



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



Zhan Ye <sup>a,\*</sup>, Christopher Kadolph <sup>a,1</sup>, Robert Strenn <sup>a,1</sup>, Daniel Wall <sup>a</sup>, Elizabeth McPherson <sup>b</sup>, Simon Lin <sup>a,2</sup>

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

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

## ARTICLE INFO

### Article history:

Received 26 August 2014

Accepted 28 March 2015

### Keywords:

Genomics  
Internet  
Sequencing  
Software

## 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.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Whole-exome sequencing (WES) and whole-genome sequencing (WGS) technologies are being rapidly implemented and integrated into clinical research and practice [3–8]. Both WGS and WES predictably contribute a broad range of incidental findings (IFs) [2]. 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 [1]. 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 [1,2]. 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 [9] determines the biological consequences of the variant and reports the Sorting Intolerant From Tolerant (SIFT) [10] and PolyPhen [11] scores. Simultaneously, the variants are annotated using the ClinVar Variation Reporter (<http://www.ncbi.nlm.nih.gov/variation/tools/reporter>) for clinical

\* Corresponding author.

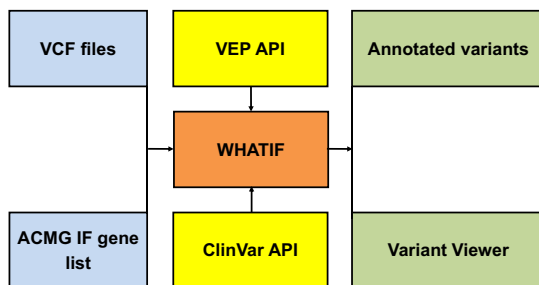
E-mail address: [ye.zhan@mcrf.mfldclin.edu](mailto:ye.zhan@mcrf.mfldclin.edu) (Z. Ye).

<sup>1</sup> These authors contributed equally.

<sup>2</sup> Current address: The Research Institute at Nationwide Children's Hospital, 700 Children's Drive Columbus, OH 43205, USA.

interpretation [12], which provides relevant clinical phenotypes and supporting evidence. In contrast to other tools that annotate all variants for an individual VCF, such as snpEff [13] and ANNOVAR [14], 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 rationale 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.



**Fig. 1.** WHATIF process. API represents application programming interfaces. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

## 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 Section 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

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

**Fig. 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.

Download English Version:

<https://daneshyari.com/en/article/504840>

Download Persian Version:

<https://daneshyari.com/article/504840>

[Daneshyari.com](https://daneshyari.com)