

# LOF & NMD

## SnpEff

Pablo Cingolani

Rob Sladek

# “Gold standard” methods

- LOF as specified in:  
(Science, 2012)
- NMD as specified in:  
(Nature, 2004)

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

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

## **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"
```