Pathogenicity Assessment of DNA Sequence Mutations

Mehmet Tevfik DORAK, MD PhD

2nd Practical Bioinformatics Course
Istanbul, 17/18 April 2017

YOUR FUTURE
STARTS WITH HOPE

Liverpool Hope University EST. 1844

Schedule

Day 1
Genome Biology in 2017
Bioinformatics Tools in Epigenetics
Pathogenicity Assessment of DNA Sequence Mutations
Introduction to Galaxy
Massive Data Sources
Practical

VCF Format

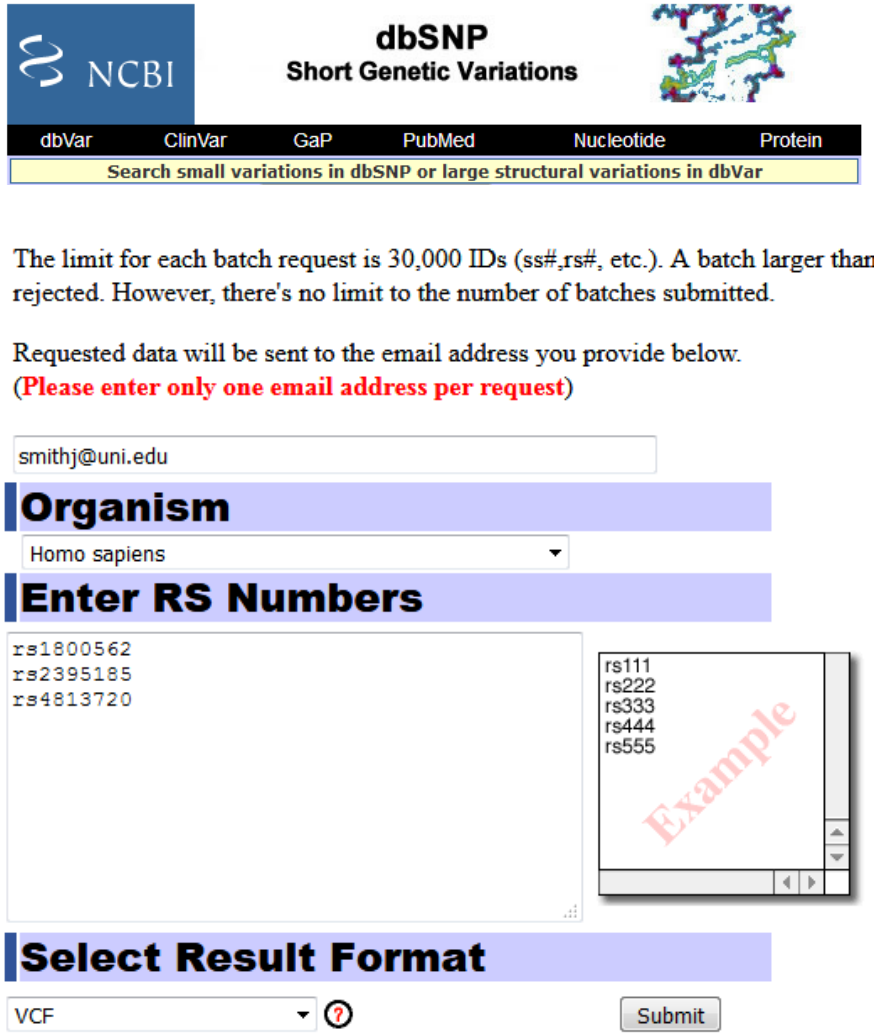
vcf_file.txt - Notepad

File Edit Format View Help

```
##fileformat=VCFv4.1
##fileDate=20131115
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO
2       166848563       .       C       G       100     PASS    AA=C;AN=2184;ERATE=0.0009;VT=SNP;RSQ=0.8819;AC=33;LDAF=0.0166;SNPSOURCE=LOWCOV;THETA=0.0013;AVGPOST=0.9949;AF=0.02;AMR_AF=0.06;AFR_AF=0.0020;EUR_AF=0.01
2       166903480       .       G       A       100     PASS    AA=G;AN=2184;AC=6;THETA=0.0005;AVGPOST=0.9996;VT=SNP;LDAF=0.0029;SNPSOURCE=LOWCOV;RSQ=0.9411;ERATE=0.0003;AF=0.0027;AFR_AF=0.01
2       166894356       .       C       T       100     PASS    AC=297;LDAF=0.1395;RSQ=0.8949;AN=2184;THETA=0.0005;VT=SNP;AVGPOST=0.9604;AA=A;ERATE=0.0008;SNPSOURCE=LOWCOV;AF=0.14;ASN_AF=0.01;AMR_AF=0.20;AFR_AF=0.03;EUR_AF=0.27
2       166198975       .       G       A       100     PASS    ERATE=0.0004;AA=C;AN=2184;LDAF=0.0010;VT=SNP;AVGPOST=0.9989;SNPSOURCE=LOWCOV;AC=1;THETA=0.0007;RSQ=0.5354;AF=0.0005;EUR_AF=0.0013
2       166852575       .       G       T       100     PASS    ERATE=0.0004;RSQ=0.4926;THETA=0.0004;AA=G;AN=2184;AVGPOST=0.9967;VT=SNP;LDAF=0.0022;SNPSOURCE=LOWCOV;AC=3;AF=0.0014;ASN_AF=0.0035;EUR_AF=0.0013
12      52082568        .       G       A       100     PASS    ERATE=0.0005;RSQ=0.5636;AA=C;AN=2184;VT=SNP;THETA=0.0006;SNPSOURCE=LOWCOV;AC=3;AVGPOST=0.9982;LDAF=0.0017;AF=0.0014;AFR_AF=0.01
12      52159534        .       T       A       100     PASS    ERATE=0.0005;AA=C;AN=2184;AVGPOST=0.9981;VT=SNP;LDAF=0.0014;SNPSOURCE=LOWCOV;AC=1;RSQ=0.4702;THETA=0.0003;AF=0.0005;AFR_AF=0.0020
9       130438189       .       G       A       100     PASS    AA=C;AN=2184;RSQ=0.8976;LDAF=0.0046;VT=SNP;AC=10;AVGPOST=0.9989;SNPSOURCE=LOWCOV;THETA=0.0010;ERATE=0.0003;AF=0.0046;AMR_AF=0.0028;AFR_AF=0.02
X       18606157        .       G       A       100     PASS    ERATE=0.0004;AC=18;THETA=0.0004;AA=G;AN=2184;VT=SNP;LDAF=0.0087;RSQ=0.8246;SNPSOURCE=LOWCOV;AVGPOST=0.9961;AF=0.01;AMR_AF=0.01;EUR_AF=0.02
9       130425622       .       C       T       100     PASS    AA=C;AN=2184;AVGPOST=0.9972;VT=SNP;AC=546;RSQ=0.9952;SNPSOURCE=LOWCOV;THETA=0.0007;LDAF=0.2496;ERATE=0.0007;AF=0.25;ASN_AF=0.17;AMR_AF=0.11;AFR_AF=0.73;EUR_AF=0.07
9       130444768       .       G       A       100     PASS    ERATE=0.0004;LDAF=0.0246;RSQ=0.9840;AA=T;AN=2184;VT=SNP;AVGPOST=0.9991;SNPSOURCE=LOWCOV;AC=53;THETA=0.0007;AF=0.02;AMR_AF=0.0028;AFR_AF=0.11
9       130434370       .       C       T       100     PASS    ERATE=0.0004;AA=G;AN=2184;THETA=0.0005;VT=SNP;LDAF=0.0424;RSQ=0.9636;SNPSOURCE=LOWCOV;AVGPOST=0.9958;AC=91;AF=0.04;ASN_AF=0.0017;AMR_AF=0.05;AFR_AF=0.01;EUR_AF=0.09
X       18598064        .       C       T       100     PASS    AA=C;RSQ=0.8413;LDAF=0.0103;AN=2184;VT=SNP;SNPSOURCE=LOWCOV;AC=21;THETA=0.0007;ERATE=0.0006;AVGPOST=0.9961;AF=0.01;ASN_AF=0.0017;AMR_AF=0.02;EUR_AF=0.02
8       133192493       .       G       A       100     PASS    RSQ=0.9920;LDAF=0.2068;AA=g;AN=2184;AVGPOST=0.9952;VT=SNP;THETA=0.0012;SNPSOURCE=LOWCOV;AC=452;ERATE=0.0006;AF=0.21;ASN_AF=0.02;AMR_AF=0.10;AFR_AF=0.63;EUR_AF=0.12
20      62076674        .       C       T       100     PASS    ERATE=0.0005;AA=C;RSQ=0.9513;AN=2184;LDAF=0.0109;VT=SNP;AC=23;SNPSOURCE=LOWCOV;THETA=0.0010;AVGPOST=0.9987;AF=0.01;AFR_AF=0.05
20      47990976        .       G       A       100     PASS    AA=G;AN=2184;AC=20;THETA=0.0005;VT=SNP;LDAF=0.0101;AVGPOST=0.9973;SNPSOURCE=LOWCOV;ERATE=0.0006;RSQ=0.8794;AF=0.01;AFR_AF=0.04;EUR_AF=0.0026
19      13414398        .       C       T       100     PASS    AC=7;AA=T;AN=2184;AVGPOST=0.9924;ERATE=0.0009;VT=SNP;THETA=0.0006;RSQ=0.5275;SNPSOURCE=LOWCOV;LDAF=0.0065;AF=0.0032;AFR_AF=0.0020;EUR_AF=0.01
12      13768545        .       C       A       100     PASS    ERATE=0.0005;AA=G;AN=2184;LDAF=0.0136;AVGPOST=0.9967;VT=SNP;RSQ=0.9028;SNPSOURCE=LOWCOV;AC=27;THETA=0.0007;AF=0.01;AMR_AF=0.0028;AFR_AF=0.05
11      33308577        .       G       A       100     PASS    ERATE=0.0004;AA=g;RSQ=0.7293;AN=2184;AVGPOST=0.9962;LDAF=0.0054;AC=8;VT=SNP;SNPSOURCE=LOWCOV;THETA=0.0010;AF=0.0037;ASN_AF=0.0017;EUR_AF=0.01
16      56385396        .       T       C       100     PASS    AA=G;AN=2184;ERATE=0.0009;LDAF=0.2289;THETA=0.0008;VT=SNP;AC=501;AVGPOST=0.9973;SNPSOURCE=LOWCOV;RSQ=0.9932;AF=0.23;ASN_AF=0.17;AMR_AF=0.16;AFR_AF=0.42;EUR_AF=0.18
22      38696896        .       T       C       100     PASS    ERATE=0.0005;AA=T;AN=2184;THETA=0.0008;LDAF=0.0010;VT=SNP;AVGPOST=0.9989;RSQ=0.5201;SNPSOURCE=LOWCOV;AC=1;AF=0.0005;AMR_AF=0.0028
5       140594621       .       A       G       100     PASS    ERATE=0.0004;LDAF=0.2697;AA=G;AN=2184;THETA=0.0008;VT=SNP;RSQ=0.9941;SNPSOURCE=LOWCOV;AVGPOST=0.9968;AC=590;AF=0.27;ASN_AF=0.17;AMR_AF=0.17;AFR_AF=0.60;EUR_AF=0.18
7       75959251        .       G       C       100     PASS    ERATE=0.0004;THETA=0.0004;RSQ=0.6753;AN=2184;VT=SNP;AA=A;LDAF=0.0007;SNPSOURCE=LOWCOV;AC=1;AVGPOST=0.9995;AF=0.0005;AFR_AF=0.0020
1       26448790        .       A       G       100     PASS    AA=C;AN=2184;LDAF=0.0032;AVGPOST=0.9972;THETA=0.0005;VT=SNP;AC=5;SNPSOURCE=LOWCOV;ERATE=0.0007;RSQ=0.6420;AF=0.0023;ASN_AF=0.0035;AFR_AF=0.0020;EUR_AF=0.0026
11      118348807       .       C       T       100     PASS    ERATE=0.0005;AC=228;AA=G;AN=2184;AVGPOST=0.9975;THETA=0.0005;VT=SNP;LDAF=0.1050;SNPSOURCE=LOWCOV;RSQ=0.9900;AF=0.10;ASN_AF=0.15;AMR_AF=0.05;AFR_AF=0.20;EUR_AF=0.03
21      45095034        .       G       T       100     PASS    ERATE=0.0005;LDAF=0.0378;AA=C;AN=2184;RSQ=0.9771;AC=84;VT=SNP;AVGPOST=0.9976;THETA=0.0006;SNPSOURCE=LOWCOV;AF=0.04;ASN_AF=0.0035;AMR_AF=0.01;AFR_AF=0.16
15      26866564        .       C       T       100     PASS    RSQ=0.9892;AA=T;AN=2184;ERATE=0.0009;AVGPOST=0.9947;VT=SNP;AC=459;LDAF=0.2094;SNPSOURCE=LOWCOV;THETA=0.0013;AF=0.21;ASN_AF=0.02;AMR_AF=0.12;AFR_AF=0.59;EUR_AF=0.15
9       130981471       .       G       C       100     PASS    ERATE=0.0005;AN=2184;AVGPOST=0.9967;VT=SNP;AA=A;RSQ=0.7481;SNPSOURCE=LOWCOV;AC=11;THETA=0.0007;LDAF=0.0050;AF=0.01;AMR_AF=0.01;AFR_AF=0.01;EUR_AF=0.0013
5       161322690       .       C       T       100     PASS    ERATE=0.0004;AA=G;AN=2184;VT=SNP;LDAF=0.0022;SNPSOURCE=LOWCOV;AVGPOST=0.9984;AC=3;RSQ=0.6874;THETA=0.0003;AF=0.0014;AFR_AF=0.01
5       14508071        .       C       T       100     PASS    ERATE=0.0004;THETA=0.0004;AA=G;AN=2184;LDAF=0.0010;VT=SNP;AVGPOST=0.9989;SNPSOURCE=LOWCOV;AC=1;RSQ=0.5224;AF=0.0005;EUR_AF=0.0013
9       140057118       .       A       C       100     PASS    AC=677;RSQ=0.9966;AA=C;AN=2184;VT=SNP;AVGPOST=0.9980;SNPSOURCE=LOWCOV;ERATE=0.0006;THETA=0.0034;LDAF=0.3090;AF=0.31;ASN_AF=0.17;AMR_AF=0.18;AFR_AF=0.76;EUR_AF=0.18
7       98654887        .       C       T       100     PASS    THETA=0.0004;AN=2184;VT=SNP;AA=A;RSQ=0.5597;AVGPOST=0.9988;LDAF=0.0011;SNPSOURCE=LOWCOV;AC=1;ERATE=0.0003;AF=0.0005;ASN_AF=0.0017
10      26446266        .       A       G       100     PASS    ERATE=0.0004;LDAF=0.0009;AA=G;AN=2184;AVGPOST=0.9990;VT=SNP;THETA=0.0033;RSQ=0.6225;SNPSOURCE=LOWCOV;AC=1;AF=0.0005;EUR_AF=0.0013
11      35336636        .       C       T       100     PASS    THETA=0.0004;AA=G;AN=2184;RSQ=0.9744;AVGPOST=0.9972;VT=SNP;SNPSOURCE=LOWCOV;LDAF=0.0416;ERATE=0.0006;AC=91;AF=0.04;AMR_AF=0.01;AFR_AF=0.17;EUR_AF=0.0026
10      62023645        .       T       C       100     PASS    AC=12;ERATE=0.0004;THETA=0.0015;AA=C;LDAF=0.0058;AN=2184;RSQ=0.9439;VT=SNP;AVGPOST=0.9993;SNPSOURCE=LOWCOV;AF=0.01;AFR_AF=0.02
9       130982295       .       G       C       100     PASS    RSQ=0.9108;THETA=0.0022;AA=G;AN=2184;AVGPOST=0.9975;VT=SNP;LDAF=0.0107;SNPSOURCE=LOWCOV;AC=21;ERATE=0.0006;AF=0.01;AMR_AF=0.01;AFR_AF=0.0041;EUR_AF=0.02
16      89598940        .       T       C       100     PASS    ERATE=0.0005;AVGPOST=0.9986;THETA=0.0015;AA=C;AN=2184;RSQ=0.9960;VT=SNP;LDAF=0.1673;SNPSOURCE=LOWCOV;AC=366;AF=0.17;ASN_AF=0.17;AMR_AF=0.09;AFR_AF=0.38;EUR_AF=0.06
1       26764708        .       G       A       100     PASS    AA=C;AVGPOST=0.9994;AN=2184;THETA=0.0005;VT=SNP;AC=9;RSQ=0.9401;LDAF=0.0044;SNPSOURCE=LOWCOV;ERATE=0.0003;AF=0.0041;AFR_AF=0.02
1       112998704       .       G       A       100     PASS    ERATE=0.0004;RSQ=0.9350;AA=C;AN=2184;THETA=0.0008;VT=SNP;AC=14;AVGPOST=0.9989;SNPSOURCE=LOWCOV;LDAF=0.0070;AF=0.01;AFR_AF=0.03
X       122778639       .       T       C       100     PASS    AC=7;LDAF=0.0035;THETA=0.0004;AA=G;AN=2184;RSQ=0.8542;VT=SNP;SNPSOURCE=LOWCOV;ERATE=0.0003;AVGPOST=0.9987;AF=0.0032;AMR_AF=0.0028;AFR_AF=0.01
5       110819814       .       G       C       100     PASS    THETA=0.0009;ERATE=0.0005;AVGPOST=0.9986;RSQ=0.5209;AA=T;AN=2184;VT=SNP;LDAF=0.0011;SNPSOURCE=LOWCOV;AC=1;AF=0.0005;ASN_AF=0.0017
11      9982858         .       G       A       100     PASS    ERATE=0.0004;AA=G;AN=2184;RSQ=0.5581;VT=SNP;AVGPOST=0.9993;SNPSOURCE=LOWCOV;AC=1;THETA=0.0013;LDAF=0.0008;AF=0.0005;EUR_AF=0.0013
17      7317981 .       G       A       100     PASS    ERATE=0.0004;AA=T;AN=2184;LDAF=0.0018;VT=SNP;THETA=0.0006;AVGPOST=0.9989;SNPSOURCE=LOWCOV;AC=4;RSQ=0.7631;AF=0.0018;AMR_AF=0.0028;EUR_AF=0.0040
1       203696551       .       T       C       100     PASS    LDAF=0.0251;AA=G;AN=2184;AC=55;AVGPOST=0.9996;VT=SNP;RSQ=0.9941;SNPSOURCE=LOWCOV;THETA=0.0010;ERATE=0.0003;AF=0.03;AMR_AF=0.01;AFR_AF=0.11
11      22770673        .       G       A       100     PASS    THETA=0.0022;AA=G;AN=2184;LDAF=0.0015;VT=SNP;AVGPOST=0.9993;SNPSOURCE=LOWCOV;ERATE=0.0003;RSQ=0.8061;AC=3;AF=0.0014;AMR_AF=0.0028;EUR_AF=0.0026
2       179638787       .       C       T       100     PASS    AA=G;AN=2184;RSQ=0.9927;THETA=0.0008;VT=SNP;LDAF=0.0461;AVGPOST=0.9989;SNPSOURCE=LOWCOV;AC=102;ERATE=0.0003;AF=0.05;AMR_AF=0.01;AFR_AF=0.20
18      9263876 .       G       A       100     PASS    ERATE=0.0004;RSQ=0.9807;AA=C;LDAF=0.0120;AN=2184;AC=26;VT=SNP;SNPSOURCE=LOWCOV;THETA=0.0010;AVGPOST=0.9995;AF=0.01;AMR_AF=0.0028;AFR_AF=0.05
10      5242207 .       G       T       802     PASS    AA=A;AC=672;AF=0.31;AFR_AF=0.75;AMR_AF=0.19;AN=2184;ASN_AF=0.17;AVGPOST=0.9978;ERATE=0.0004;EUR_AF=0.18;LDAF=0.3070;RSQ=0.9967;THETA=0.0008;VT=INDEL
X       3559896 .       G       A       100     PASS    AVGPOST=0.9994;LDAF=0.0466;AA=G;AN=2184;RSQ=0.9944;VT=SNP;SNPSOURCE=LOWCOV;AC=102;ERATE=0.0003;THETA=0.0003;AF=0.05;AMR_AF=0.01;AFR_AF=0.20
17      42226361        .       G       A       100     PASS    ERATE=0.0004;AA=G;AN=2184;VT=SNP;SNPSOURCE=LOWCOV;AC=1;THETA=0.0013;RSQ=0.6223;AVGPOST=0.9992;LDAF=0.0008;AF=0.0005;AFR_AF=0.0020
5       167922407       .       C       T       100     PASS    LDAF=0.0016;AA=G;AN=2184;RSQ=0.6663;THETA=0.0005;VT=SNP;SNPSOURCE=LOWCOV;AVGPOST=0.9987;AC=2;ERATE=0.0007;AF=0.0009;AMR_AF=0.0028;AFR_AF=0.0020
17      80051289        .       C       T       100     PASS    THETA=0.0009;LDAF=0.0013;AA=G;AN=2184;AVGPOST=0.9990;VT=SNP;SNPSOURCE=LOWCOV;RSQ=0.6745;AC=2;ERATE=0.0007;AF=0.0009;EUR_AF=0.0026
7       86522284        .       C       T       100     PASS    ERATE=0.0004;AA=C;AN=2184;VT=SNP;THETA=0.0006;AVGPOST=0.9989;SNPSOURCE=LOWCOV;AC=1;RSQ=0.4492;LDAF=0.0008;AF=0.0005;AFR_AF=0.0020
11      120007964       .       G       C       100     PASS    AN=2184;THETA=0.0005;VT=SNP;LDAF=0.0119;AA=A;ERATE=0.0008;SNPSOURCE=LOWCOV;AVGPOST=0.9968;AC=25;RSQ=0.8931;AF=0.01;AMR_AF=0.01;AFR_AF=0.05
4       47405630        .       T       C       100     PASS    AA=C;LDAF=0.3118;AN=2184;THETA=0.0005;VT=SNP;AVGPOST=0.9970;SNPSOURCE=LOWCOV;RSQ=0.9951;ERATE=0.0007;AC=681;AF=0.31;ASN_AF=0.17;AMR_AF=0.19;AFR_AF=0.76;EUR_AF=0.18
2       196682472       .       C       G       100     PASS    ERATE=0.0004;AA=G;AN=2184;AC=6;VT=SNP;THETA=0.0006;RSQ=0.9618;AVGPOST=0.9998;LDAF=0.0029;SNPSOURCE=LOWCOV;AF=0.0027;AFR_AF=0.01
9       3266287 .       A       G       100     PASS    AC=12;ERATE=0.0004;RSQ=0.9690;AA=G;AN=2184;AVGPOST=0.9996;LDAF=0.0057;VT=SNP;SNPSOURCE=LOWCOV;THETA=0.0013;AF=0.01;AFR_AF=0.02
10      89717712        .       C       T       100     PASS    ERATE=0.0004;AN=2184;AC=6;VT=SNP;RSQ=0.8597;LDAF=0.0034;AA=A;SNPSOURCE=LOWCOV;THETA=0.0010;AVGPOST=0.9987;AF=0.0027;AFR_AF=0.01
```

VCF Format from a SNP List

A list of up to 30,000 SNPs (rsIDs) can be converted to a VCF file (hg38 only) using this tool. A link to the VCF file is e-mailed when the conversion is complete.



Liverpool Hope University

Whole Genome: dbWGFP

dbWGFP — A database and web server of human whole-genome single nucleotide variants and their functional predictions

JIANGLAB | HOME | SERVICES | DOWNLOADS | HELP

- Introduction
- Services
- Downloads
- Help

Introduction

The recent advancement of the next generation sequencing technology has enabled the fast and low-cost detection of all genetic variants spreading across entire human genomes, making the application of whole-genome sequencing a tendency in studies of disease-causing genetic variants. Nevertheless, there still lacks a repository that collects predictions of functionally damaging effects of human genetic variants, though it has been well recognized that such predictions plays a central role in the analysis of whole-genome sequencing data.

To fill in this gap, we developed **dbWGFP** (*a database of human whole-genome single nucleotide variants and their functional predictions*) that contains functional predictions and annotations of more than 8.5 billion possible human whole-genome single nucleotide variants. Specifically, this database integrates 32 functional predictions calculated by 13 popular computational methods, 15 conservation features derived from 4 conservation calculation approaches, and 44 valuable annotations obtained from the ENCODE project, accompanied with a highly efficient search program.

dbWGFP offers two web services for retrieving predictions and annotations for human whole-genome single nucleotide variants. In the step-by-step mode, you can upload a file containing variants and retrieve results online. In the batch mode, you can upload a file containing your email address and variants and retrieve results via your email.

dbWGFP offers two versions for downloading. The lite version includes prediction scores for human whole-genome single nucleotide variants. The full version includes both predictions and annotations. Both versions include a search program that can extract predictions and/or annotations in a highly efficient way. Different versions of dbWGFP are also archived for easy access.

DATABASE

Database, 2016, 1–11
doi: 10.1093/database/baw024
Original article

Original article

dbWGFP: a database and web server of human whole-genome single nucleotide variants and their functional predictions

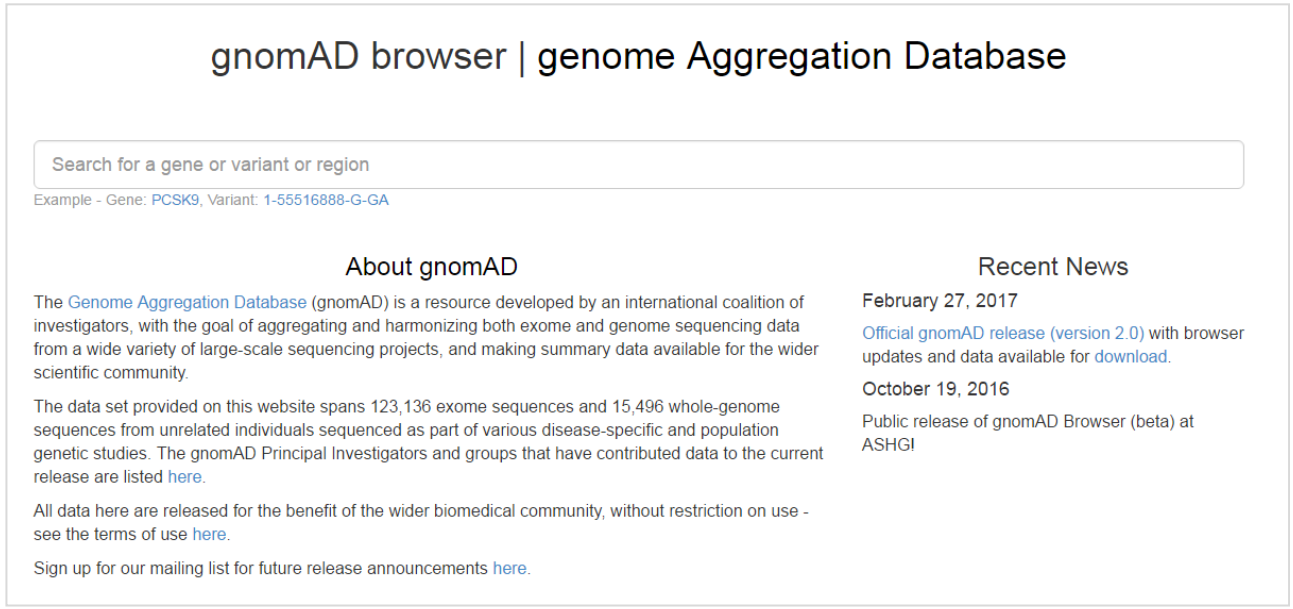
Jiaxin Wu, Mengmeng Wu, Lianshuo Li, Zhuo Liu, Wanwen Zeng and Rui Jiang*

Whole Genome: dbWGFP

Table 1. Computational methods for predicting functionally damaging effects or conservation properties of single nucleotide variants

Method	Version	Source	Website
Grantham	Sep-74	CADD	—
SIFT	Aug-11	dbNSFP	http://sift.jcvi.org
PolyPhen-2	v2.2.2	dbNSFP	http://genetics.bwh.harvard.edu/pph2
LRT	Nov-09	dbNSFP	http://www.genetics.wustl.edu/jflab/lrt_query.html
MutationTaster	Mar-13	dbNSFP	http://www.mutationtaster.org
Mutation Assessor	Release 2	dbNSFP	http://mutationassessor.org
FATHMM	v2.3	dbNSFP	http://fathmm.biocompute.org.uk
RadialSVM	v2.4	dbNSFP	http://sites.google.com/site/jpopgen/dbNSFP
LR	v2.4	dbNSFP	http://sites.google.com/site/jpopgen/dbNSFP
CADD	v1.0	CADD	http://cadd.gs.washington.edu
GWAVA	v1.0	GWAVA	https://www.sanger.ac.uk/sanger/StatGen_Gwava
MSRV	Aug-07	MSRV	http://bioinfo.au.tsinghua.edu.cn/msrv
SinBaD	Nov-12	SinBaD	http://tingchenlab.cmb.usc.edu/sinbad
phastCons	Nov-09	UCSC	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/phastCons46way
PhyloP	Nov-09	UCSC	http://hgdownload.soe.ucsc.edu/goldenPath/hg19/phyloP46way
GERP++	May-11	GERP	http://mendel.stanford.edu/SidowLab/downloads/gerp
SiPhy	v0.5	SiPhy	http://www.broadinstitute.org/genome_bio/siphy

Whole Genome: gnomAD



Contributing projects

- 1000 Genomes
- 1958 Birth Cohort
- ALSGEN
- Alzheimer's Disease Sequencing Project (ADSP)
- Atrial Fibrillation Genetics Consortium (AFGen)
- Estonian Genome Center, University of Tartu (EGCUT)
- Bulgarian Trios
- Finland-United States Investigation of NIDDM Genetics (FUSION)
- Finnish Twin Cohort Study
- FINN-ADGEN
- FINRISK
- Framingham Heart Study
- Génome Québec - Genizon Biobank
- Genomic Psychiatry Cohort
- GoT2D
- Genotype-Tissue Expression Project (GTEx)
- Health2000
- Inflammatory Bowel Disease:
  - Helsinki University Hospital Finland
  - NIDDK IBD Genetics Consortium
  - Quebec IBD Genetics Consortium
- Jackson Heart Study
- Kuopio Alzheimer Study
- LifeLines Cohort
- MESTA
- METabolic Syndrome In Men (METSIM)
- Finnish Migraine Study
- Myocardial Infarction Genetics Consortium (MIGen):
  - Leicester Exome Seq
  - North German MI Study
  - Ottawa Genomics Heart Study
  - Pakistan Risk of Myocardial Infarction Study (PROMIS)
  - Precocious Coronary Artery Disease Study (PROCARDIS)
  - Registre Gironi del COR (REGICOR)
  - South German MI Study
  - Variation in Recovery: Role of Gender on Outcomes of Young AMI Patients (VIRGO)
- National Institute of Mental Health (NIMH) Controls
- NHLBI-GO Exome Sequencing Project (ESP)
- NHLBI TOPMed
- Schizophrenia Trios from Taiwan
- Sequencing Initiative Suomi (SiSu)
- SIGMA-T2D
- Swedish Schizophrenia & Bipolar Studies
- T2D-GENES
  - GoDARTS
- T2D-SEARCH
- The Cancer Genome Atlas (TCGA)

Whole Genome: gnomAD (SNP)

Variant: 6:26093141 G / A

	Exomes	Genomes	Total
Filter	Pass	Pass	
Allele Count	8157	1169	9326
Allele Number	246016	30926	276942
Allele Frequency	0.03316	0.03780	0.03367
dbSNP	rs1800562		
UCSC	6-26093141-G-A		
ClinVar	Click to search for variant in Clinvar		

Genotype Quality Metrics

Site Quality Metrics

Report this variant

Annotations

This variant falls on 14 transcripts in 1 genes:

missense

- HFE Transcripts

non coding transcript exon

- HFE Transcripts

intron

- HFE - ENST00000352392

Note: This list may not include additional transcripts in the same gene that the variant does not overlap.

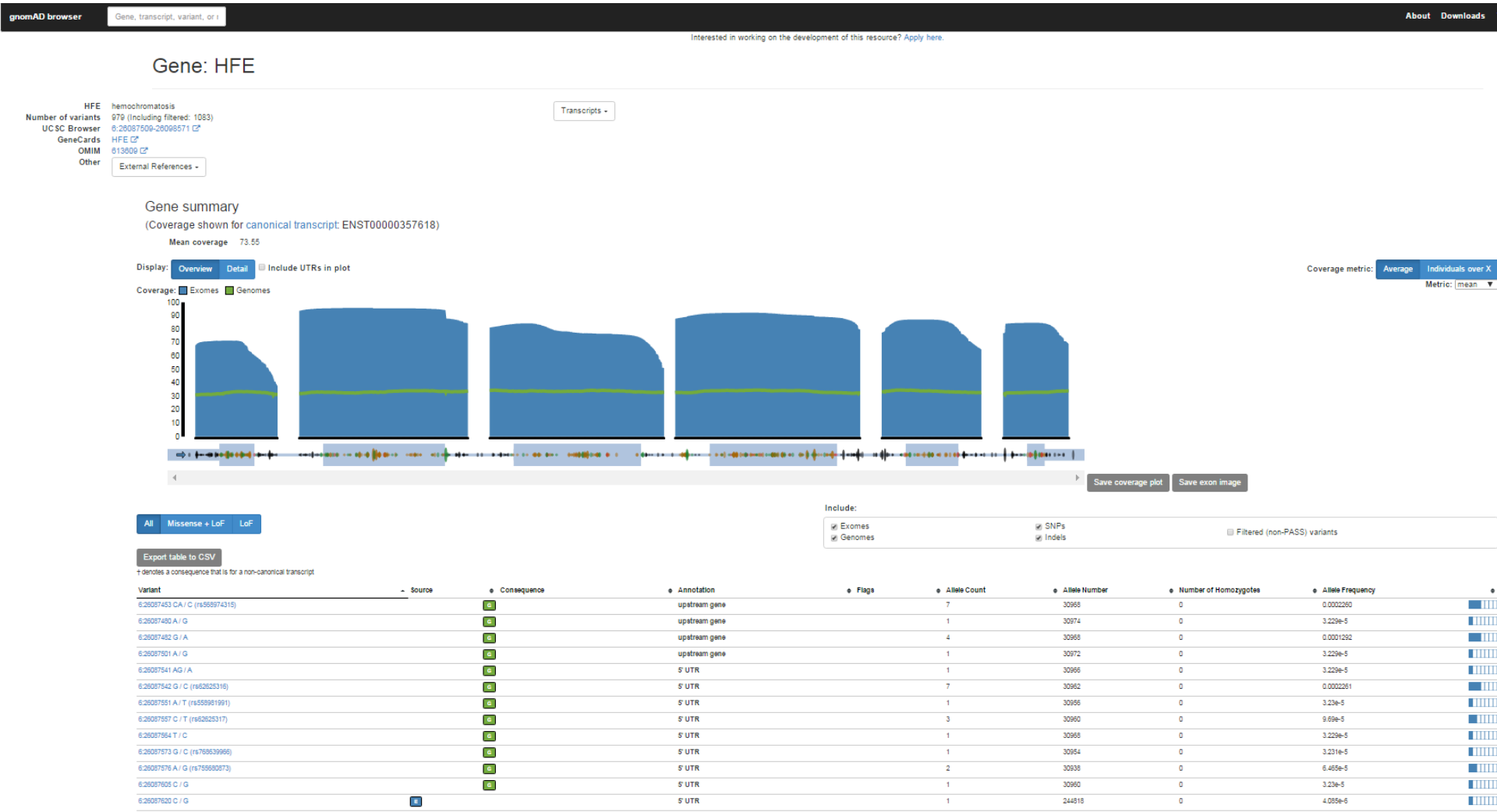
Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (Non-Finnish)	7275	126464	235	0.05753
Other	232	6464	5	0.03589
European (Finnish)	902	25778	20	0.03499
Latino	485	34414	5	0.01409
Ashkenazi Jewish*	119	10152	2	0.01172
African	241	24020	0	0.01003
South Asian	69	30782	0	0.002242
East Asian	3	18868	0	0.0001590
Total	**9326**	**276942**	**267**	**0.03367**

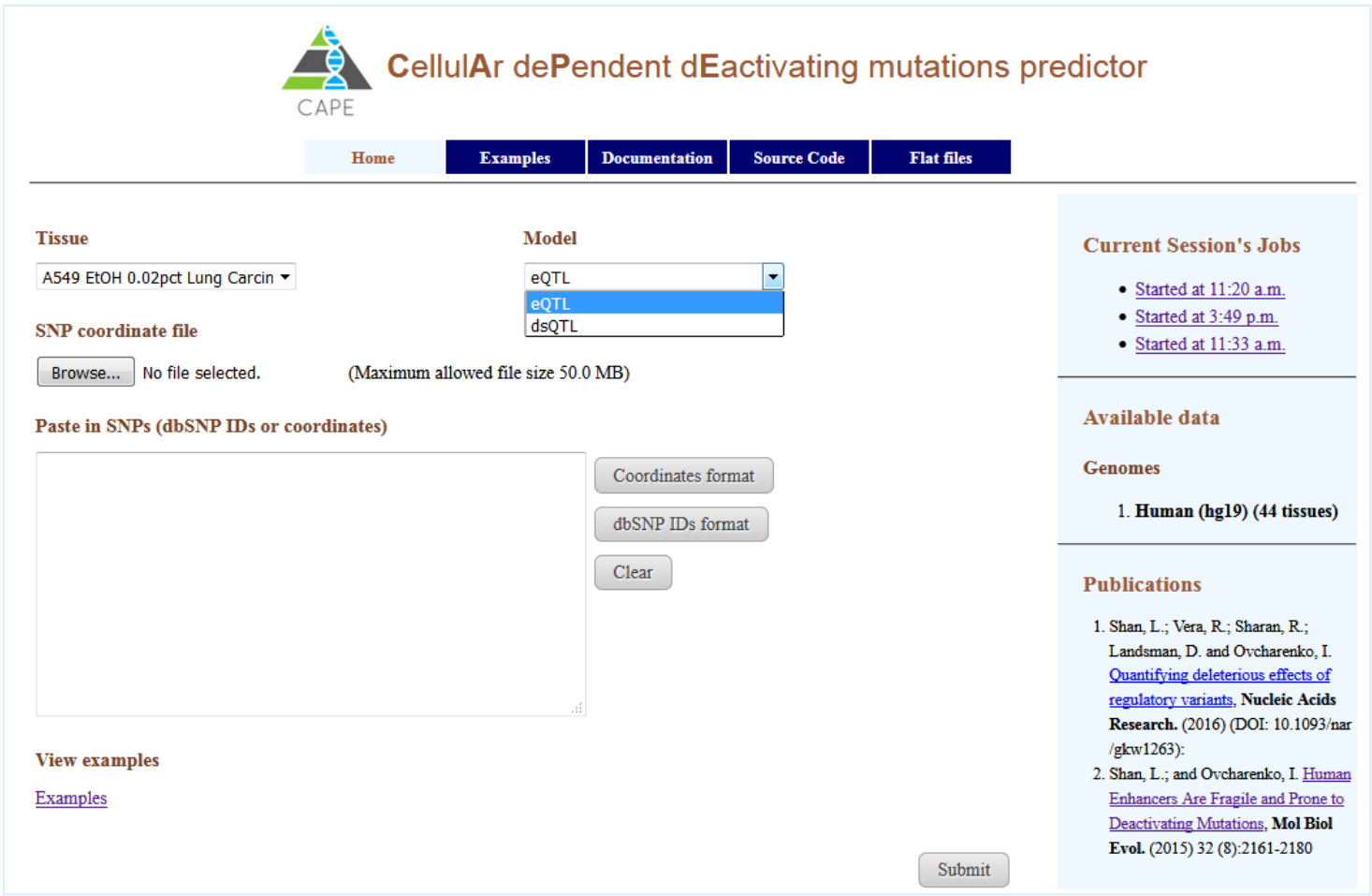
Include: ☑ Exomes ☑ Genomes

* For detailed analysis of Ashkenazi Jewish frequency see the IBD Exomes Browser.

Whole Genome: gnomAD (gene)



Regulatory Variants



Published online 15 December 2016

Nucleic Acids Research, 2017, Vol. 45, No. 5 ***2307–2317***
doi: 10.1093/nar/gkw1263

Quantifying deleterious effects of regulatory variants

Shan Li[1], Roberto Vera Alvarez[1], Roded Sharan[2], David Landsman[1] and Ivan Ovcharenko[1,*]

[1]Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA and [2]School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

FunSeq2 for Known and Unknown Variants

FunSeq2 - A flexible framework to prioritize regulatory mutations from cancer genome sequencing

Analysis | Results | Downloads | Documentation | FAQ | Whole Genome Query

Overview

This tool is specialized to prioritize somatic variants from cancer whole genome sequencing. It contains two components : 1) building data context from various resources; 2) variants prioritization. We provide downloadable scripts for users to customize the data context (found under 'Downloads'). The variants prioritization step is downloadable, and also implemented as a web server (Right Panel), along with the pre-processed data context.

Instructions

✣ Input File - BED or VCF formatted. Click the "green" button to add multiple files. With multiple files, the tool will do recurrent analysis.(Note: for BED format, user can put variants from multiple genomes into one file, see Sample input file .)

✣ Recurrence DB - User can select particular cancer types from the database. The DB will continue to be updated with newly-available WGS data.

✣ Gene List - Option to analyze variants associated with a particular set of genes. Note: Please use Gene Symbols, with one row per gene.

✣ Differential Gene Expression Analysis - Option to detect differentially expressed genes in RNA-Seq data. Two files are needed: an expression file and a class label file. Please refer to Expression input files for instructions on how to prepare these files.

✣ Note: In addition to on-site calculation, we also provide scores for all possible noncoding SNVs of GRCh37/hg19 under 'Downloads' (without annotation and recurrence analysis).

Input File: (only for hg19 SNVs)

Choose File | No file chosen

BED or VCF files as input. Sample input file

Output Format: bed

MAF: 0

Minor allele frequency threshold to filter polymorphisms from 1KG (value 0~1)

Cancer Type from Recurrence DB: Summary table

All Cancer Types

Add a gene list (Optional)

Add differential gene expression analysis (Optional)

Upload

Integrative Annotation of Variants from 1092 Humans: Application to Cancer Genomics

Ekta Khurana[1,2,*], Yao Fu[1,*], Vincenza Colonna[3,4,*], Xinmeng Jasmine Mu[1,*], Hyun Min Kang[5], Tuuli Lappalainen[6,7,8], Andrea Sboner[9,10], Lucas Lochovsky[1], Jieming Chen[1,11], Arif Harmanci[1,2], Jishnu Das[12,13], Alexej Abyzov[1,2], Suganthi Balasubramanian[1,2], Kathryn Beal[14], Dimple Chakravarty[9], Daniel Challis[15], Yuan Chen[3], Declan Clarke[16], Laura Clarke[14], Fiona Cunningham[14], Uday S. Evani[15], Paul Flicek[14], Robert Fragoza[13,17], Erik Garrison[18], Richard Gibbs[15], Zeynep H. Gümüş[10,19], Javier Herrero[14], Naoki Kitabayashi[9], Yong Kong[2,20], Kasper Lage[21,22,23,24,25], Vaja Liluashvili[10,19], Steven M. Lipkin[26], Daniel G. MacArthur[22,27], Gabor Marth[18], Donna Muzny[15], Tune H. Pers[24,28,29], Graham R. S. Ritchie[14], Jeffrey A. Rosenfeld[30,31,32], Cristina Sisu[1,2], Xiaomu Wei[13,26], Michael Wilson[1,33], Yali Xue[3], Fuli Yu[15], 1000 Genomes Project Consortium[†], Emmanouil T. Dermitzakis[6,7,8], Haiyuan Yu[12,13], Mark A. Rubin[9], Chris Tyler-Smith[3,†], Mark Gerstein[1,2,34,‡]

Liverpool Hope University EST. 1844

RVS for Known and Unknown (MS) Variants

RVS | Queries | API | Data sources | About | Contact Us | DIVAS

the Reference Variant Store

The Reference Variant Store holds information on more than 520 million genetic variants using various methods of annotation. Sources for variants in RVS are re-sequencing projects such as the 1000 Genomes, ESP6500, UK10K, TCGA, Scripps Wellderly; clinical annotation databases such as ClinVar and HGMD; and hypothetical, amino-acid changing single-point mutations from dbNSFP. We provide annotations on low level effects, phenotypes and diseases, population frequencies, and predictive functional impact scores.

Available queries

Our query interface is available via the "Queries" menu. Entry points are searches by specific coordinate, dbSNP, or gene; or a comparison of cohorts based on populations in RVS. The first four queries are also wrapped in our REST API.

Query by dbSNP: enter dbSNP IDs (rs-numbers) to search RVS for matching variants.
Query by coordinates: enter specific chromosomal locations with alternate alleles to search RVS for genetic variants; or enter chromosomal regions to search RVS for all variants falling into those regions.
Query by gene: enter one or more genes and, optionally, genetic variants to search RVS.
Query by phenotype: enter a phenotype or disease name, and optionally a gene to search.
Query by cohorts: choose two cohorts from the available datasets and find the differences in allele frequencies with respect to selected genes.

Hakenberg *et al. BMC Bioinformatics* (2016) 17:24
DOI 10.1186/s12859-015-0865-9

BMC Bioinformatics

DATABASE Open Access

CrossMark

Integrating 400 million variants from 80,000 human samples with extensive annotations: towards a knowledge base to analyze disease cohorts

Jörg Hakenberg[1,3*], Wei-Yi Cheng[1,4], Philippe Thomas[2,5], Ying-Chih Wang[1], Andrew V. Uzilov[1] and Rong Chen[1*]

RVS | Queries | API | Data sources | About | Contact Us | DIVAS

RVS queries

Query by coordinates | Query by dbSNP | Query by gene list | Query by phenotype | Query by cohorts (beta)

Paste a list of max. 10 variants using coordinates and ref/alt alleles. Alternatively, enter a chromosomal region to retrieve all variants found in that region.
Example for a single nucleotide variant: `chr7:140453136A>T`
Example for a chromosomal region: `7:140430000-140440000` — we accept regions of up to 10,000bp.

6:32400000-32450000

Limit: ○ exonic ○ non-synonymous ◉ all
Output: ☑ Display only variants observed in 1000KG Ph3, ExAC, ESP6500, UK10K ALSPAC/TWINS, and/or Scripps Wellderly
☐ Display results for the canonical isoform only (where applicable)
Format: ◉ HTML ○ JSON ○ Excel

Search | Click here to paste examples

Liverpool Hope University EST. 1844

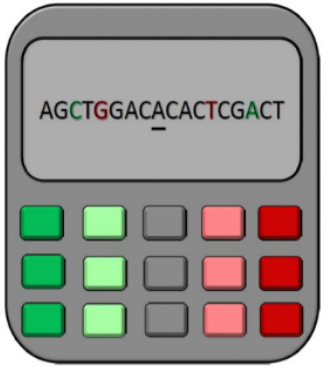
ClinGen Pathogenicity Calculator

ClinGen
Clinical Genome Resource

PATHOGENICITY CALCULATOR

Article describing ClinGen Pathogenicity Calculator is now available online from Genome Medicine.

LOGIN



WHAT IS THE CLINGEN PATHOGENICITY CALCULATOR?

The shift from genetic testing of individual genes to exome and genome sequencing has been accompanied by new challenges in genome interpretation. The American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) have published Standards and Guidelines for the Interpretation of Sequence Variants. To enable wide application of the ACMG/AMP and similar guidelines and the development of collective knowledge by the community, ClinGen has developed the ClinGen Pathogenicity Calculator. By automating the formal reasoning, the Calculator eliminates errors in rule application and makes it possible to automatically calculate provisional conclusions based on latest evidence. Moreover, the Calculator makes reasoning explicit by documenting applicable rules, evidence codes, and links to supporting data. By explicitly communicating the reasoning behind a conclusion about pathogenicity of any specific variant, the Calculator enables critical evaluation of the reasoning and facilitates resolution of conflicting conclusions.

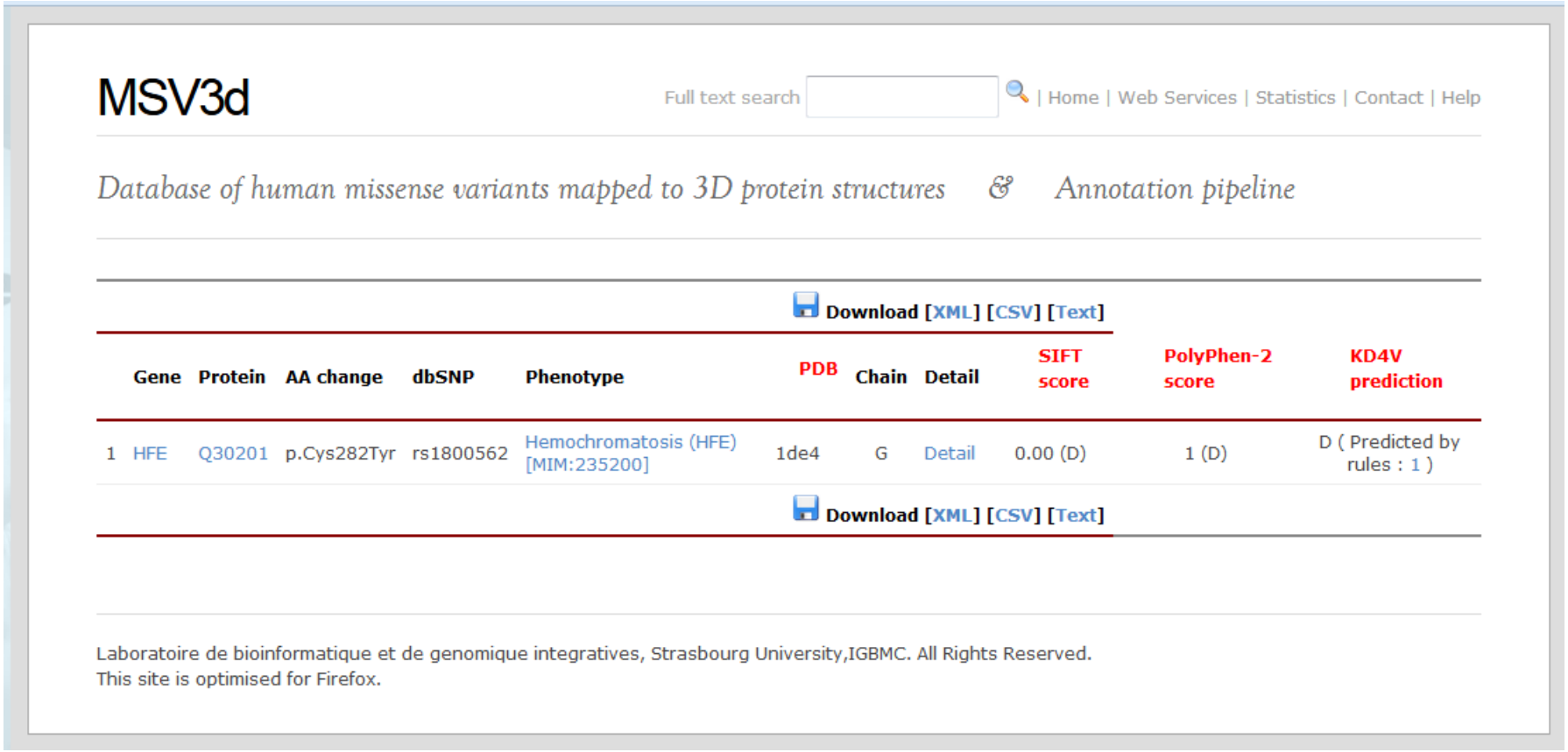
SOFTWARE Open Access

ClinGen Pathogenicity Calculator: a configurable system for assessing pathogenicity of genetic variants

Ronak Y. Patel[1*†], Neethu Shah[1†], Andrew R. Jackson[1], Rajarshi Ghosh[1], Piotr Pawliczek[1], Sameer Paithankar[1], Aaron Baker[1], Kevin Riehle[1], Hailin Chen[1], Sofia Milosavljevic[1], Chris Bizon[2], Shawn Rynearson[3], Tristan Nelson[4], Gail P. Jarvik[9], Heidi L. Rehm[5,6], Steven M. Harrison[5], Danielle Azzariti[5], Bradford Powell[7], Larry Babb[8], Sharon E. Plon[1], Aleksandar Milosavljevic[1*] and on behalf of the ClinGen Resource

Liverpool Hope University EST. 1844

Exonic Variants: Known SNPs



MSV3d

Full text search | Home | Web Services | Statistics | Contact | Help

Database of human missense variants mapped to 3D protein structures & Annotation pipeline

Download [XML] [CSV] [Text]

	Gene	Protein	AA change	dbSNP	Phenotype	PDB	Chain	Detail	SIFT score	PolyPhen-2 score	KD4V prediction
1	HFE	Q30201	p.Cys282Tyr	rs1800562	Hemochromatosis (HFE) [MIM:235200]	1de4	G	Detail	0.00 (D)	1 (D)	D (Predicted by rules : 1)

Download [XML] [CSV] [Text]

Laboratoire de bioinformatique et de genomique integratives, Strasbourg University,IGBMC. All Rights Reserved.
This site is optimised for Firefox.

http://decrypthon.igbmc.fr/msv3d/cgi-bin/humsavar?rsid=rs1800562

dbNSFP

OXFORD

Human Molecular Genetics, 2015, Vol. 24, No. 8 2125–2137

doi: 10.1093/hmg/ddu733
Advance Access Publication Date: 30 December 2014
Original Article

ORIGINAL ARTICLE

Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies

Chengliang Dong[1,2,†], Peng Wei[4,6,†], Xueqiu Jian[5], Richard Gibbs[7], Eric Boerwinkle[4,5,7], Kai Wang[1,2,3,*] and Xiaoming Liu[4,5,*]

Abstract

Accurate deleteriousness prediction for nonsynonymous variants is crucial for distinguishing pathogenic mutations from background polymorphisms in whole exome sequencing (WES) studies. Although many deleteriousness prediction methods have been developed, their prediction results are sometimes inconsistent with each other and their relative merits are still unclear in practical applications. To address these issues, we comprehensively evaluated the predictive performance of 18 current deleteriousness-scoring methods, including 11 function prediction scores (PolyPhen-2, SIFT, MutationTaster, Mutation Assessor, FATHMM, LRT, PANTHER, PhD-SNP, SNAP, SNPs&GO and MutPred), 3 conservation scores (GERP++, SiPhy and PhyloP) and 4 ensemble scores (CADD, PON-P, KGGSeq and CONDEL). We found that FATHMM and KGGSeq had the highest discriminative power among independent scores and ensemble scores, respectively. Moreover, to ensure unbiased performance evaluation of these prediction scores, we manually collected three distinct testing datasets, on which no current prediction scores were tuned. In addition, we developed two new ensemble scores that integrate nine independent scores and allele frequency. Our scores achieved the highest discriminative power compared with all the deleteriousness prediction scores tested and showed low false-positive prediction rate for benign yet rare nonsynonymous variants, which demonstrated the value of combining information from multiple orthologous approaches. Finally, to facilitate variant prioritization in WES studies, we have pre-computed our ensemble scores for 87 347 044 possible variants in the whole-exome and made them publicly available through the ANNOVAR software and the dbNSFP database.

dbNSFP

DATABASES

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs

Xiaoming Liu,[1,2]* Chunlei Wu,[3] Chang Li,[2] and Eric Boerwinkle[1,2,4]

Jpopgen

Search this site

Navigation

Jpopgen - a collection of tools for DNA sequence analyses

Sitemap

About the maintainer

This site is maintained by:

Xiaoming Liu, Ph.D.

Assistant Professor,
Human Genetics Center,
School of Public Health,
The University of Texas Health Science Center at Houston.

E-mail: xmliu.uth{at}gmail.com

Homepage

dbNSFP

INTRODUCTION:

dbNSFP is a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants (nsSNVs) in the human genome. Its current version is based on the Gencode release 22 / Ensembl version 79 and includes a total of 83,422,341 nsSNVs and ssSNVs (splicing-site SNVs). It compiles prediction scores from 20 prediction algorithms (SIFT, Polyphen2-HDIV, Polyphen2-HVAR, LRT, MutationTaster2, MutationAssessor, FATHMM, MetaSVM, MetaLR, CADD, VEST3, PROVEAN, FATHMM-MKL coding, fitCons, DANN, GenoCanyon, Eigen coding, Eigen-PC, M-CAP, REVEL, MutPred), 6 conservation scores (PhyloP x 2, phastCons x 2, GERP++ and SiPhy) and other related information including allele frequencies observed in the 1000 Genomes Project phase 3 data, UK10K cohorts data, ExAC consortium data and the NHLBI Exome Sequencing Project ESP6500 data, various gene IDs from different databases, functional descriptions of genes, gene expression and gene interaction information, etc.

Some dbNSFP contents (may not be up-to-date though) can also be accessed through variant tools, ANNOVAR, KGGSeq, UCSC Genome Browser's Variant Annotation Integrator, Ensembl Variant Effect Predictor, SnpSift and HGMD. Please cite our papers (see below) if you used dbNSFP contents through those tools.

Please note some component score/content of dbNSFP has specific requirements or licence for non-academic usage. dbNSFP does not grant the non-academic usage of those scores/contents, so please contact the original score/content providers for that purpose.

Please join our Email group for news and updates from dbNSFP.

For whole genome annotation, we recommend our whole genome annotation pipeline WGSA, in which dbNSFP is a component resource.

We thank Drs. CS (Jonathan) Liu from Softgenetics and Kai Wang from USC for providing hosting space.

We welcome developers of functional prediction methods to provide their predictions and scores to the database. Please contact Dr. Xiaoming Liu (xmliu.uth{at}gmail.com).

CITATION:

1. Liu X, Jian X, and Boerwinkle E. 2011. dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. *Human Mutation*. 32:894-899.
2. Liu X, Jian X, and Boerwinkle E. 2013. dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. *Human Mutation*. 34:E2393-E2402.
3. Liu X, Wu C, Li C and Boerwinkle E. 2016. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-synonymous and Splice Site SNVs. *Human Mutation*. 37:235-241. [preprint]

dbNSFP

DATABASES

Human Mutation
OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs

Xiaoming Liu,[1,2]* Chunlei Wu,[3] Chang Li,[2] and Eric Boerwinkle[1,2,4]

Index of dbNSFP prediction files

Name	Size	Last Modified	
dbnsfp_chr1_mutpred.txt.gz	195 MB	8/31/16	4:55:00 PM
dbnsfp_chr2_mutpred.txt.gz	143 MB	8/31/16	4:56:00 PM
dbnsfp_chr3_mutpred.txt.gz	108 MB	8/31/16	4:57:00 PM
dbnsfp_chr4_mutpred.txt.gz	75 MB	8/31/16	4:58:00 PM
dbnsfp_chr5_mutpred.txt.gz	89 MB	8/31/16	4:58:00 PM
dbnsfp_chr6_mutpred.txt.gz	97 MB	8/31/16	4:59:00 PM
dbnsfp_chr7_mutpred.txt.gz	90 MB	8/31/16	5:00:00 PM
dbnsfp_chr8_mutpred.txt.gz	66 MB	8/31/16	5:00:00 PM
dbnsfp_chr9_mutpred.txt.gz	79 MB	8/31/16	5:01:00 PM
dbnsfp_chr10_mutpred.txt.gz	76 MB	8/31/16	4:48:00 PM
dbnsfp_chr11_mutpred.txt.gz	114 MB	8/31/16	4:49:00 PM
dbnsfp_chr12_mutpred.txt.gz	100 MB	8/31/16	4:50:00 PM
dbnsfp_chr13_mutpred.txt.gz	34 MB	8/31/16	4:50:00 PM
dbnsfp_chr14_mutpred.txt.gz	61 MB	8/31/16	4:50:00 PM
dbnsfp_chr15_mutpred.txt.gz	64 MB	8/31/16	4:51:00 PM
dbnsfp_chr16_mutpred.txt.gz	79 MB	8/31/16	4:51:00 PM
dbnsfp_chr17_mutpred.txt.gz	106 MB	8/31/16	4:52:00 PM
dbnsfp_chr18_mutpred.txt.gz	30 MB	8/31/16	4:52:00 PM
dbnsfp_chr19_mutpred.txt.gz	122 MB	8/31/16	4:53:00 PM
dbnsfp_chr20_mutpred.txt.gz	47 MB	8/31/16	4:55:00 PM
dbnsfp_chr21_mutpred.txt.gz	20 MB	8/31/16	4:55:00 PM
dbnsfp_chr22_mutpred.txt.gz	41 MB	8/31/16	4:55:00 PM
dbnsfp_chrX_mutpred.txt.gz	72 MB	8/31/16	5:01:00 PM
dbnsfp_chrY_mutpred.txt.gz	566 KB	8/31/16	5:01:00 PM
README	4 KB	9/1/16	2:55:00 PM

WGSA

DATABASES

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs

Xiaoming Liu,[1,2]* Chunlei Wu,[3] Chang Li,[2] and Eric Boerwinkle[1,2,4]

Jpopgen

Search this site

Navigation

Jpopgen - a collection of tools for DNA sequence analyses

Sitemap

About the maintainer

This site is maintained by:

Xiaoming Liu, Ph.D.

Assistant Professor,
Human Genetics Center,
School of Public Health,
The University of Texas Health Science Center at Houston.

E-mail: xmliu.uth{at}gmail.com

Homepage

dbNSFP

INTRODUCTION:

dbNSFP is a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants (nsSNVs) in the human genome. Its current version is based on the Gencode release 22 / Ensembl version 79 and includes a total of 83,422,341 nsSNVs and ssSNVs (splicing-site SNVs). It compiles prediction scores from 20 prediction algorithms (SIFT, Polyphen2-HDIV, Polyphen2-HVAR, LRT, MutationTaster2, MutationAssessor, FATHMM, MetaSVM, MetaLR, CADD, VEST3, PROVEAN, FATHMM-MKL coding, fitCons, DANN, GenoCanyon, Eigen coding, Eigen-PC, M-CAP, REVEL, MutPred), 6 conservation scores (PhyloP x 2, phastCons x 2, GERP++ and SiPhy) and other related information including allele frequencies observed in the 1000 Genomes Project phase 3 data, UK10K cohorts data, ExAC consortium data and the NHLBI Exome Sequencing Project ESP6500 data, various gene IDs from different databases, functional descriptions of genes, gene expression and gene interaction information, etc.

Some dbNSFP contents (may not be up-to-date though) can also be accessed through variant tools, ANNOVAR, KGGSeq, UCSC Genome Browser's Variant Annotation Integrator, Ensembl Variant Effect Predictor, SnpSift and HGMD. Please cite our papers (see below) if you used dbNSFP contents through those tools.

Please note some component score/content of dbNSFP has specific requirements or licence for non-academic usage. dbNSFP does not grant the non-academic usage of those scores/contents, so please contact the original score/content providers for that purpose.

Please join our Email group for news and updates from dbNSFP.

For whole genome annotation, we recommend our whole genome annotation pipeline WGSA, in which dbNSFP is a component resource.

- WGSA: WGSA is an annotation pipeline for human genome re-sequencing studies, to facilitate the functional annotation step of whole genome sequencing (WGS). Currently WGSA supports the annotation of SNVs and indels locally without remote database requests, allowing it to scale up for large WGS studies.

1. Liu X, White S, Peng B, Johnson AD, Brody JA, Li AH, Huang Z, Carroll A, Wei P, Gibbs R, Klein RJ and Boerwinkle E. 2016. WGSA: an annotation pipeline for human genome sequencing studies. *Journal of Medical Genetics* 53:111-112.

WGSA: https://sites.google.com/site/jpopgen/wgsa

Liverpool Hope University EST. 1844

WGSA

JMG

WGSA: an annotation pipeline for human genome sequencing studies

Xiaoming Liu, Simon White, Bo Peng, Andrew D Johnson, Jennifer A Brody, Alexander H Li, Zhuoyi Huang, Andrew Carroll, Peng Wei, Richard Gibbs, Robert J Klein and Eric Boerwinkle

Jpopgen

Search this site

Navigation

Jpopgen - a collection of tools for DNA sequence analyses

Sitemap

About the maintainer

This site is maintained by:

Xiaoming Liu, Ph.D.

Assistant Professor,
Human Genetics Center,
School of Public Health,
The University of Texas Health Science Center at Houston.

E-mail: xmliu.uth{at}gmail.com

Homepage

WGS Annotator (WGSA) - an annotation pipeline for human genome re-sequencing studies

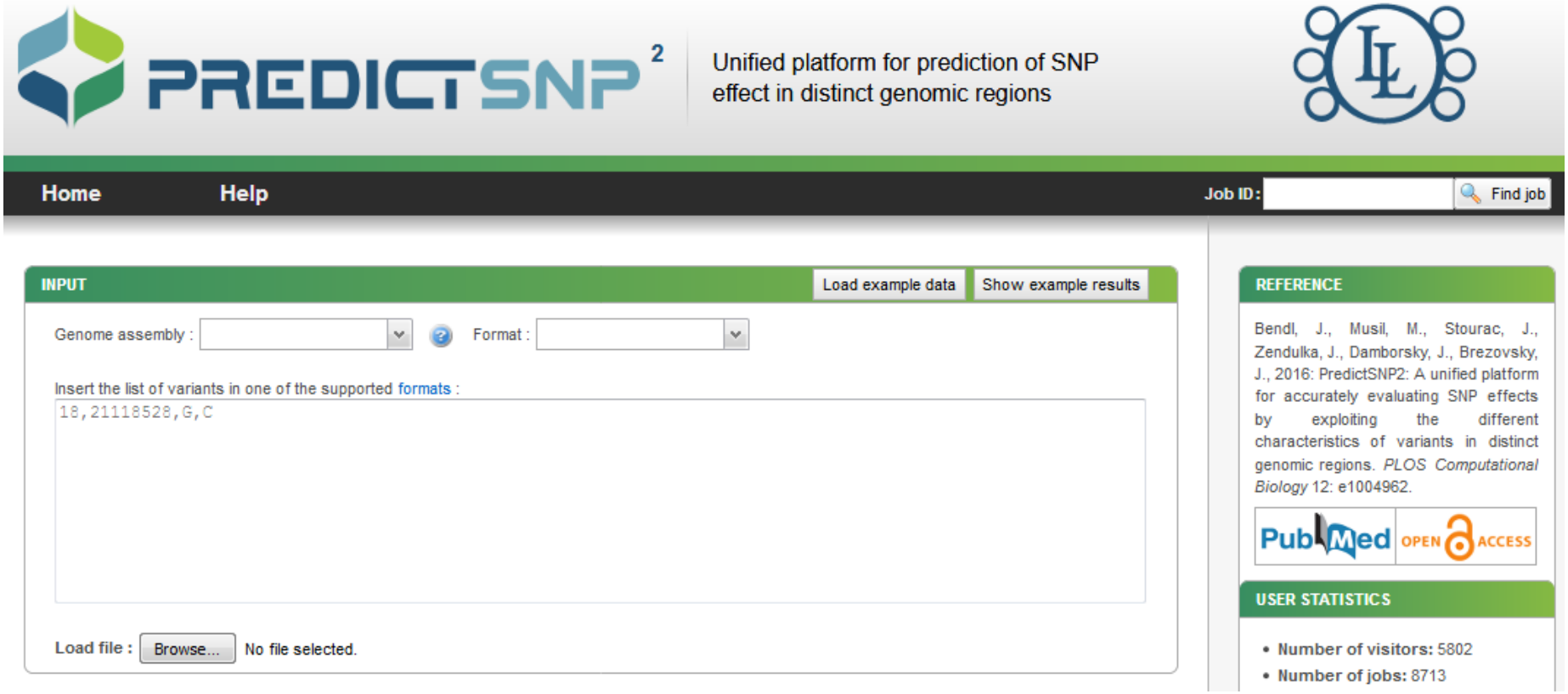
WGSA is an annotation pipeline for human genome re-sequencing studies, to facilitate the functional annotation step of whole genome sequencing (WGS). Currently WGSA supports the annotation of SNVs and indels locally without remote database requests, allowing it to scale up for large WGS studies.

For gene-model based annotation, WGSA integrates the outputs from three annotation tools (ANNOVAR, SnpEff and VEP) for RefSeq and Ensembl GENCODE gene models, plus ANNOVAR outputs using UCSC knowGene model, and provides a summary of variant consequences from the seven annotation results. To further speed up the process for large-scale WGS studies, we have pre-computed annotations for all potential human SNVs (a total of 8,584,031,106) based on human reference hg19 and use it as a local database (in the resources/precomputed folder). For SNV-centric resources, WGSA integrated 13 functional prediction scores (CADD, FATHMM-MKL, Funseq, Funseq2, RegulomeDB, DANN, fitCons x 4, genoCanyon, Eigen, Eigen-PC), 8 conservation scores (GERP++, PhyloP x 3, phastCons x 3, SyPhy), allele frequencies from 4 large-scale re-sequencing studies (1000G, EP6500, ExAC, UK10K), variants in 4 disease related databases (ClinVar, COSMIC, GWAS_catalog, GRASP2), among others (see list of resources). For regulatory region-centric resources, WGSA integrated cell type specific transcription factor binding sites, DNAse1 hypersensitivity regions and chromosome activity predictions from multiple epigenomics projects (see list of resources). WGSA also contains rich functional annotations for non-synonymous SNVs and genes from our dbNSFP database.

WGSA is provided both as an Amazon Machine Image (AMI) ready to run out-of-the-box and a downloadable version. WGSA and its resources are freely available for academic usage. Licenses are required for non-academic usage some of the resources, such as ANNOVAR, Polyphen2, CADD, DANN and VEST3 (in dbNSFP). WGSA does not grant the non-academic usage of those resources, so please contact the original content provider for that purpose.

You can sign up for our email list for future release announcements.
Please report any bug you found or any suggestions or comments to xmliu.uth{at}gmail.com or xiaoming.liu{at}uth.tmc.edu.

PredictSNP2

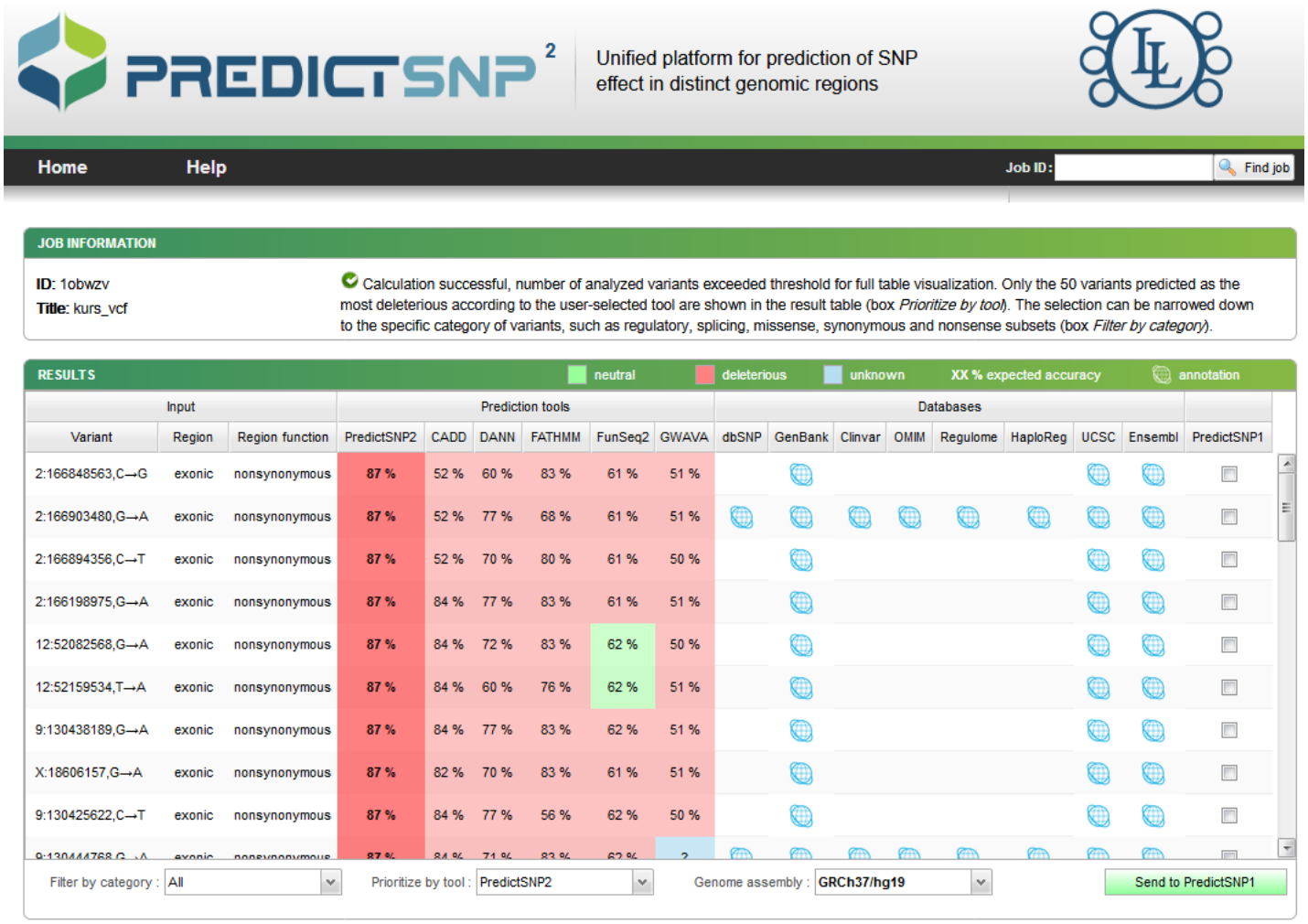


PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions

Jaroslav Bendl, Miloš Musil, Jan Štourač, Jaroslav Zendulka, Jiří Damborský, Jan Brezovský

Liverpool Hope University

PredictSNP2



PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions

Jaroslav Bendl, Miloš Musil, Jan Štourač, Jaroslav Zendulka, Jiří Damborský, Jan Brezovský

Liverpool Hope University

… Looking forward

Day 1
Genome Biology in 2017
Bioinformatics Tools in Epigenetics
Pathogenicity Assessment of DNA Sequence Mutations
Introduction to Galaxy
Massive Data Sources
Practical

YOUR FUTURE
STARTS WITH HOPE

Liverpool Hope University EST. 1844

YOUR FUTURE
STARTS WITH HOPE

Liverpool Hope University EST. 1844