

LOF & NMD SnpEff

Pablo Cingolani

“Gold standard” methods

- LOF as specified in:
(Science, 2012)

A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes

Daniel G. MacArthur,^{1,2*} Suganthi Balasubramanian,^{3,4} Adam Frankish,¹ Ni Huang,¹ James Morris,¹ Klaudia Walter,¹ Luke Jostins,¹ Lukas Habegger,^{3,4} Joseph K. Pickrell,⁵ Stephen B. Montgomery,^{6,7} Cornelis A. Albers,^{1,8} Zhengdong D. Zhang,⁷ Donald F. Conrad,¹⁰ Gerton Lunter,¹¹ Hancheng Zheng,¹² Qasim Ayub,¹ Mark A. DePristo,¹³ Eric Banks,¹³ Min Hu,¹ Robert E. Handsaker,^{13,14} Jeffrey A. Rosenfeld,¹⁵ Menachem Fromer,¹³ Mike Jin,³ Ximeng Jasmine Mu,^{3,4} Ekta Khurana,^{3,4} Kai Ye,¹⁶ Mike Kay,¹ Gary Ian Saunders,¹ Marie-Marthe Suner,³ Toby Hunt,¹ If H. A. Barnes,¹ Clara Amid,^{1,17} Denise R. Carvalho-Silva,¹ Alexandra H. Bignell,¹ Catherine Snow,¹ Bryndis Yngvadottir,¹ Suzannah Bumpstead,¹ David N. Cooper,¹⁸ Yali Xue,¹ Irene Gallego Romero,^{1,5} 1000 Genomes Project Consortium, Jun Wang,¹² Yingrui Li,¹² Richard A. Gibbs,¹⁹ Steven A. McCarroll,^{13,14} Emmanouil T. Dermitzakis,⁷ Jonathan K. Pritchard,^{5,20} Jeffrey C. Barrett,¹ Jennifer Harrow,¹ Matthew E. Hurles,¹ Mark B. Gerstein,^{3,4,21†} Chris Tyler-Smith[†]

- NMD as specified in:
(Nature, 2004)

NONSENSE-MEDIATED mRNA DECAY: SPLICING, TRANSLATION AND mRNP DYNAMICS

Lynne E. Maquat

How to use

Simply add '-lof' to the command line

```
java -Xmx4g -jar snpEff.jar -v \
    -lof \
    GRCh37.68 \
    file.vcf.gz > file.eff.vcf
```

Output

- SnpEff adds 'LOF' and 'MND' tags to INFO fields (column 8 in VCF format).
- LOF and NMD tag format:

Gene | ID | num_transcripts | percent_affected

Where:

- Gene : Gene name
- ID : Gene ID (usually ENSEMBL)
- Num_transcripts : Number of transcripts in this gene
- percent_affected : Percentage of transcripts affected by this variant.

Output examples: LOF

Given an effect

EFF=..., SPLICE_SITE_DONOR(HIGH||||639|ILDR2|protein_coding|CODING|ENST00000271417|1)

The corresponding LOF tag is

LOF=ILDR2|ENSG00000143195|7|1.00

- Gene name : **ILDR2**
- Gene ID : **ENSG00000143195**
- Num_transcripts : There are **7** transcripts in this gene
- percent_affected : **100%** of transcripts are affected.

Output examples: NMD

Given an effect

EFF=STOP_GAINED(HIGH|NONSENSE|taT/taG|Y269*|321|RHD|protein_coding|CODING|ENST00000454452|)...

The corresponding NMD tag is

NMD=RHD|ENSG00000187010|10|0.80

- Gene name : **RHD**
- Gene ID : **ENSG00000187010**
- Num_transcripts : There are **10** transcripts in this gene
- percent_affected : **80%** of transcripts are affected.

How to obtain a list of LOF variants

1) Calculate effects using '-lof' command line option

```
java -Xmx4g -jar snpEff.jar -v -lof GRCh37.66 file.vcf.gz > file.eff.vcf
```

2) Filter using SnpSift, e.g.:

- a. Get all entries having LOF with genes that have more than 50% of transcripts affected

```
cat file.eff.vcf | java -jar SnpSift.jar filter "LOF[*].PERC > 0.5"
```

- b. Get all entries having NMD with genes that have more than 3 transcripts

```
cat file.eff.vcf | java -jar SnpSift.jar filter "NMD[*].NUMTR > 3"
```