338 0 0,001 3 1 122 0,003 rer. Clicical Significane lata Clargory 2 0 0,000 rer. Clicical Significane lata Clargory 2 0 0,000 regeneration 1 12 0,000 regeneration 1 13 0,000 ref. Clicical Significane lata Clargory 2 0 0,000 ref. Clicical Significane lata Clargory 2 1 0,000 ref. Clicical Significane ref. Clicical Significane ref.	Code Largery Code Largery Code Largery mic 1 Categorie Counts Proportions mic 1 212 0.0055 wintersem 122 0.0055 wintersem 123 0.0056 wintersem 123 0.0055 wintersem 123 0.0055 wintersem 123 0.0055 wintersem 123 0.0056 wintersem 123 0.0051 wintersem 124 0.0051 wintersem 126 0.0051 wintersem 126 0.0051 wintersem 126 0.0051 wintersem 126 0.0051 </td <td>Joubletons</td> <td></td> <td>2250 0.1</td>	Joubletons		2250 0.1
San Clargory and Clargory by Bangan and Clargory and Clargory by Bangan and Clargory by Bangan and Clargory by Bangan and Clargory by Clargory	Constantion Propertion onic 744 0.0010 R 0.0010 R 0.0010 R 0.0010 R 0.0000 Wintersam 122 0.0000 Wintersam 123 0.0000 Wintersam 133 0.0000 Wintersam 143 0.0001 Mintersame 143 0.00011 Mintersa			
aint Categories Courts Proportions reice 734 0.0010 aint Categories Counts Proportions pressorate 0.0000 Cinvac: Cinvac: Cinvac: Significance aint Categories Counts Pressorate pressorate 0.00000 Cinvac: Cinvac: Cinvac: Significance aint Categories Counts Pressorate Cinvac: Cinvac: Cinvac: Significance Cinvacinclonal Score Categories <td>Carlos ordes Counts Proportions rsi 318 0.0015 rsi 318 0.0015 rsite 122 0.0085 statem 214 0.0045 statem 214 0.0055 statem 213 0.0055 statem 213 0.0055 statem 212 0.0065 statem 214 0.0212</td> <td>encode Category</td> <td></td> <td></td>	Carlos ordes Counts Proportions rsi 318 0.0015 rsi 318 0.0015 rsite 122 0.0085 statem 214 0.0045 statem 214 0.0055 statem 213 0.0055 statem 213 0.0055 statem 212 0.0065 statem 214 0.0212	encode Category		
nic 724 0.0310 nic 1336 0.0215 nic 1336 0.0041 wistream 124 0.0015 cing 9 0.0005 wistream 214 0.0145 cing 9 0.0005 wistream 0.0005 Cinex: Clinical Significance and Clategories Courts Preportions g Response 0 0.0005 king Clategories 0.0005 Cinex: Clinical Significance and Clategories Courts Preportions ping night 346 0.0224 citical Clategories Courts Preportions ping nonymous 200 0.0111 preportiers 130 0.0007 ping nonymous 200 0.0111 preportiers 130 0.0007 ping nonymous 200 0.0112 preportiers 130 0.0007 ping nonymous 200 0.0112 preportiers Courts Proportiers Dident Function 140	oric 7.54 0.0510 IR 318 0.0015 IR 318 0.0015 IR 0.0005 0.0005 Varianti 122 0.0005 0.0005 Varianti 122 0.0005 0.0005 Varianti 122 0.0005 0.0005 Varianti 123 0.0005 0.0003 Major Parlogenic 0 0.0005 Varianti 123 0.0003 0.0003 Major Parlogenic 0 0.0003 Major Parlogenics 0.0003 0.00012 Mator Categories 0.00012 Major Parlogenics 0.00012 Major Parlogenics 0.00012 Major Parlogenics 0.00012 Major Parlogenics 0.00012 Major Parlogenics 0.00012 Major	Variant Categories Cou	nts Proportion	15
windtream 12 0.0003 theam 214 0.0145 cing 9 0.0006 tart Calegories Counts Proportions g Response 0 0.0000 hy Panoponic 5 0.0000 citical Consequences Counts Proportions tant Categories Counts Proportions conservation 165 0.0112 tant Categories Counts Proportions conservation 166 0.0112 tant Categories Counts Proportions conservation 166 0.0112 tant Categories Counts Proportions conservation 166 0.0112 conservation	windtham 121 0.0003 atheam 214 0.0145 atheam 214 0.0145 atheam 214 0.0145 atheam 214 0.0145 atheam 214 0.0156 var. Chical Significance Cinvar. Chical Significance atheam 605 0.0000 https://min 605 0.0400 ay Benjoni 605 0.0400 ay Benjoni 133 0.0000 citical Categories Consta Frequencies attact Categories 0.0301 Constance Frequencies Consequences 143 0.0001 Frequencies Consequences attact Categories Consequences 165 0.0114 Frequencies Consequences Consequences 162 0.0001 Frequencies Consequences Consequences Notegories Consequences 163 0.0124 Consequences Consequences Consequences 166 0.0136 <td< td=""><td></td><td></td><td></td></td<>			
windtream 12 0.0003 theam 214 0.0145 cing 9 0.0006 tart Calegories Counts Proportions g Response 0 0.0000 hy Panoponic 5 0.0000 citical Consequences Counts Proportions tant Categories Counts Proportions conservation 165 0.0112 tant Categories Counts Proportions conservation 166 0.0112 tant Categories Counts Proportions conservation 166 0.0112 tant Categories Counts Proportions conservation 166 0.0112 conservation	windtham 121 0.0033 atheam 214 0.0145 atheam 214 0.0145 atheam 214 0.0006 var. Chical Significance Cinvar. Clinical Significance riatal Categories Counts Preportions ig Response 0 0.0000 hyperporties 7 0.0003 atheam 46 0.0234 varcom 233 0.0005 ristic Categories Counts Preportions officing Interpretations 133 0.0000 ristic Categories Counts Proportions OFF 75 0.0001 manymorymous 502 0.0330 OFF 75 0.0001 Intel Categories Counts Proportions OFF 75 0.0001 Intel Categories 102 0.0002 VGE Enhancers 118 0.0012 Intel Categories Counts Proportions VGE Enhancers 155 0.0033 VG Denotin Function Kees <	UTR	318 0.02	15
windtream 12 0.0003 theam 214 0.0145 cing 9 0.0006 tart Calegories Counts Proportions g Response 0 0.0000 hy Panoponic 5 0.0000 citical Consequences Counts Proportions tant Categories Counts Proportions conservation 165 0.0112 tant Categories Counts Proportions conservation 166 0.0112 tant Categories Counts Proportions conservation 166 0.0112 tant Categories Counts Proportions conservation 166 0.0112 conservation	windtham 121 0.003 atheam 214 0.0145 kining 9 0.0006 var. Chical Significance Chical Categories Count riat Categories Counts Proportions ig Response 0.0005 Counts var. Chical Significance Counts Proportions var. Significance Counts<			
Interaction 214 0.0145 cing 9 0.0000 arr. Clinical Significance Cluva: Proportions g Response 0 0.0000 hy Parlogenic 5 0.0003 signi encome 0.0000 Proportions g Response 0 0.0000 hy Parlogenic 5 0.0003 signi encome 0.0000 Proportions g Response Counts Proportions p Response Counts Proportions p Response 143 0.0000 number Missense 143 0.0001 p Proportions 0.0112 Oncome p Proportions 0.0112 Oncome p Proportions 102 0.00009 GE Proneters 102 0.00009 GE Proneters 102 0.00009 GE Proneters 102 0.0006 Colorevariation 1661 0.1124 Colorevariation 1661 0.1124 Colorevariation<	stream 214 0.0145 scing 9 0.0006 var: Clinical Significance Clinical Significance Clinical Significance static Classoprie 0.0000 0.0000 thouganic 5 0.0003 sky Patropenic 5 0.0003 sky Patropenic 5 0.0003 sky Patropenic 5 0.0000 cicloal Consequences Courts Proportions origin 346 0.0007 major 0.0000 0.0112 sky Patropenics Courts Proportions origin 143 0.0007 monymous 269 0.0112 sky Phonotons 18 0.0012 httegrative Functional Socie Category >= 10 Categories Courts relat Categories Courts Proportions C2 Court Aucteotide Diversity 0.057 0.5453 C3 Court Nacteotide Diversity 0.500 0.7722 C4 Cataent brance function 4496 0.0336 <t< td=""><td></td><td></td><td></td></t<>			
cing 9 0.0008 er: Cincial Significance Cinver: Cincial Significance faita Catagories Counts Proportions gispone 5 0.0003 hyle Pathogenia 55 0.0003 hyle Pathogenia 53 0.0006 trictical Interpretations 133 0.0007 citatical Consequences Counts Proportions faita Catagories Counts Proportions for proportions 562 0.0312 typen proportions 135 0.0112 typen proportions 136 0.0007 tart Catagories Counts Proportions gis promotions 138 0.0007 tart Catagories Counts Proportions gis promotions 138 0.0007 promotions 138 0.0007 tart Catagories Counts Proportions Coonservation 148 0.0007 Coonservation 148 0.0007 Coonservation 148 0.0007 Coonservation 148 0.0007	eing 9 0.0006 wer. Clinical Significance Clinical Significance inter Categories Counts Preportions kay Response 0 0.0000 https://wei.wei.wei.wei.wei.wei.wei.wei.wei.wei.			
ar. Cinical Significance Circuit Proportions g Response 0 0.0009 hy Pathogenic 5 0.0003 hy Pathogenic 3.0 0.0158 hy Pathogenic 0.0011 Poportions hy Pathogenic 0.0011 Poportions hy Pathogenic 0.0011 Poportions hy Pathogenic 0.0011 Poportions hy Pathogenics 0.0011 Poportions hy Pathogenics 0.0011 Poportions hy Pathogenics 0.0011 Poportions Di Potentoris 102 0.0009 GL Chanacowin 18 0.0012 Integrative Functional Store Categorys >= 0 Poportions Mathone Donality 7068 0.00136 Di Pomoritor To Store Store Categorys >= 15 Poportions Mathone Donality 0.002 0.0014 <td>var: Clicical Significance Circuits Proportions into Classopica Counts Proportions into into classopica 0.00030 aly Plantopanic 5 0.00030 atrice and consequences Counts Proportions intericing Interpretation 133 0.0000 ctical Alucation Socie Counts Proportions interpretations 143 0.00077 interpretations 143 0.00077 interpretations 148 0.0012 interpretation 148 0.0012 interpretation 148 0.0012 interpretation 148 0.0012 interpretation 148 0.00</td> <td>Splicing</td> <td></td> <td></td>	var: Clicical Significance Circuits Proportions into Classopica Counts Proportions into into classopica 0.00030 aly Plantopanic 5 0.00030 atrice and consequences Counts Proportions intericing Interpretation 133 0.0000 ctical Alucation Socie Counts Proportions interpretations 143 0.00077 interpretations 143 0.00077 interpretations 148 0.0012 interpretation 148 0.0012 interpretation 148 0.0012 interpretation 148 0.0012 interpretation 148 0.00	Splicing		
Arr. Classical Significance Counts Proportions g Response Counts Proportions g Response 0 0.0000 hy Benjoni 5 0.0003 tign 0.005 0.0469 hy Benjoni 346 0.0234 convon 233 0.0198 tincical Consequences Counts Proportions tintic Categories Counts Proportions p Response 143 0.0007 nonymous 502 0.0340 typhen Probably Damaging 124 0.0064 GE Promoters 118 0.0012 tant Categories Counts Proportions C Conservation 166 0.0112 typhen Probably Damaging 124 0.0064 GE Promoters 102 0.0064 C Conservation 166 0.0112 typhen Probably Damaging 124 0.0064 C Conservation 166 0.0412 C Conservation 166 0.0412 C Conservation 166 0.0412	riant Categories Counts Proportions ing Response 0 0.0000 ally Pathogenic 5 0.0000 ally Pathogenic 5 0.0000 ally Pathogenic 5 0.0000 ally Pathogenic 5 0.0000 ally Pathogenic 13 0.0000 alloting Intergritation 133 0.0000 actional Consequences infait Categories Counts Proportions OF 7 00 000112 https://www.fuelcoide.org/ Categories 102 0.00000 Alloting Counts Proportions OF Protectional Score Category >= 10 riant Categories Counts Proportions OF Optimer Functional Score Category >= 10 riant Categories Counts Proportions OF Optimer S 102 0.00000 Alloting Counts Proportions OF Optimer S 100 0.0451 OF Optimer S 10	phicing	5 0.000	10
Arr. Classical Significance Counts Proportions g Response Counts Proportions g Response 0 0.0000 hy Benjoni 5 0.0003 tign 0.005 0.0469 hy Benjoni 346 0.0234 convon 233 0.0198 tincical Consequences Counts Proportions tintic Categories Counts Proportions p Response 143 0.0007 nonymous 502 0.0340 typhen Probably Damaging 124 0.0064 GE Promoters 118 0.0012 tant Categories Counts Proportions C Conservation 166 0.0112 typhen Probably Damaging 124 0.0064 GE Promoters 102 0.0064 C Conservation 166 0.0112 typhen Probably Damaging 124 0.0064 C Conservation 166 0.0412 C Conservation 166 0.0412 C Conservation 166 0.0412	Ware Classed Significance age Response Counter Preportione ug Response 0 0.0000 holgenic 57 0.0003 nign 065 0.0400 w/ Bengin 346 0.0234 nicitic laterpretatione 133 0.0080 ectoral Consequences Interpretatione 155 0.0012 nicital Categories Counts Preportions 0.0012 hyfthen Probably Damaging 124 0.0004 GE Promoters 188 0.0012 histogative Functional Score Category >= 10 Interpretation 166 0.0412 Ge Promoters 188 0.0012 Histogative Functional Score Category >= 10 Interpretation 166 0.0412 0.0004 GE Promoters 188 0.0012 Holgenshive Functional Score Category >= 10 Interpretation 166 0.0412 Ge Departing Functional Score Category >= 10 Interpretation Factor 666 0.0451 Ge Departing Functional Score Category >= 15 Interpretation Factor 666 0.0451 Ge Dep			
Link Categories Counts Propertions g Respons 0 0.0000 g Respons 0.0000 0.0000 g Respons 0.0000 0.0000 g Respons 0.00234 0.0000 g Respons 0.0000 0.0141 g Respons 0.0000 0.0141 g Respons 0.0000 0.0141 g Respons 165 0.0112 g Respons 165 0.0112 g Respons 165 0.012 g Respons 165 0.012 g Respons 166 0.0124 g Respons 166 0.0126 g Respons 166 0.0126 g Respons 166 0.0126 g Respons 166 0.0126 g Respons 0.0201 g Respons	riant Categories Counts Proportions Other integration Other integ	linvar: Clinical Significan	e	
g Response hogenic 5 0.0000 hy Pathogenic 5 0.0000 hy Benign 346 0.0224 nrown 233 0.0185 ticing Interpretation 333 0.0090 ticing Interpretation 333 0.0090 ticing Interpretation 333 0.0090 ticing Interpretation 56 0.0000 ticing Interpretation 57 0.0000 ticing Interpretation 50 0.0090 ticing Interpretation 50 0.0090 0.0011 ticing Interpretation 50 0.0090 0.0012 ticing Interpretation 50 0.0090 0.0012 ticing Interpretation 50 0.0012 ticing Interpretation 50 0.00451 0.0090 0.00451 0.0012 ticing Interpretation 50 0.00451 0.0012 ticing Interpretation 10 0.0090 0.00451 0.0012 ticing Interpretation 10 0.0012 ticing Interpretation 0.0012 ticing Interpretation	μg Response 0 0.0000 thogenic 57 0.0039 with Pathogenic 5 0.0003 with Pathogenic 0.00011 with Pathogenic 0.00012 h Integrative Functional Score Category >= 10 mata Categories Count With Pathogenic 0.0012 h Mathogenic Pathogenic 0.0012 <td>ariant Categories</td> <td></td> <td>ortions</td>	ariant Categories		ortions
Topolatic 57 0.0039 by Pathogenic 5 0.0003 tign 606 0.0409 by Benign 636 0.0234 crown 233 0.0158 tilliciting Interpretation 133 0.0090 attal Categories Counts Proportions 0F 75 0.0061 0roymous 502 0.0340 tryptive Misenene 143 0.0090 0 Conservation 166 0.0112 tryptive Misenene 143 0.0091 200 Developing 0.0141 Toeleterious 165 0 Conservation 1661 0.0112 tryptive Misenene 143 0.0012 tripticational Score Category >= 10 Score Promoters CACE Prevees attal Categories Counts Proportions C Protein Function 496 0.0453 C Conservation 1661 0.1124 C Deparentics 885 0.0655 C Interpretics 885 0.0643 D Provinity To TSS-TES Matconenentry Michaelee Co	Integratic 57 0.0339 at/ Pathogenic 5 0.0039 at/ Pathogenic 5 0.0039 at/ Pathogenic 5 0.0039 at/ Pathogenic 5 0.0039 at/ Pathogenic 33 0.0039 at/ Pathogenic 133 0.0039 at/ Pathogenic 75 0.0051 at/ Pathogenic 75 0.0051 argenorymous 502 0.0340 sruptive Missensa 143 0.0007 argenorymous 209 0.01411 TD Deleterious 165 0.0112 AyPen Probably Damaging 124 0.0009 MGE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 CdE Ensemp AC Passing Paragenetics Citic Conservation 169 0.0451 CdE Conservation Citic Conservation 169 0.0451 CdE Conservation Citic Conservation 169 0.0451 CdE Conservation Citic Conservation 169<	Drug Response		
by Brain 346 0.0234 crown 233 0.0158 introgram 233 0.0168 introgram 233 0.0168 introgram 233 0.0168 introgram 233 0.0168 introgram 75 0.0061 introgram 502 0.0340 rymphon 502 0.0340 rymphon 165 0.0112 Phen Probably Damaging 102 0.0069 GE Promotery 18 0.0012 Integrative Functional Score Category >= 10 iant Categories Counts Conservation 1661 0.1124 Counts Versity 805 0.0336 C Conservation 1661 0.1124 C Donal Nucleotide Diversity 805 0.0451 C Mutatio Density <	and Benjin 346 0.0234 known 223 0.0158 mint Categories 23 0.0158 mint Categories Counts Proportions OF 75 0.0051 monymous 502 0.0340 sruptive Missense 143 0.0007 monymous 209 0.0141 TP Deletarious 155 0.0112 AyPhen Probably Damagin 124 0.0084 KGE Promoters 102 0.0069 KGE Enhancers 18 0.0012 httegratics 835 0.0565 C2 Local Nucleotide Diversity 8057 0.5453 C2 Inscription Factor 666 0.04615 C2 Inscription Factor 666 0.04615 C3 Transcription Factor 666 0.04615 C4 Mapability To TSS-TES 3484 0.2388 MDD PHRED 1551 0.1050 httegratics Category >= 15 mint Categories Counts Proportions C2 Protein Function 496 0.0338 C4 Chatation Density 7080 0.04782 C4 Mapability To TSS-TES 3484 0.2388 MDD PHRED 1551 0.1050 httegratics 643 0.0381 C4 Chatation Density 55 0.00441 C4 Deplementes 412 0.0279 C4 Conservation 563 0.0381 C4 Conservation 563 0.0381	athogenic		
by Brain 346 0.0234 crown 233 0.0158 introgram 233 0.0168 introgram 233 0.0168 introgram 233 0.0168 introgram 233 0.0168 introgram 75 0.0061 introgram 502 0.0340 rymphon 502 0.0340 rymphon 165 0.0112 Phen Probably Damaging 102 0.0069 GE Promotery 18 0.0012 Integrative Functional Score Category >= 10 iant Categories Counts Conservation 1661 0.1124 Counts Versity 805 0.0336 C Conservation 1661 0.1124 C Donal Nucleotide Diversity 805 0.0451 C Mutatio Density <	and Benjin 346 0.0234 known 233 0.0158 mint Categories 233 0.0158 mint Categories Counts Proportions OF 75 0.0051 monymous 502 0.0340 sruptive Missense 143 0.0007 monymous 209 0.0141 TP Deletarious 155 0.0112 AyPhen Probably Damagin 124 0.0064 KGE Enhancers 18 0.0012 httegrative Functional Score Category >= 10 miant Categories Counts Proportions C2 Orden Function 466 0.0453 C2 Conservation 1661 0.1124 C2 Epigenetics 855 0.0565 C3 Local Nucleotide Diversity 8057 0.5453 C4 Mustion Density 7080 0.04782 C4 Transcriptions Factor 666 0.0451 C4 Transcriptions Factor 666 0.0451 C4 Transcriptions Factor 666 0.0451 C4 Dratin Density 7080 0.04782 C4 Dratin Density 7080 0.0453 C4 Dratin Density 7080 0.0556 C4 Dratin Density 7080 0.0453 C4 Drati	-		
by Brain 346 0.0234 crown 233 0.0158 introgram 233 0.0168 introgram 233 0.0168 introgram 233 0.0168 introgram 233 0.0168 introgram 75 0.0061 introgram 502 0.0340 rymphon 502 0.0340 rymphon 165 0.0112 Phen Probably Damaging 102 0.0069 GE Promotery 18 0.0012 Integrative Functional Score Category >= 10 iant Categories Counts Conservation 1661 0.1124 Counts Versity 805 0.0336 C Conservation 1661 0.1124 C Donal Nucleotide Diversity 805 0.0451 C Mutatio Density <	av) Binjin 346 0.0234 known 233 0.0158 mint Categories 233 0.0158 mint Categories Counts Proportions OF 75 0.0051 many Categories 143 0.0097 many mous 209 0.0141 Protectional Score Category >= 10 mint Categories 102 0.0099 KGE Enhancers 18 0.0012 Av/Phen Protokuk) Damaging 124 0.0064 KGE Franciers 102 0.0069 KGE Enhancers 18 0.0012 mint Categories Counts Proportions C2 Orden Function 466 0.0451 C2 Engigenetics 853 0.0666 C0 Local Nucleotide Diversity 8057 0.5453 C3 Conservation 1661 0.1124 C3 Transcription Factor 666 0.0451 C3 Transcription Factor 666 0.0451 C4 Mapability To TSS-TES 3484 0.2358 MD PHRED 1551 0.1050 h Integories Counts Proportions C4 Protein Function 496 0.0336 C4 Crosservation 1661 0.0152 C4 Transcription Factor 666 0.0451 C4 Mapability To TSS-TES 3484 0.2358 MD PHRED 1551 0.1050 h Integories Counts Proportions C4 Orden Integories Counts Proportions C4 Conservation 1669 0.0453 C4 Chancer 1551 0.1050 h Integories Counts Proportions C4 Conservation 553 0.0381 C4 Conservation			
Control 233 0.0158 Hicking Interpretations 133 0.0090 titon Consequences 133 0.0097 tory mous 502 0.0340 tory mous 124 0.0084 GE Promoters 102 0.0069 GE Enhancers 18 0.0012 totat Consequences Consequences Consequences Conservation 1661 0.1124 Conservation 1661 0.0124 Conservation 1661 0.0124 Conservation 1661 0.0124 Conservation 1661 0.0124 Consencicotide Diversity 0.0553 <	Known 233 0.0158 nilicting Interpretations 133 0.0090 Actional Consequences Functional Consequences Functional Consequences riant Categories Counts Proportions 0 OF 75 0.0061 Non- Non- programmus 502 0.0340 Non- Non- programmus 209 0.0112 Non- Non- sympher Probably Damaging 124 0.0064 Non- Non- NGE Promoters 102 0.0069 Non- Non- NGE Enhancers 18 0.0012 Nithergenitics Non- No Conservation 1661 0.1124 Non- Non- No Conservation 1661 0.1124 Non- Non- No Conservation 1661 0.1124 Non- Non- No Conservation 666 0.0453 Non- Non- Non- No Matalon Density 7089 0.0473 Non- Non- Non-			
Integrative functional Score Category >= 10 For portions Conservation 166 0.0012 Conservation 166 0.0012 Phen Probably Damaging 124 0.0069 GE Promoters 18 0.0012 Coloresequences Coloresequences Coloresequences Integrative Functional Score Category >= 10 Coloresequences Coloresequences Coloresequences Coloresequences Coloresequences Coloresequences Coloresev	Initiality Interpretations 133 0.0090 Initial Categories Counts Proportions OF 75 0.0061 snuptive Missense 143 0.0097 nonymorymous 5062 0.0340 snuptive Missense 143 0.0097 nonymous 165 0.0112 Vighten Probably Damaging 124 0.0084 Vighten Probably Damaging 124 0.0085 Vighten Probably Damaging 124 0.0084 Vighten Probably Damaging 124 0.0084 Vighten Probably Damaging 124 0.00356 Vighten Probably Damaging 0.0124 Vighten Probably Damaging 124 0.0356 Vighten Probably Damaging 124 0.0453 Vighten Probablity 669 0.0453 <			
Charact Categories Counts Proportions 26 75 0.0051 1000000000000000000000000000000000000	Control Consequences Functional Consequences riant Categories Counts Proportions OF 75 0.0051 inspinorymous 502 0.0040 supplie Messnes 143 0.0097 monymous 208 0.0112 supplie Messnes 165 0.0112 supplie Messnes 124 0.0084 GE Formotiers 102 0.0069 NGE Formotiers 18 0.0012 Nategories Counts Proportions CO Conservation 1661 0.1124 CO Eprotein Function 505 0.0565 Co Local Nucleotide Diversity 0.0570 Co Conservation 1661 0.1124 CO Eprotein Function 505 0.0565 Co Local Nucleotide Diversity 0.0575 0.0565 Co Local Nucleotide Diversity 0.0575 0.0565 Co Local Nucleotide Diversity 0.0565 0.0565 Co Local Nucleotide Diversity 0.0565 0.0565 Co Conservation			
counts Proportions pr 75 0.0051 hsynonymous 502 0.0340 nuptive Missense 143 0.0097 horing 209 0.0141 Toeletorious 105 0.0121 When Probably Damaging 124 0.0084 GE Promoters 102 0.0069 GE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 faint Categories Counts fordien Function 496 0.0336 Conservation 1661 0.1124 Chategories Counts Proportions Conservation 1661 0.1124 Chategories Counts Proportions Conservation 1661 0.1124 Chategories Counts Proportions Chategories Counts Proportions Chategories Counts Proportions Chategories Counts Proportions Chategories Counts Proportions<	Character Counts Proportions OF 75 0.0051 Instruct Categories Counts Proportions OF 75 0.0051 Instruction 143 0.0077 Instruction 165 0.0112 Apphen Probably Damaging 124 0.0084 KGE Fromoters 102 0.0069 NGE Enhancers 18 0.0012 Initial Categories Counts Proportions C2 Engenetics 635 0.0365 C3 Conservation 1661 0.1124 C4 Decal Nucleotide Diversity 635 0.0565 C3 Conservation 1661 0.0451 C4 Decal Nucleotide Diversity 635 0.0565 C4 Decal Nucleotide Diversity 656 0.0451 C4 Decal Nucleotide Diversity 656 0.0451 C4 Decal Nucleotide Diversity 656 0.0361 C4 Decal Nucleotide Diversity 656 0.0361 C4 Decal Nucleotide Diversity 656 0.0361			0.0090
Counts Proportions PF 75 0.0061 ruptive Missense 143 0.0097 ruptive Missense 143 0.0097 romymous 209 0.0141 T Defentious 165 0.0069 GE Promoters 102 0.0069 GE Promoters 12 0.0012 Lintegrative Functional Score Category >= 10 iant Categories Counts Proportions C Conservation 1661 0.1124 Conservation 1661 0.1124 C Epigenetics 835 0.0565 Counts Proportions Counts Proportions C Matation Density 7080 0.4782 Conservation 1651 0.0163 C Prokinity To TSS-TES 3484 0.2386 Ocon Nectored Proportions Proportions C Prokinity To TSS-TES 3484 0.2386 Ocon Nectored Proportions Proportions C Prokinity To TSS-TES 3484 0.2386 Ocon Nectored Proportions Proportions Pro	Column Propertions Column Propertions OF 75 0.0051 rinsynonymous 502 0.0340 sruptive Missense 143 0.0097 Tonorymous 209 0.0141 FT Deleterious 165 0.0012 Mythen Probably Damaging 102 0.0064 VigPen Probably Damaging 102 0.0066 VigPen Probably Damaging 102 0.0066 VigPen Probably Damaging 102 0.0066 VigPen Probably Damaging 0.0012 High Integrative Functional Score Category >= 10 Integrative Function 496 0.0336 APC Anaeosime Fister Vig Conservation 1661 0.1124 Micro Damaity Vig Conservation 666 0.0451 Micro Damaity Vig Conservation 563 0.0381 Micro Damaity Vig Conservation 563	Inctional Consequences		
pF 75 0.0061 hspncrymous 502 0.0340 nuptive Missense 143 0.0007 normous 209 0.0141 T Deleterious 165 0.0112 yPhen Probabily Damaging 124 0.0066 GE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 High Integrative Guidenting tant Categories Counts Proportions C Protein Function 496 0.0336 C Conservation 1661 0.1124 C Mutation Density 669 0.0453 C Protein Function 496 0.0336 C Protein Function 496 0.0453 C Conservation 1661 0.1124 C Mutation Density 669 0.0453 C Protein Function 496 0.0336 C Protein Function 496 0.0336 C Protein Function 1551 0.0024 C Protein Function 496 0.0336 C Conservation 1551 0.0024 C Integrative Functional Score Category >= 15 <t< td=""><td>OF 75 0.0051 misynorymous 502 0.0340 arupive Missense 143 0.0097 morymous 209 0.0111 FT Deleterious 165 0.0112 When Probably Damaging 124 0.0069 NGE Francers 102 0.0069 NGE Enhancers 18 0.0012 hintegrative Functional Score Category >= 10 riant Categories Counts riant Categories Counts Proportions NC Conservation 1661 0.1124 CC Conservation 1669 0.0453 NC Mappability 7080 0.4792 CC Addition Density 7080 0.4792 CC Addition Density 7080 0.4533 NDD PHRED 1551 0.1050 hintegrative Functional Score Category >= 15 ration Categories Counts riant Categories Counts Proportions CC Conservation 563 0.0036 CC Conservation 563 0.0036</td><td>ariant Categories</td><td>Counts</td><td>Proportions</td></t<>	OF 75 0.0051 misynorymous 502 0.0340 arupive Missense 143 0.0097 morymous 209 0.0111 FT Deleterious 165 0.0112 When Probably Damaging 124 0.0069 NGE Francers 102 0.0069 NGE Enhancers 18 0.0012 hintegrative Functional Score Category >= 10 riant Categories Counts riant Categories Counts Proportions NC Conservation 1661 0.1124 CC Conservation 1669 0.0453 NC Mappability 7080 0.4792 CC Addition Density 7080 0.4792 CC Addition Density 7080 0.4533 NDD PHRED 1551 0.1050 hintegrative Functional Score Category >= 15 ration Categories Counts riant Categories Counts Proportions CC Conservation 563 0.0036 CC Conservation 563 0.0036	ariant Categories	Counts	Proportions
initial control	Integrative Generation Integrative Generation Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 15 Integrative Functional Score Category == 15 Integrative Function 196 0.00381 IC Conservation 563 0.002279	LOF	75	0.0051
botymous 209 0.0141 T Deleterious 165 0.0112 yPhen Probably Damaging 124 0.0069 GE Promoters 102 0.0069 GE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 10 iand Categories Counts Proportions C Potein Function 496 0.0336 C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Cocal Nucleotide Diversity 8057 0.5453 C Mappability 669 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 15 10.050 Integrative Functional Score Category >= 15 164 0.0336 C Conservation 563 0.0381 0.06453 C Conservation 563 0.0386 0.06453 C Conservation 563 0.0386 0.06463 C Conservation 563 0.0386 0.06279 C Local Nucleotide Diversity	monymous 209 0.0141 FT Deleterious 165 0.0112 MyPhen Probably Damaging 124 0.0069 MGE Ernancers 102 0.0069 MGE Enhancers 18 0.0012 Initial Categories Counts Proportions CC Conservation 1661 0.1124 Counts Proportions ACC Environs CC Conservation 1661 0.1124 Counts Proportions ACC Environs Counts 835 0.0566 Counts 8057 0.5453 Counts 669 0.0453 CC Environs ACC Environs ACC Environs ACC Environs Count Mucleoidie Diversity 669 0.0453 CC Environs Counts Proportions Count Mucleoidie Diversity 669 0.0453 Count Mucleoidie Diversity 669 0.0453 Count Mucleoidie Diversity 659 0.0024 Count Mucleoidie Diversity 659 0.0024 Count Mucleoidie Diversity 650 0.0024	onsynonymous	502	0.0340
nonymous 209 0.0141 T Deletarious 165 0.0112 yPhen Probably Damaging 124 0.0069 GE Enhancers 102 0.0069 GE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 15 Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Intanscription Factor 666 0.0451 C Transcription Factor 666 0.0451 C Transcription Factor 666 0.0453 D PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Integrative Functional Score Category >= 15 Integrative Function 496 0.0336 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65	nonymous 209 0.0141 TT Deletrious 165 0.0112 MyPhen Probabily Damaging 124 0.0069 KQE Enhancers 102 0.0069 KQE Enhancers 102 0.0069 KQE Enhancers 18 0.0012 Hintegrative Functional Score Category >= 10 ACC Enhancers ACC Enhancers riant Categories Counts Proportions Co Conservation 1661 0.1124 Co Coal Nucleotide Diversity 8057 0.5453 Co Coal Nucleotide Diversity 6059 0.0453 Co Conservation 1551 0.1050 Integrative Functional Score Category >= 15 ACC Onservation 666 Model 0.03361 Co Conservation 666 Co Transcription Function 496 0.03361 Co Conservation Co Conservation 563 0.0381 Co Conservation 666 Co Conservation 563 0.0381 Co Conservation 666 Co Conservation 563 0.0381 Co	isruptive Missense	143	0.0097
1 Decisional Code 103 0.0012 CASE Enhancers 102 0.0069 GE Promoters 102 0.0069 GE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 Integrative Function 496 Conservation 1661 0.1124 Conservation 1661 0.0453 Conservation 1551 0.1050 Integrative Functional Score Category >= 15 0.0336 Iant Categories Counts Proportions Conservation 563 0.0336	Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 10 Integrative Functional Score Category >= 15 APC Prevention Table Diversity Conservation 1661 Conservation 666 Conservation 1661 Conservation 666 Conservation 666 Conservation 666 Conservation 151 Conservation 151 Conservation 151 Conservation 563 Conservation </td <td>ynonymous</td> <td>209</td> <td>0.0141</td>	ynonymous	209	0.0141
WPhen Probably Damaging 124 0.0084 GE Promoters 102 0.0069 GE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 High Integrati Sci Protein Function 496 0.0336 C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Conservation 1666 0.0451 C Transcription Factor 666 0.0451 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0451 C Protein Function 496 0.0336 C Conservation 563 0.0381 C Depigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.22253 C Transcription Factor 328 0.0222 C Local Nucleotide Diversity 65 0.0044 C Mutation Density<	hyphen Probably Damaging 124 0.0084 GE Promoters 102 0.0069 NGE Enhancers 18 0.0012 h Integrative Functional Score Category >= 10 High Integrative Comparison riant Categories Counts Proportions CC Conservation 1661 0.1124 CC Epigenetics 835 0.0565 CC Local Nucleotide Diversity 8057 0.5453 CC Mappability 669 0.0433 CC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 h Integrative Functional Score Category >= 15 riant Categories Counts riant Categories Counts Proportions C Conservation 563 0.00381	IFT Deleterious	165	0.0112
GE Promoters 102 0.0069 GE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 High Integrati Integrative Function 496 0.0336 C Protein Function 496 0.0365 C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Local Nucleotide Diversity 8057 0.5453 C Mutation Density 7080 0.4792 C Proximity To TSS-TES 3484 0.2358 DD P HRED 1551 0.1050 Integrative Functional Score Category >= 15 Integrative Function 496 0.0336 C Onservation 663 0.0453 0.0150 Integrative Functional Score Category >= 15 Integrative Functional Score Category >= 15 Integrative Functional Score Category >= 15 Proportions C Conservation 653 0.0336 Onestago Proportions C Conservation 563 0.0326 Onestago Proportions C Conservation 563 0.0336 Onestago Pro	GE Promoters 102 0.0069 NGE Enhancers 18 0.0012 Integrative Functional Score Category >= 10 If integrative Functional Score Category >= 10 riant Categories Counts Proportions NC Protein Function 496 0.0336 NC Depteredics 835 0.0565 NC Local Nucleotide Diversity 6057 0.5453 NC Mutation Density 7080 0.4792 Ye Transcription Factor 666 0.0451 NC Materion Density 7080 0.4792 Ye Transcription Factor 666 0.0453 ND PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Integrative Functional Score Category >= 15 rian Categories Counts Proportions Ye Protein Function 496 0.0336 Ye Conservation 563 0.0341 Ye Conservation 563 0.0326 Ye Conservation 563 0.0326 Ye Conservation 563 0.0326 Ye Conservation 563 0.0326 Ye Conservation 563 <	olyPhen Probably Dama	ging 124	0.0084
GE Enhancers180.0012Integrative Functional Score Category >> 10ProportionsC Protein Function4960.0336C Conservation16610.1124C Epigenetics8350.0565C Local Nucleotide Diversity80570.5453C Mutation Density70800.4792C Transcription Factor6660.0453C Protein Function4960.0336C Protein Function500.0453D P HRED15510.1050Integrative Functional Score Category >> 15Integrative FunctionProportionsC Protein Function4960.0336C Conservation5630.0381C Protein Function4960.0336C Conservation5630.0381C Protein Function4960.0336C Conservation5630.0381C Epigenetics4120.0279C Local Nucleotide Diversity650.0044C Mutation Density3290.2223C Transcription Factor3280.0222C Transcription Factor3280.0222C Transcription Factor3280.0222C Mapability1520.0103	AGE Enhancers180.0012Integrative Functional Score Category >= 10High Integrativeriant CategoriesCountsProportionsC2 Protein Function16610.0124C2 Conservation16610.0124C2 Conservation16610.0124C2 Colcal Nucleotide Diversity80570.5453C2 Cocal Nucleotide Diversity6690.0453C2 Proximity To TSS-TES34840.2358NDD PHRED15510.1050Integrative Functional Score Category >= 15Proportionsriant CategoriesCountsProportionsC2 Conservation5630.0361C2 Conservation5630.0361C2 Conservation5630.0361C2 Conservation5630.0361C2 Conservation5630.0222C2 Cons	AGE Promoters		
Integrative Functional Score Category >> = 10 Figh Integrative Functional Categories Counts Proportions C Protein Function 1661 0.01366 Counts Proportions Counts APC Expension AP	h Integrative Functional Score Category >= 10 Proportiona Proportion	CAGE Enhancers		
Circularity Functional Score Category >= 10 Stant Categories Counts Proportions C Protein Function 1661 0.0336 C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Local Nucleotide Diversity 8057 0.6453 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Stant Categories Counts Proportions C Conservation 663 0.0336 APC Protein Function 496 0.0336 C Conservation 663 0.0381 APC Conservation APC Proportions C Conservation 663 0.0381 APC Conservation APC Inservice Factor C Deplemetics 412 0.0279 APC Inservice Factor APC Inservice Factor C Conservation 653 0.0044 APC Inservice Factor APC Inservice Factor APC Inservic	Integrative Functional Score Category >= 10 riant Categories Counts Proportions CC Protein Function 496 0.0336 CC Conservation 1661 0.1124 CC Epigenetics 835 0.0565 CL Cocal Nucleotide Diversity 8057 0.5453 CC Mutation Density 7080 0.4792 CC Transcription Factor 666 0.0451 CC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Friant Categories Counts Proportions CC Conservation 563 0.0381 0.0279 0.0479 CC Conservation 563 0.0381 0.0279 0.0479 CC Conservation 563 0.0381 0.0279 0.0279 CC Cocal Nucleotide Diversity 65 0.0044 0.0222 0.01033 CC Conservation 563 0.0222 0.0222 0.0222 0.0222 CC Mappability 152 0.0103 0.0222		10	0.0012
Circularity Functional Score Category >= 10 Stant Categories Counts Proportions C Protein Function 1661 0.0336 C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Local Nucleotide Diversity 8057 0.6453 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Stant Categories Counts Proportions C Conservation 663 0.0336 APC Protein Function 496 0.0336 C Conservation 663 0.0381 APC Conservation APC Proportions C Conservation 663 0.0381 APC Conservation APC Inservice Factor C Deplemetics 412 0.0279 APC Inservice Factor APC Inservice Factor C Conservation 653 0.0044 APC Inservice Factor APC Inservice Factor APC Inservic	Integrative Functional Score Category >= 10 riant Categories Counts Proportions CP Protein Function 496 0.0336 CC Conservation 1661 0.1124 CC Epigenetics 835 0.0565 CL Cocal Nucleotide Diversity 8057 0.5453 CC Mutation Density 7080 0.4792 CC Transcription Factor 666 0.0451 CP Consumption Factor 666 0.0453 CC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Friant Categories Counts Proportions CC Conservation 563 0.0381 0.0279 0.0279 CC Local Nucleotide Diversity 65 0.0044 0.0222 0.0222 Chanenything To TSS-TES 1170 0.0792 0.0004 0 CP rotein Function 456 0.0004 0.0004 0.0004 0.0004 CC Conservation 563 0.00044 0.00044 0.00044 0			
Canto Categories Counts Proportions C Protein Function 496 0.0336 C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Local Nucleotide Diversity 8057 0.5453 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0451 C Mappability 669 0.0453 C Protein Function 496 0.0336 C Protein Function 50 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 France Categories C Conservation 563 0.0381 C Conservation 563 0.0381 C Conservation 563 0.0044 C Epigenetics 412 0.0229 C Conservation 563 0.0044 C Mutation Density 3329 0.22253 C Transcription Factor 328 0.0222 C Mappability	Initian Categories Counts Proportions PC Protein Function 496 0.0336 PC Conservation 1661 0.1124 VC Epigenetics 835 0.0565 PC Local Nucleotide Diversity 8057 0.5453 PC Mutation Density 7080 0.4792 PC Transcription Factor 666 0.0453 PC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integrative Functional Score Category >= 15 riant Categories Counts Proportions PC Protein Function 496 0.0336 PC Protein Function 496 0.0336 PC Conservation 563 0.0381 PC Protein Function 496 0.0336 PC Conservation 563 0.0381 PC Conservation 563 0.0381 PC Conservation 563 0.0381 PC Conservation 563 0.0044 PC Mutation Density 3329 0.22	gh Integrative Functiona	al Score Catego	ry >= 10
C Protein Function 496 0.0336 C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Cola Nucleotide Diversity 8057 0.5453 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0453 C Proximity To TSS-TES 3484 0.2336 DD P HRED 1551 0.1050 Integrative Functional Score Category >= 15 Conservation C Conservation 563 0.0386 C Conservation 563 0.0386 C Conservation 563 0.0386 C Conservation 563 0.0386 C Depignetics 412 0.0279 C Coal Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2223 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	CP Orbitin Function 496 0.0336 CC Conservation 1661 0.1124 CC Epigenetics 835 0.0565 CC Local Nucleotide Diversity 8057 0.5453 PC Mutation Density 7080 0.4792 PC Transcription Factor 666 0.0453 PC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 h Integrative Functional Score Category >= 15 Franscription riant Categories Counts Proportions PC Conservation 563 0.0386 PC Conservation 563 0.0386 PC Conservation 563 0.02279 PC Local Nucleotide Diversity 65 0.0044 PC Epigenetics 412 0.0279 PC Conservation 563 0.0386 PC Transcription Factor 328 0.0222 PC Mapability 65 0.0044 PC Transcription Factor 328 0.0222 PC Mapability 152 0.0103 PC Transcription Factor 328 0.0222 PC Mapabi	ariant Categories	Counts	Proportions
C Conservation 1661 0.1124 C Epigenetics 835 0.0565 C Local Nucleotide Diversity 8057 0.5453 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0453 C Morpability 669 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Figh Integration C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.00441 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	C2 Conservation 1661 0.1124 V2 Cepigenetics 835 0.0565 V2 Local Nucleotide Diversity 8057 0.5453 V2 Conservation Density 7080 0.4792 V2 Charascription Factor 666 0.0453 V2 Cranscription Factor 666 0.0453 V2 Conservation 1551 0.1050 h Integrative Functional Score Category >= 15 France Category >= 15 riant Categories Counts Propertions V2 Conservation 563 0.0381 V2 Conservation 563 0.0044 V2 Conservation 563 0.00219 V2 Conservation 563 0.0024 V2 Conservation 583 0.0222 V2 Mutation Density 329 0.2253 V2 Transcription Factor 328 0.0222 V2 Mappability 152 0.0103 V2 Proximity T	PC Protein Function		
C Epigenetics 835 0.0565 C Local Nucleotide Diversity 8057 0.5453 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0451 C Mapability 669 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integration C Conservation 663 0.0381 C Conservation 663 0.0381 C Conservation 663 0.0044 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mapability 152 0.0103	PC Epigenetics 835 0.0565 PC Local Nucleotide Diversity 8057 0.5453 PC Mutation Density 7080 0.4792 PC Transcription Factor 666 0.0451 PC Mapability 669 0.0463 PC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 h Integrative Functional Score Category >= 15 Friant Categories Counts PC Tortein Function 496 0.0336 PC Conservation 563 0.00381 PC Excel Nucleotide Diversity 65 0.0044 PC Transcription Factor 328 0.0222 PC Mutation Density 3329 0.2253 PC Transcription Factor 328 0.0222 PC Mapability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	PC Conservation	1661	0.1124
Cl Local Nucleotide Diversity 8057 0.5453 C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0451 C Mapability 669 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integration C Conservation 663 0.0381 C Conservation 663 0.0381 C Conservation 663 0.0044 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	NC Local Nucleotide Diversity 8057 0.5453 NC Mutation Density 7080 0.4792 NC Transcription Factor 666 0.0451 NC Mapability 669 0.0453 NC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Friance Category >= 15 riant Categories Counts Proportions NC Conservation 563 0.0381 NC Expigenetics 412 0.0279 NC Local Nucleotide Diversity 65 0.0044 NC Mutation Density 3329 0.22253 NC Transcription Factor 328 0.0222 NC Mapability 152 0.0103 NC Proximity To TSS-TES 1170 0.0792	PC Epigenetics		
C Mutation Density 7080 0.4792 C Transcription Factor 666 0.0453 C Mappability 669 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integrative Functional Score Category >= 15 C Conservation 496 0.0336 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	CC Mutation Density 7080 0.4792 CC Transcription Factor 666 0.0451 RC Papability 669 0.0453 RC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integrative Functional Score Category >= 15 High Integrative Functional Score Category >= 15 Protein Function 496 0.0336 0.0381 PC Conservation 563 0.0381 PC Explorentics 412 0.0279 PC Cocal Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.22253 PC Matation Factor 328 0.0222 PC Mapability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792			
C Transcription Factor 666 0.0451 C Mappability 669 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integrative Function C Onservation 663 0.0386 C Conservation 563 0.0386 C Epigenetics 412 0.0279 C Cocal Nucleotide Diversity 65 0.00444 C Mutation Density 3329 0.22253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	Caranscription Factor 666 0.0451 CADD Parently 669 0.0453 CCD Parently 669 0.0453 CCD Parently 669 0.0453 CCD Parently 1551 0.1050 Integrative Functional Score Category >= 15 High Integrative Functional Score Category >= 15 riant Categories Counts Proportions CC Conservation 496 0.0336 CC Conservation 563 0.0381 CC Conservation 563 0.0044 CC Cocal Nucleotide Diversity 65 0.0044 CALCO Preservition Factor 328 0.0222 CAD Preservition Factor 328 0.0222 CAD Proventity 152 0.0103 CP Construction Factor 328 0.0222 CAD Proventity 152 0.0103 CP Construction Table 0.0792			
C Mappability 669 0.0453 C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integrative Functional Conservation APC Protein Function C Protein Function 496 0.0386 APC Conservation C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Tanscription Factor 328 0.0222 C Mappability 152 0.0103	PC Mappability 669 0.0453 PC Proximity To TSS-TES 3484 0.2358 NDD PHRED 1551 0.1050 Hintegrative Functional Score Category >= 15 High Integrative Functional Score Category >= 15 High Integrative Functional Score Category >= 15 C Protein Function 496 0.0336 PC Conservation 563 0.0381 PC Experiences 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.22253 PC Mappability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792			
C Proximity To TSS-TES 3484 0.2358 DD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integration Iant Categories Counts Proportions C Protein Function 496 0.0336 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Tanscription Factor 328 0.0222 C Mapability 152 0.0103	PC Proximity To TSS-TES 3484 0.2358 ADD PHRED 1551 0.1050 In Integrative Functional Score Category >= 15 Figh Integrative Functional Score Category >= 15 riant Categories Counts Proportions PC Protein Function 496 0.0336 PC Conservation 563 0.0381 PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.22253 PC Mapability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792			
DD P HRED 1551 0.1050 Integrative Functional Score Category >= 15 High Integration tant Categories Counts Proportions C Protein Function 496 0.0336 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.22253 C Transcription Factor 328 0.0222 C Mapability 152 0.0103	ADD PHRED 1551 0.1050 Integrative Functional Score Category >= 15 Figh Integrative Functional Score Category >= 15 rint Categories Counts Proportions PC Protein Function 496 0.0336 PC Conservation 563 0.0381 PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.22253 PC Transcription Factor 328 0.0222 PC Mappability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	APC Mappability		
High Integrative High Integrative Integrative Functional Score Category >= 15 Integrative Functional Score Category >= 15 Integrative Functional Colspan="2">Colspan="2">Integrative Function C Protein Function 496 0.0336 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 329 0.2253 C Transcription Factor 328 0.0222 C Mapapability 152 0.0103	High Integrative Functional Score Category >> 15riant CategoriesCountsProportionsCP Drotein Function4960.0336C Conservation5630.0381C Epigenetics4120.0279C Local Nucleotide Diversity650.0044C Mutation Density33290.2253C Mappability1520.0103C Proximity To TSS-TES11700.0792			
Integrative Functional Score Category >= 15 iant Categories Counts Proportions C Protein Function 496 0.036 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mapability 152 0.0103	Initial Categories Counts Proportions PC Protein Function 496 0.0386 PC Conservation 563 0.0381 PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.2253 PC Mapability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	CADD PHRED	1551	0.1050
Integrative Functional Score Category >= 15 iant Categories Counts Proportions C Protein Function 496 0.036 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mapability 152 0.0103	Initial Categories Counts Proportions PC Protein Function 496 0.0336 PC Conservation 563 0.0381 PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.22253 PC Mapability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792			
C Protein Function 496 0.0336 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	APC Potenti Function 496 0.0336 PC Conservation 563 0.0381 PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.2223 PC Mappability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	ph Integrative Functiona	al Score Catego	y >= 15
C Protein Function 496 0.0336 C Conservation 563 0.0381 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	PC Protein Function 496 0.0336 PC Conservation 563 0.0381 PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.2253 PC Transcription Factor 328 0.0222 PC Mappability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	ariant Categories	Counts	Proportions
C Conservation 563 0.0361 C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	CC Conservation 563 0.0361 PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.2253 PC Transcription Factor 328 0.0222 PC Mappability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	C Protein Function	496	0.0336
C Epigenetics 412 0.0279 C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	PC Epigenetics 412 0.0279 PC Local Nucleotide Diversity 65 0.0044 PC Mathino Density 3329 0.2253 PC Mathino Factor 328 0.0222 PC Mappability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	PC Conservation		
C Local Nucleotide Diversity 65 0.0044 C Mutation Density 3329 0.2253 C Transcription Factor 328 0.0222 C Mappability 152 0.0103	PC Local Nucleotide Diversity 65 0.0044 PC Mutation Density 3329 0.2253 PC Transcription Factor 328 0.0222 PC Mapability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792			
C Mutation Density 3329 0.2253 APC Posimily To TSS-TES C Transcription Factor 328 0.0222 CADD PHRED C Mappability 152 0.0103	CC Mutation Density 3329 0.2253 APC Painting To TSS-TES PC Transcription Factor 328 0.0222 CADD PHRED 0 PC Mappability 152 0.0103 0.0792 0 PC Proximity To TSS-TES 1170 0.0792 0 0			
C Mutation Density 3329 0.2253 APC Posimily To TSS-TES C Transcription Factor 328 0.0222 CADD PHRED C Mappability 152 0.0103	CC Mutation Density 3329 0.2253 APC Pasimity To TISE-TES Count of the test of the test of t			
C Transcription Factor 328 0.0222 0 C Mappability 152 0.0103 0	CC Transcription Factor 328 0.0222 PC Mappability 152 0.0103 PC Proximity To TSS-TES 1170 0.0792	PC Mutation Density	3329	0.2253
	PC Proximity To TSS-TES 1170 0.0792	PC Transcription Factor	328	0.0222
	PC Proximity To TSS-TES 1170 0.0792		152	0.0103
G FIOAIIIIU II I				
	ADD PHRED 603 0.0408			

Figure 6. Features of the annotation Genomics Data Structure (aGDS) format: This figure shows the features of the aGDS format and the process of creating aGDS files by combining functional annotations with genotype data.

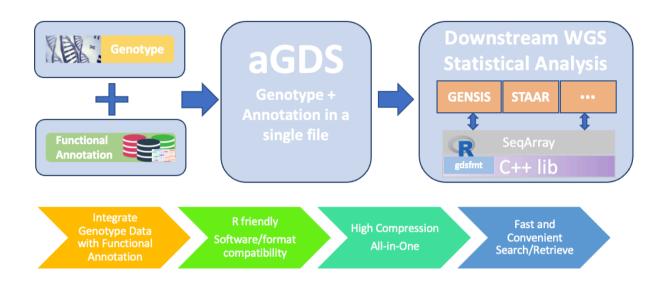
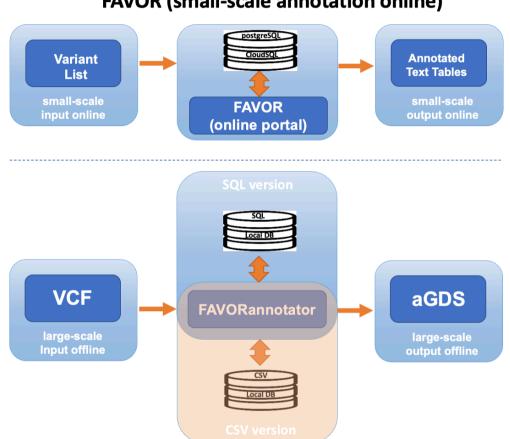


Figure 7. Graphical Representation of the features of FAVOR batch annotation and FAVORannotator: For small-scale annotation (up to 10,000 variants), batch annotation can be used for online annotation at the FAVOR website. For large-scale annotation, e.g., hundreds of millions of variants in a Whole Genome/Exome Sequencing (WGS/WES) study, FAVORannotator can be used for annotation in a local cluster or a cloud platform, e.g., Amazon Web Services (AWS) and Google Cloud Platform (GCP). FAVORannotator uses the backend database FAVOR, which is available in the SOL or CVS formats, and outputs an aGDS file that integrates genotype and annotation data in a single file.



FAVOR (small-scale annotation online)

FAVORannotator (Large-scale annotation):

R-script to annotate any WGS/WES study using the FAVOR database

Figure 8. Cloud-Native FAVORannotator Workflow. The interface of the FAVORannotator Workflow on Terra.bio.

Itera Workspaces > FAVORannotator-AnVIL/FAVORannotator > workflows > FAVORannotator FAVORannotator							COVID-19 Data & Tools	Cloud Environment
DASHBOARD	DATA	NOTEBOOKS	WORKFLOWS	JOB HISTORY				(;)
← Back to list								
Snapshot: 1 Source: FAVOR Synopsis: No documento Run workflo	AnVIL/FAVORar ation provided w with inputs de ww(s) with inputs	nnotator/12 fined by file paths defined by data table te intermediate outpr NPUTS ==		_	try with more memory (0		CANCEL SAVE
Hide optional i	nputs					Download json Drag or cli	ck to upload json	SEARCH INPUTS
Task name	L	,	/ariable		Туре	Attribute		
FAVORanno	tator		CHRN		Int	1		{}
FAVORanno	tator	1	nputaGDS		File	"gs://UKBB200kWES.chr1.gds"		▷ {}
FunctionalA	nnotation		FAVORannotator		File	"gs://FAVORannotatorTerra.R"		()