



Published in final edited form as:

Comput Biol Med. 2016 January 1; 68: 165–169. doi:10.1016/j.compbio.2015.03.028.

## WHATIF: an open-source desktop application for extraction and management of the incidental findings from next-generation sequencing variant data

Zhan Ye<sup>1,\*,#</sup>, Christopher Kadohph<sup>1,#</sup>, Robert Strenn<sup>1,#</sup>, Daniel Wall<sup>1</sup>, Elizabeth McPherson<sup>2</sup>, and Simon Lin<sup>1,3</sup>

<sup>1</sup>Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI 54449

<sup>2</sup>Department of Medical Genetics Services, Marshfield Clinic Research Foundation, Marshfield, WI 54449

### Abstract

**Background**—Identification and evaluation of incidental findings in patients following whole exome (WGS) or whole genome sequencing (WGS) is challenging for both practicing physicians and researchers. The American College of Medical Genetics and Genomics (ACMG) recently recommended a list of reportable incidental genetic findings. However, no informatics tools are currently available to support evaluation of incidental findings in next-generation sequencing data.

**Methods**—The Wisconsin Hierarchical Analysis Tool for Incidental Findings (WHATIF), was developed as a stand-alone Windows-based desktop executable, to support the interactive analysis of incidental findings in the context of the ACMG recommendations. WHATIF integrates the European Bioinformatics Institute Variant Effect Predictor (VEP) tool for biological interpretation and the National Center for Biotechnology Information ClinVar tool for clinical interpretation.

**Results**—An open-source desktop program was created to annotate incidental findings and present the results with a user-friendly interface. Further, a meaningful index (WHATIF Index) was devised for each gene to facilitate ranking of the relative importance of the variants and estimate the potential workload associated with further evaluation of the variants. Our WHATIF application is available at: <http://tinyurl.com/WHATIF-SOFTWARE>

**Conclusions**—The WHATIF application offers a user-friendly interface and allows users to investigate the extracted variant information efficiently and intuitively while always accessing the up to date information on variants via application programming interfaces (API) connections. WHATIF's highly flexible design and straightforward implementation aids users in customizing the source code to meet their own special needs.

\*Corresponding author: ye.zhan@mcrf.mfldclin.edu.

<sup>3</sup>(Current address) The Research Institute at Nationwide Children's Hospital, 700 Children's Drive Columbus, OH 43205

<sup>#</sup>These authors contributed equally.

## 1 INTRODUCTION

Whole-exome sequencing (WES) and whole-genome sequencing (WGS) technologies are being rapidly implemented and integrated into clinical research and practice (Wang *et al.*, 2013; Gonzaga-jaregu *et al.*, 2012; Marian 2012, Bick *et al.*, 2011; Rabbania *et al.*, 2014; Zatai *et al.*, 2014). Both WGS and WES predictably contribute a broad range of incidental findings (IFs) (Clayton *et al.*, 2013). According to the American College of Medical Genetics and Genomics (ACMG), IFs are defined as secondary findings, not related to the primary indication for ordering the sequencing, but potentially of clinical importance to the patient and ordering physician (Green *et al.*, 2013). As a guideline to clinicians and researchers confronted with evaluating IFs from genomic sequencing data, the ACMG recently recommended the reporting of 56 genes and 64 disorders (Green *et al.*, 2013; Clayton *et al.*, 2013). At the time there are no informatics tools available to facilitate this process.

The Wisconsin Hierarchical Analysis Tool for Incidental Findings (WHATIF) was developed to support annotation of all ACMG IF genes included in the Variant Call Format (VCF) file and the hierarchical browsing of the results from human organism studies. The variants are annotated using the Variant Effect Predictor (VEP) for biological interpretation (<http://useast.ensembl.org/info/docs/tools/vep/index.html>). VEP (McLaren *et al.*, 2010) determines the biological consequences of the variant and reports the Sorting Intolerant From Tolerant (SIFT) (Kumar *et al.*, 2009) and PolyPhen (Adzhubei *et al.*, 2010) scores. Simultaneously, the variants are annotated using the ClinVar Variation Reporter (<http://www.ncbi.nlm.nih.gov/variation/tools/reporter>) for clinical interpretation (Landrum *et al.*, 2014), which provides relevant clinical phenotypes and supporting evidence. In contrast to other tools that annotate all variants for an individual VCF, such as snpEff (Clingolani *et al.*, 2012) and ANNOVAR (Wang *et al.*, 2010), the WHATIF application specifically targets, the ACMG-reportable IF variants. Also unlike the command line-based tools such as snpEff and ANNOVAR, the WHATIF application provides a user friendly interface which allows users to define and select the exact information they wish to investigate. In addition, the users can add their own comments and rational based on the patient's clinical information, while browsing the variants and supporting evidence.

Our contributions are twofold. First, we created an open-source desktop program to annotate IFs and to present the results with a user-friendly interface. Source code has been provided which encompasses the entire process, beginning with the initial steps of preprocessing the raw VCF file through to the variant browser function for researchers wish to make their own customizations to the WHATIF application. Second, we devised a meaningful index (WHATIF Index) for each gene to facilitate ranking of the relative importance of the variants and estimate the potential workload associated with further evaluation of the variants.

## 2 WHATIF WORKFLOW

The WHATIF software uses a two-stage process. First, variants within the gene lists are extracted from the VCF file (blue boxes in Fig. 1). Second, VEP and ClinVar are queried

and the results returned from both databases are summarized hierarchically at both gene and variant levels (Orange boxes in Fig. 1). Thereafter, the user may investigate the annotated results using the WHATIF variant browser (green boxes in Fig. 1). The process of how WHATIF is operationalized is shown in Fig. 1.

## 2.1 Extraction of IF gene variants from VCF file

In this step, the program extracts variants corresponding to the ACMG-reportable IF genes using a VCF file generated by a WGS or WES study. As shown in Fig. 2, the last number in the column of 'index' (detailed further in section 3) shows the number of variants for each gene to be submitted to the VEP and ClinVar databases. The results are stored and used as the input for the next step.

## 2.2 Search information from VEP and ClinVar

For each variant extracted from step 2.1, WHATIF performs the analysis by submitting queries using VEP application programming interfaces (API) for biological interpretation and ClinVar API for clinical interpretation. For each variant, all relevant records for both biological and clinical interpretation are processed by the software and summarized in tabular form at both the gene-disease level and variant level.

## 2.3 Presentation of results

Representative results are presented hierarchically in an interactive format in three sortable tables, gene-disease table, variant level table and variant detail table. We introduced the concept of a three-component WHATIF index to help users to balance and sort out the following three concerns: clinically relevant variants with pathogenic functions, biologically important variants with biological functions, and all variants that might or might not functionally relevant due to the limitation of the current knowledge base. Fig. 2 shows the gene-disease level summary table in which WHATIF reports a high level summary of genes with the associated diseases, the number of variants characterized by clinical and biological functions and total variants extracted from VCF file. For each IF gene-disease pair, a link from MedGen website (<http://www.ncbi.nlm.nih.gov/medgen>) is provided for clinical and genetic details for the specific combination. Fig. 3 shows the variant level summary table in which WHATIF reports the gene associated with the variant, location of the variant, reference, alternative alleles, and zygosity. Users are able to sort the columns of this table. This allows the user to prioritize the list according to their preference with the default option being to display of all the variant(s) with clinical and biological functions at the top of the table. Further, the search function is also provided in this table to allow the user to search and filter by chromosome and location of variant. On selecting a row in this table, the user is directed to the variant detail screen (Fig. 4). The variant details screen contains four sections: the variant properties, clinical transcripts, biological transcripts from VEP and ClinVar and notes. Given space constraints, the number of data elements displayed is limited. However, users are able to display additional data elements through the use of the filters buttons. Users may display all the fields returned by both web services by adding or removing items in the filters sections for both clinical (ClinVar) and biological (VEP) results. WHATIF also allows the user to add notes concerning each variant by selecting the section for notes. When note(s) are added, a symbol is created in front of the variant, as

shown in Fig 3. When note(s) are saved, it can be retrieved the next time the user opens the same project. The column of notes has a clip board symbol. Sorting is also enabled for this table so that user could directly check out all the variants with notes.

### 3 APPLICATION

To demonstrate application of WHATIF, we analyzed four VCF files of WGS data. These samples have been analyzed and reported and are available for research purposes from the 1000 Genomes Project (1000 Genomes Project Consortium, *et al.*, 2012). The link is [ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120316\\_phase1\\_integrated\\_release\\_version2/](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120316_phase1_integrated_release_version2/).

A summary of the number of variants extracted from the VCF files, number of records and significant variants identified from both ClinVar and VEP are reported in Table 1. The VCF file used for our figures is sample NA19776. Variant counts for each gene and the corresponding WHATIF index (column labeled 'index') for both VEP and ClinVar APIs are summarized in Fig 2.

The construct of the WHATIF Index is displayed as a string with three numbers separated by colons, where the first number is the count of variants with significant clinical records returned by ClinVar. In this context, clinical significance is defined by the ClinVar clinical significance categories of Pathogenic, Likely pathogenic, Histocompatibility, and Drug Response. The second number represents the count of variants with significant biological records returned by VEP. 'Significance' in this context evaluates biological consequences in the context of table 2.

The third number represents the total number of variants submitted for the gene. For example, the genes SDHD, SCN5A, TSC1 in Fig. 2 shows 1 significant variant from ClinVar, 2, 2 and 7 significant variants from VEP and a total of 23, 161 and 34 variants, respectively. Whereas the WHATIF Index was designed to identify the most significant genes without requiring user review of all the variant level data, the user retains the option to select a gene, and review all of the variant level data (Fig 3.). For example, by selecting a specific variant of interest listed in Fig. 3, further details associated with the selected variant are displayed in Fig. 4. Through the addition of notes or the application of filters, the user is provided various options for customizing how the relevant information is displayed.

### 4 DISCUSSION

Our WHATIF application in its current form will offer clinicians as well as researchers the opportunity to perform analyses of single nucleotide variations (SNVs) using individual VCFs from next-generation sequencing (NGS) on their desktop computer. The graphical and user friendly interface will help users extract and navigate massive amounts of data while efficiently supporting user annotations. Maintenance of the software and the associated databases is minimal since our software utilizes API calls to retrieve data from both the European Bioinformatics Institute (EBI) and National Center for Biotechnology (NCBI) in real time. The WHATIF Index summarizes information for both clinical and biological

relevant variants requiring further investigation or confirmation. Search function will be included in the next release of the WHATIF software.

There are two common types of variations identified during a NGS experiment: SNVs and insertions/deletions (INDELs). Several studies have demonstrated high rates of inconsistency in INDEL calling applying various popular bioinformatics tools compared to the SNVs calling (O'Rawe *et al.*, 2013 and Shigemizu *et al.*, 2013). Both SNVs and INDELs have shown clinically meaningful and important functions. Due to the inconsistency of the calling algorithms, WHATIF is currently limited to the interpretation of SNVs. Future versions will seek to incorporate the capacity to analyze INDELs and other genetic variations. This will occur when the results of generating INDELs are more reliable. We utilized the APIs of VEP and ClinVar to annotate the VCF file. By using APIs we can ensure the WHATIF application will always display the latest data from the EBI and the NCBI. Concurrent with the preparation of this manuscript, a group from Stanford University reported release of a software called PATH-SCAN (Daneshjou *et al.* 2014). PATH-SCAN is a command line-based python software package for annotating the input variant files using ClinVar and ACMG gene list. This software requires users to have intermediate computer skills and knowledge of how to install a python based application to download all the necessary databases in order to perform analyses. Further, PATH-SCAN does not provide an interactive graphical interface and user-friendly navigation of the retrieved information nor is it clear whether functionality is provided for user annotation. An additional drawback of using pre-downloaded databases is the requirement for frequent checking and updating to keep information current and accurate. Another group from Spain also came up with a web-based interactive tool for whole-exome sequencing studies called BiERapp (Alemán *et al.* 2014). This web-based tool will allow user to upload a VCF and provides multiple levels of filtering using, allele frequencies, gene regions, segregation etc. to prioritize the variants. The primary purpose of BiERapp is to search for de novo mutation of disease-gene pairs rather than incidental findings. To use BiERapp, users are required to upload their VCF file to the computation server which might limit the usage of clinical samples, and to have an extensive genetic background in order to setup the filters for screening. Further, users are limited in their ability to customize the program to extract or display information that is relevant to their practice. In addition BiERapp does not store the user's files on the server, requiring users to upload the VCF file every single time they wish to analyze the data.

The open-source nature of WHATIF enables continuously improvement by the broad scientific and clinical genetics community. Examples of customizations include changing the gene list from the ACMG IF genes to any preferred gene list and the addition of additional supplemental information from the input VCF file such as read depth and the number of reads from both forward and reverse directions. For those users who utilize Mac/Linux/Unix-based systems rather than Windows-based systems, use of Virtualbox (Oracle VM Virtualbox, [www.virtualbox.org](http://www.virtualbox.org)) is recommended to establish a windows virtual machine to support installation and use of the WHATIF application. Users will still need a Windows license when using a virtual machine. For those users analyzing sequencing data from Complete Genomics platform, the variant file (.tsv) can be easily transformed into a

VCF file using their free available package: cgatools (Complete Genomics, <http://cgatools.sourceforge.net/>).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The project described was partially supported by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427. Authors are solely responsible for the content and do not represent the official views of the NIH. The authors would like to acknowledge the philanthropic support of Marshfield Clinic Research Foundation's Award (SP code: MCP10211) to Dr. Elizabeth McPherson. We thank Dr. Ingrid Glurich for critical review and Marie Fleisner for proofreading and formatting.

## References

- Green RC, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013; 15:565–574. [PubMed: 23788249]
- Clayton EW, et al. Managing incidental genomic findings: legal obligations of clinicians. *Genet Med*. 2013; 15:624–629. [PubMed: 23448723]
- Wang Z, et al. The role and challenges of exome sequencing in studies of human diseases. *Front Genet*. 2013; 4:160. [PubMed: 24032039]
- Gonzaga-Jauregui C, et al. Human genome sequencing in health and disease. *Annu Rev Med*. 2012; 63:35–61. [PubMed: 22248320]
- Marian AJ. Challenges in medical applications of whole exome/genome sequencing discoveries. *Trends Cardiovasc Med*. 2012; 22:219–223. [PubMed: 22921985]
- Bick D, Dimmock D. Whole exome and whole genome sequencing. *Curr Opin Pediatr*. 2011; 23:594–600. [PubMed: 21881504]
- Rabbania B, et al. The promise of whole-exome sequencing in medical genetics. *J Hum Genet*. 2014; 59:5–15. [PubMed: 24196381]
- Zatati MH, et al. Reporting results from whole-genome and whole-exome sequencing in clinical practice: a proposal for Canada? *J Med Genet*. 2014; 51:68–70. [PubMed: 24078715]
- McLaren W, et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010; 26:2069–2070. [PubMed: 20562413]
- Kumar P, et al. Predicting the effects of coding non-synonymous variations on protein function using the SIFT algorithm. *Nature Protocols*. 2009; 4:1073–1081. [PubMed: 19561590]
- Adzhubei IA, et al. A method and server for predicting damaging missense mutations. *Nature Methods*. 2010; 7:248–249. [PubMed: 20354512]
- Landrum MJ, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014; 42:D980–D985. [PubMed: 24234437]
- Clingolani P, et al. 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)*. 2012; 6:80–92. [PubMed: 22728672]
- Wang K, et al. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38:e164. [PubMed: 20601685]
- 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012 Nov 1; 491(7422):56–65. [PubMed: 23128226]
- O'Rawe, Jason, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and whole genome sequencing. *Genome Medicine*. 2013; 5:28. [PubMed: 23537139]
- Shigemizu, Daichi, et al. A practical method to detect SNVs and indels from whole genome and exome sequencing data. *Scientific Reports*. 2013; 3:2161. [PubMed: 23831772]

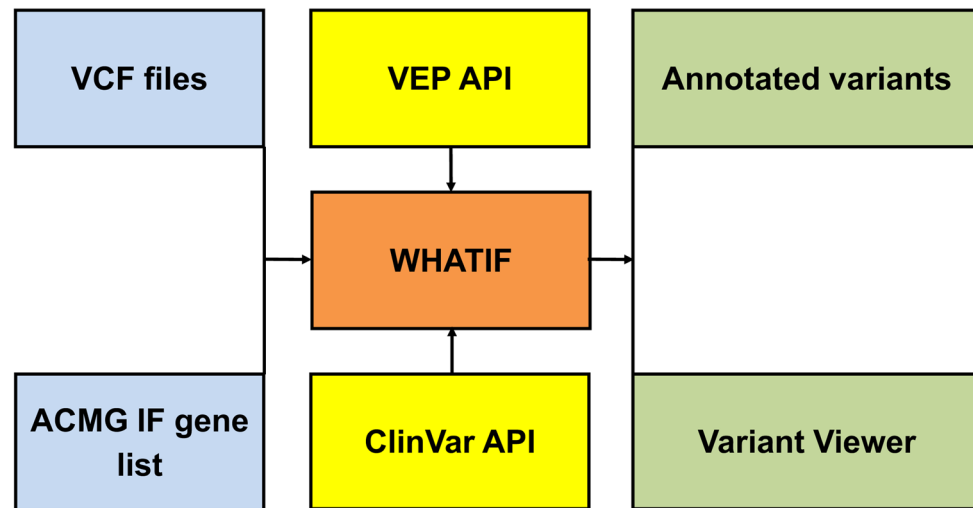
- Daneshjou, Roxana, et al. PATH-SCAN: a reporting tool for identifying clinically actionable variants. *Pac Symp Biocomput.* 2014:229–40. [PubMed: 24297550]
- Alemán A, et al. A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome sequencing studies. *Nucleic Acids Res.* 2014 Jul.42:W88–93. Web Server issue. [PubMed: 24803668]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 1.**  
WHATIF process. API represents application programming interfaces.



whatif variant viewer  
HOME VARIANTS DETAILS

Variant selection by Gene & Disease

Selected Gene Symbol Disease Name Index MedGen

<input checked="" type="checkbox"/>	SDHD	Paragangliomas 1	1:2:23	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=SDHD">http://www.ncbi.nlm.nih.gov/medgen/?term=SDHD</a>
<input checked="" type="checkbox"/>	SCN5A	Long QT syndrome 3	1:2:161	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=SCN5A">http://www.ncbi.nlm.nih.gov/medgen/?term=SCN5A</a>
<input checked="" type="checkbox"/>	SCN5A	Brugada syndrome 1	1:2:161	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=SCN5A">http://www.ncbi.nlm.nih.gov/medgen/?term=SCN5A</a>
<input type="checkbox"/>	TSC1	Tuberous sclerosis 1	1:1:21	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=TSC1">http://www.ncbi.nlm.nih.gov/medgen/?term=TSC1</a>
<input checked="" type="checkbox"/>	APOB	Familial hypercholesterolemia	0:7:34	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=APOB">http://www.ncbi.nlm.nih.gov/medgen/?term=APOB</a>
<input type="checkbox"/>	PCSK9	Hypercholesterolemia, autosomal dominant, 3	0:5:71	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=PCSK9">http://www.ncbi.nlm.nih.gov/medgen/?term=PCSK9</a>
<input type="checkbox"/>	MYLK	Aortic aneurysm, familial thoracic 7	0:4:270	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=MYLK">http://www.ncbi.nlm.nih.gov/medgen/?term=MYLK</a>
<input type="checkbox"/>	PMS2	Lynch syndrome	0:3:34	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=PMS2">http://www.ncbi.nlm.nih.gov/medgen/?term=PMS2</a>
<input type="checkbox"/>	CACNA1S	Malignant hyperthermia	0:3:142	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=CACNA1S">http://www.ncbi.nlm.nih.gov/medgen/?term=CACNA1S</a>
<input type="checkbox"/>	DSP	Arrhythmogenic right ventricular cardiomyopathy, type 8	0:2:88	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=DSP">http://www.ncbi.nlm.nih.gov/medgen/?term=DSP</a>
<input type="checkbox"/>	COL3A1	Ehlers-Danlos syndrome, type 4	0:2:37	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=COL3A1">http://www.ncbi.nlm.nih.gov/medgen/?term=COL3A1</a>
<input type="checkbox"/>	TMEM43	Arrhythmogenic right ventricular cardiomyopathy, type 5	0:2:28	<a href="http://www.ncbi.nlm.nih.gov/medgen/?term=TMEM43">http://www.ncbi.nlm.nih.gov/medgen/?term=TMEM43</a>

4 of 57 Diseases selected

Finish

**Figure 2.**

Gene-disease table. The WHATIF index is displayed as a string with three counts separated by colons. For example, the gene SDHD has 1 clinically significant variants from ClinVar, 2 biologically significant variants from VEP and a total of 23 variants.

whatif variant viewer

HOME VARIANTS DETAILS

Filter by Chromosome & Location: ex: 11, 3:3, 11:1-3

Notes	GeneLocation	Chromosome & Location	ReferenceAllele	ObservedAllele	Zygosity	Significant (Bio)	Significant (Cln)
	111957497-111990353	11:111958677	A	G	heterozygous	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	111957497-111990353	11:111962706	C	T	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111963860	A	G	heterozygous	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111964761	T	C	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111966307	T	C	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111966651	G	A	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111968063	G	A	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111968139	G	A	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111968227	C	G	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111972656	T	C	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111973603	T	G	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111974051	A	G	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111976320	G	A	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111977508	G	A	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111977919	C	T	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111983680	G	A	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111984131	A	C	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111984887	T	G	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111987364	C	A	homozygous alt	<input type="checkbox"/>	<input type="checkbox"/>
	111957497-111990353	11:111989558	A	T	heterozygous	<input type="checkbox"/>	<input type="checkbox"/>

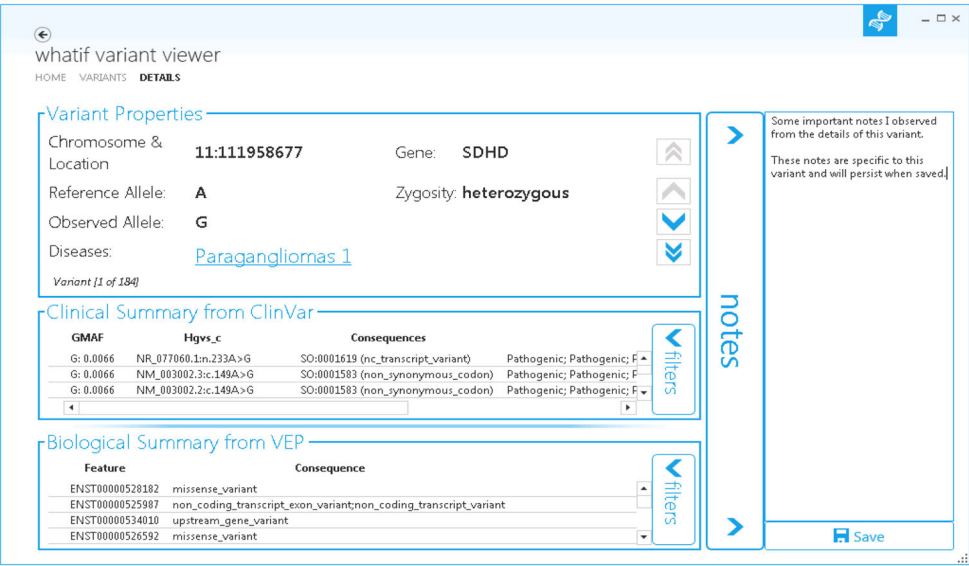
184 of 5488 Variants selected

Page 1 of 10

Gene & Disease Selection >>

**Figure 3.**

Variant table. The table reported all the basic information of a particular variant which were extracted from VCF file, All the column headers and notations follow the VCF standard. The column of Significant (Bio) and Significant (Cln) are indicators for whether the variant has significant features extracted from VEP and ClinVar database respectively.



**Figure 4.** Variant details table. This table contains information extracted from both VCF file for variant properties, VEP for biological summary and ClinVar for clinical summary.

Table 1

Summary counts of all four samples

Samples	Total # of <sup>a</sup> variants within the ACMG IF genes	Total # records of <sup>b</sup> ClinVar	Total # records of <sup>c</sup> VEP	# of sig ClinVar <sup>d</sup> Variants	# of sig VEP <sup>e</sup> Variants
NA19776	5488	20714	27536	3	53
HG00463	6194	20923	29579	0	72
NA12889	5955	21758	31244	1	62
NA19789	6212	22173	31752	4	66

<sup>a</sup> . Total number of variants extracted from the given sample VCF files  
<sup>b</sup> . Total number of records returned back from ClinVar database  
<sup>c</sup> . Total number of records returned back from VEP database  
<sup>d</sup> . Total number of significant variants of clinical significance within all the variants  
<sup>e</sup> . Total number of significant variants of biological significant within all the variants

Table 2

Biological consequences displayed are considered as biologically significant.

frameshift variant
stop_lost
stop_gained
incomplete_terminal_codon_variant
splice_donor_variant
inframe_deleteion
missense_variant
transcript_amplification
TFBS_ablation
TFBS_amplification
splice_acceptor_variant
splice_region_variant
transcript_ablation
initiator_codon_variant
inframe_insertion
TF_binding_site_variant
regulatory_region_variant
regulatory_region_ablation
regulatory_region_amplification