

Clinvar_predictions

Svetlana Milrud

09 05 2022

Cninvar + Gnomad

```
predictions<-read_excel("~/Clinvar_eve.xlsx")
```

Control sample

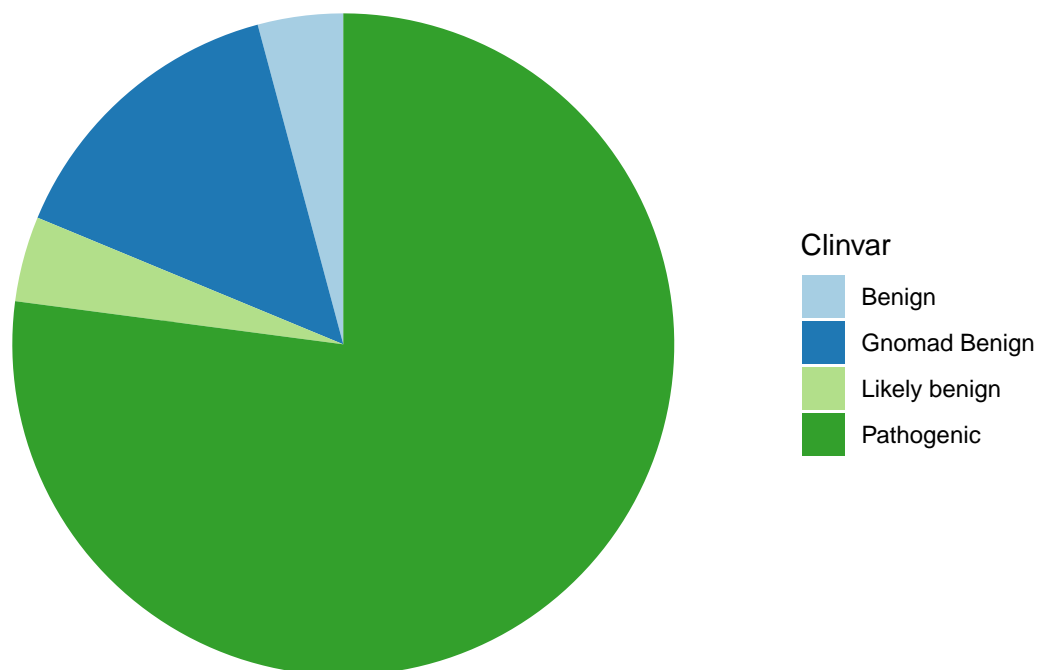
```
table(predictions$ClinVar)
```

```
##
##      Benign Gnomad Benign Likely benign      Pathogenic
##           2           7           2           37
```

Because of small number of Benign variants in ClinVar database, variants with higher allele frequency were chosen from GnomAD as GnomAD Benign

| | A | B | C | D | E | F | G | H | I |
|----|-------|----------------|------------------|-----------------------|---------------|--------------|----------|-------------------------|---|
| 1 | Chrom | Transcript Con | VEP Annotation | ClinVar Clinical Sig | ClinVar Varia | Allele Count | Allele F | Allele Frequency | |
| 2 | 99 | c.1173A>G | missense_variant | Benign | 135038 | 18372 | 279610 | 0.06570580451342942 | |
| 3 | 169 | c.2198A>G | missense_variant | Benign | 456537 | 117 | 279944 | 0.0004179407310033435 | |
| 4 | 165 | c.2080G>T | missense_variant | Uncertain significano | 572793 | 38 | 280572 | 0.00013543760603338894 | |
| 5 | 71 | c.895A>G | missense_variant | Likely benign | 1157452 | 31 | 249486 | 0.00012425546924476724 | |
| 6 | 46 | c.548A>G | missense_variant | Likely benign | 1129752 | 28 | 248800 | 0.00011254019292604502 | |
| 7 | 96 | c.1130C>G | missense_variant | Uncertain significano | 403909 | 20 | 280668 | 0.00007125856884290337 | |
| 8 | 11 | c.140A>G | missense_variant | Uncertain significano | 403906 | 17 | 280188 | 0.00006067354776078919 | |
| 9 | 137 | c.1544A>G | missense_variant | Uncertain significano | 403903 | 13 | 235744 | 0.00005514456359440749 | |
| 10 | 33 | c.436G>A | missense_variant | Uncertain significano | 526641 | 12 | 249464 | 0.000048103133117403714 | |
| 11 | 16 | c.178C>A | missense_variant | Conflicting interpre | 526635 | 12 | 279824 | 0.000042884098576247925 | |
| 12 | 30 | c.400C>T | missense_variant | Uncertain significano | 456546 | 10 | 280784 | 0.000035614564932474785 | |
| 13 | 145 | c.1749G>A | missense_variant | Uncertain significano | 999157 | 1 | 31324 | 0.00003192440301366364 | |
| 14 | 178 | c.2296C>G | missense_variant | | | 1 | 31344 | 0.000031904032669729455 | |
| 15 | 172 | c.2217G>T | missense_variant | | | 1 | 31354 | 0.00003189385724309498 | |
| 16 | 142 | c.1651C>G | missense_variant | | | 1 | 31356 | 0.00003189182293659906 | |
| 17 | 171 | c.2215A>G | missense_variant | | | 1 | 31360 | 0.00003188775510204082 | |
| 18 | 47 | c.565C>A | missense_variant | | | 1 | 31372 | 0.00003187555782226189 | |
| 19 | 181 | c.2392G>A | missense_variant | Uncertain significano | 971035 | 1 | 31372 | 0.00003187555782226189 | |
| 20 | 194 | c.2655C>G | missense_variant | | | 1 | 31372 | 0.00003187555782226189 | |
| 21 | 24 | c.265C>T | missense_variant | | | 1 | 31384 | 0.00003186336986999745 | |
| 22 | 66 | c.832A>G | missense_variant | | | 1 | 31386 | 0.000031861339450710506 | |
| 23 | 103 | c.1199C>T | missense_variant | | | 1 | 31388 | 0.00003185930929017459 | |
| 24 | 131 | c.1514G>A | missense_variant | | | 1 | 31390 | 0.00003185727938834024 | |
| 25 | 69 | c.850A>G | missense_variant | | | 1 | 31394 | 0.00003185322036057845 | |
| 26 | 189 | c.2539C>G | missense_variant | | | 1 | 31398 | 0.000031849162367029746 | |
| 27 | 74 | c.905T>C | missense_variant | | | 1 | 31400 | 0.00003184713375796178 | |
| 28 | 125 | c.1484A>C | missense_variant | | | 1 | 31402 | 0.0000318451054072989 | |
| 29 | 155 | c.1880A>G | missense_variant | | | 6 | 248084 | 0.00002418535657277374 | |
| 30 | 120 | c.1438T>C | missense_variant | | | 6 | 249336 | 0.0000240639137549331 | |
| 31 | 133 | c.1528C>A | missense_variant | Uncertain significano | 456529 | 6 | 280532 | 0.000021387934353300158 | |
| 32 | 27 | c.361A>G | missense_variant | Uncertain significano | 246688 | 6 | 280582 | 0.000021384123001475505 | |
| 33 | 195 | c.2667A>G | missense_variant | Uncertain significano | 640626 | 5 | 247484 | 0.00002020332627563802 | |
| 34 | 62 | c.817A>G | missense_variant | Uncertain significano | 571273 | 5 | 248150 | 0.00002014910336490026 | |
| 35 | 179 | c.2309G>A | missense_variant | | | 5 | 248688 | 0.000020105513736086984 | |
| 36 | 146 | c.1878A>C | missense_variant | Uncertain significano | 403911 | 5 | 248814 | 0.0000200953333533533 | |

Figure 1: Variants from GnomAD: first 10 were taken



Calculating Statistics - Sensitivity, Specificity, Accuracy

SF - python script ([https://github.com/Svetave/PIK3CA/blob/main/SF%20for%20clinvar%20variants%20\(PIK3CA\).ipynb](https://github.com/Svetave/PIK3CA/blob/main/SF%20for%20clinvar%20variants%20(PIK3CA).ipynb))

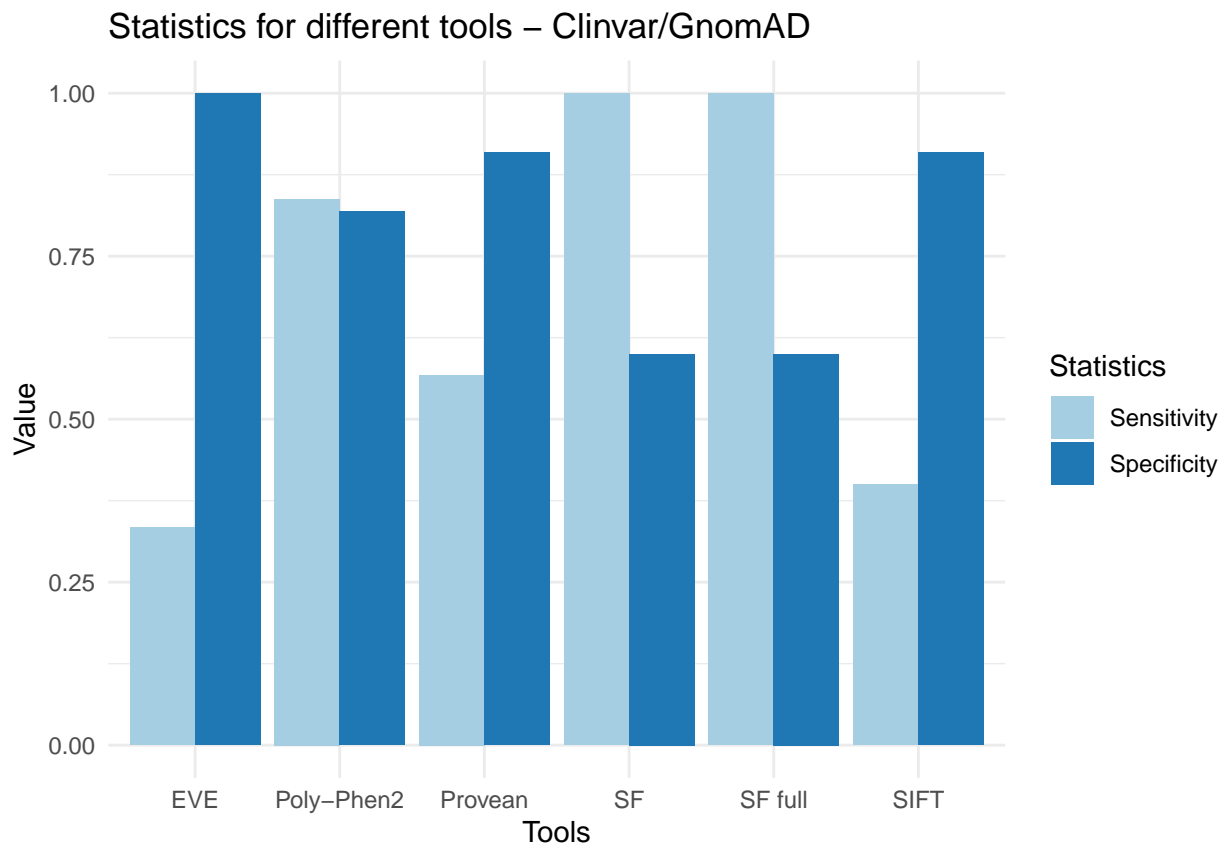
- For *SF* only 306 sequences were taken from final multiple alignment (312 sequences)
- For *SF full* all 312 sequences were taken from final multiple alignment

Statistics - Sensitivity, Specificity

| Statistics | Tools | Value |
|-------------|------------|-----------|
| Sensitivity | Provean | 0.5675676 |
| Specificity | Provean | 0.9090909 |
| Sensitivity | Poly-Phen2 | 0.8378378 |
| Specificity | Poly-Phen2 | 0.8181818 |
| Sensitivity | SIFT | 0.4000000 |
| Specificity | SIFT | 0.9090909 |
| Sensitivity | EVE | 0.3333333 |
| Specificity | EVE | 1.0000000 |
| Sensitivity | SF | 1.0000000 |
| Specificity | SF | 0.6000000 |
| Sensitivity | SF full | 1.0000000 |
| Specificity | SF full | 0.6000000 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| XP_027025573.1 LOW QUALITY PROTEIN: PIK3CA Tachysurus | 992 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_034065896.1 PIK3CA isoform-like isoform X1 Gymnodraco | 992 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_046713774.1 PIK3CA Silurus meridionalis | 992 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_011941631.1 PREDICTED: LOW QUALITY PROTEIN: PIK3CA | 992 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_019388065.1 PREDICTED: PIK3CA Crocodylus porosus | 944 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_044217278.1 PIK3CA isoform-like isoform X2 Thunnus alb: | 992 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | - | - |
| XP_031519421.1 PIK3CA isoform X2 Papio anubis | 929 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | K | S | F | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| XP_033067648.1 PIK3CA isoform X2 Trachypithecus francoisi | 929 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | K | S | F | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| XP_006787079.1 PIK3CA isoform-like Neolamprologus brichar | 937 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_041261722.1 PIK3CA isoform X2 Onychostruthus taczano | 870 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_026967239.1 PIK3CA isoform X2 Lagenorhynchus obliquid | 835 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| XP_029563208.1 PIK3CA isoform X2 Salmo trutta | 774 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_004479033.1 PIK3CA Dasypus novemcinctus | 715 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_010180872.1 PREDICTED: PIK3CA, partial Mesitornis unico | 832 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| XP_012494777.1 PREDICTED: LOW QUALITY PROTEIN: PIK3CA | 868 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_015349795.1 PREDICTED: PIK3CA Marmota marmota mar | 918 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_006892248.1 PREDICTED: PIK3CA Elephantulus edwardii | 733 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| XP_010134443.1 PREDICTED: PIK3CA, partial Buceros rhinoce | 765 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| XP_035971787.1 PIK3CA, partial Halichoerus grypus | 619 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_010214201.1 PREDICTED: PIK3CA, partial Tinamus g | 637 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| XP_029651546.1 LOW QUALITY PROTEIN: PIK3CA isoform | 859 | K | H | A | K | L | F | I | T | L | F | T | L | M | L | S | C | G | I | P | E | L | Q | S | M | D | D | I | G | Y | L | R | K | T | L | A | V | E | K | S | E | Q | E | A |
| XP_013082251.1 PREDICTED: PIK3CA isoform-like isoform | 850 | K | H | A | N | L | F | I | T | L | F | T | M | M | L | Q | C | G | I | P | E | L | Q | S | L | D | D | I | G | Y | L | R | K | T | L | A | V | E | K | T | E | E | K | A |
| XP_036914671.1 PIK3CA Sturnira hondurensis | 484 | Q | H | A | N | L | F | I | N | L | F | S | M | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | A | L | D | K | T | E | Q | E | A |
| XP_024208688.1 LOW QUALITY PROTEIN: PIK3CA isoform | 423 | Q | H | A | N | L | F | I | N | L | F | S | V | M | L | G | S | G | M | P | E | L | Q | S | C | D | D | I | A | Y | I | R | K | T | L | V | L | D | K | T | E | Q | E | A |
| XP_034805214.1 LOW QUALITY PROTEIN: PIK3CA isoform | 423 | Q | H | A | N | L | F | I | N | L | F | S | V | M | L | G | S | G | M | P | E | L | Q | S | F | D | D | I | A | Y | I | R | K | T | L | V | L | D | K | T | E | Q | E | A |

Figure 2: Removed sequences are shown in bold(for SF)



EVE model:

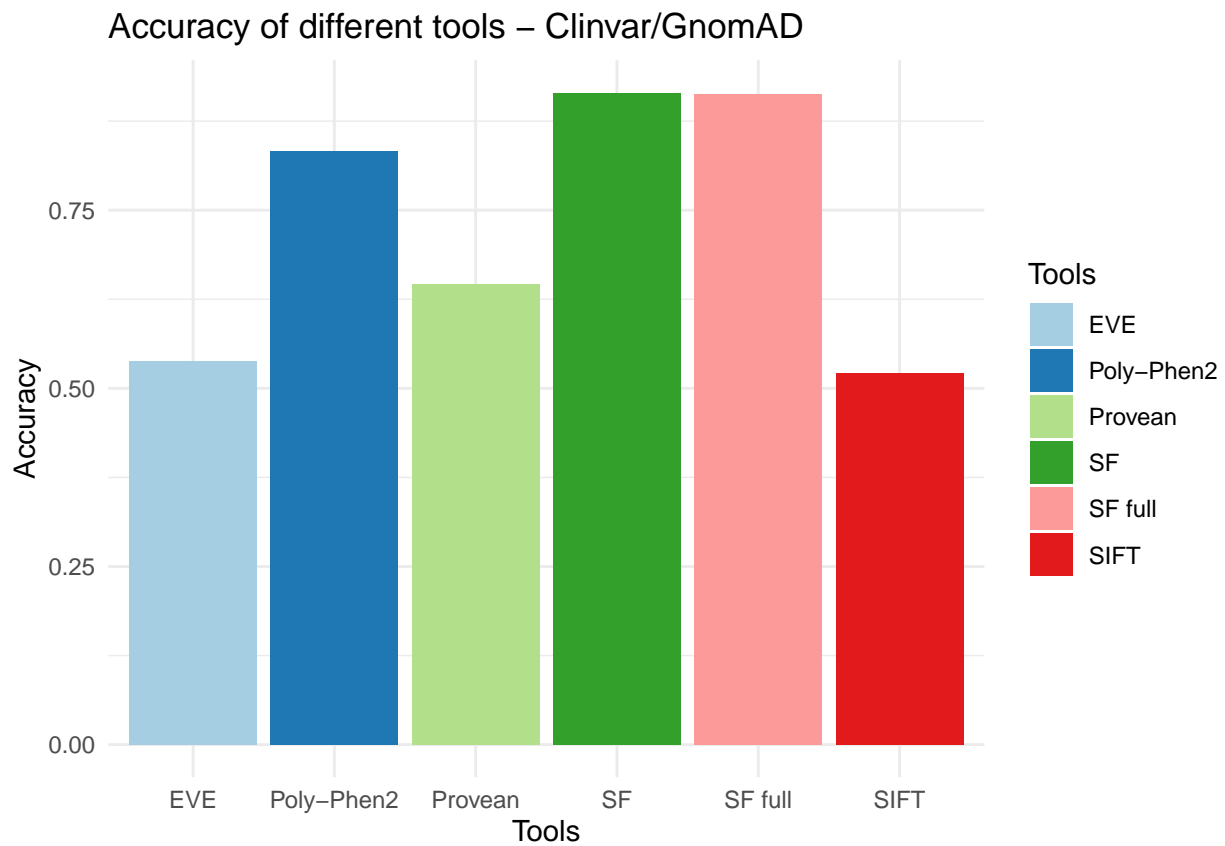
- Eve: 1-121 positions and 726-730 positions do not have scores
- For 13 from 48 - no information
- 9 variants from 48 - uncertain significance

SF full:

Despite the fact, that *SF full* shows the same statistics as *SF*, variants, that were characterized as “Pathogenic” by *SF*, in *SF full*, they have prediction “Likely Pathogenic”.

Statistics - Accuracy

| Accuracy | Tools |
|-----------|------------|
| 0.6458333 | Provean |
| 0.8333333 | Poly-Phen2 |
| 0.5217391 | SIFT |
| 0.9148936 | SF |
| 0.5384615 | EVE |
| 0.9130435 | SF full |



Clinvar without Gnomad

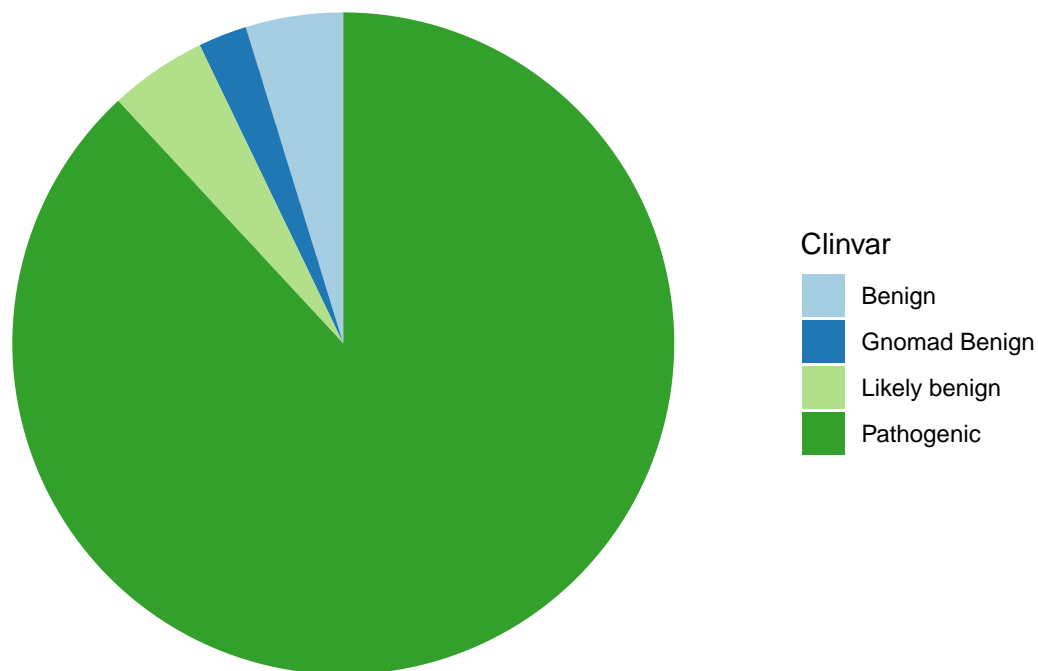
Control sample

```
table(predictions_no$ClinVar)
```

```
##
##      Benign Gnomad Benign Likely benign  Pathogenic
##          2          1          2          37
```

Because variants from GnomAD database do not have sufficient allele frequency the same procedure was made without GnomAD Benign variants

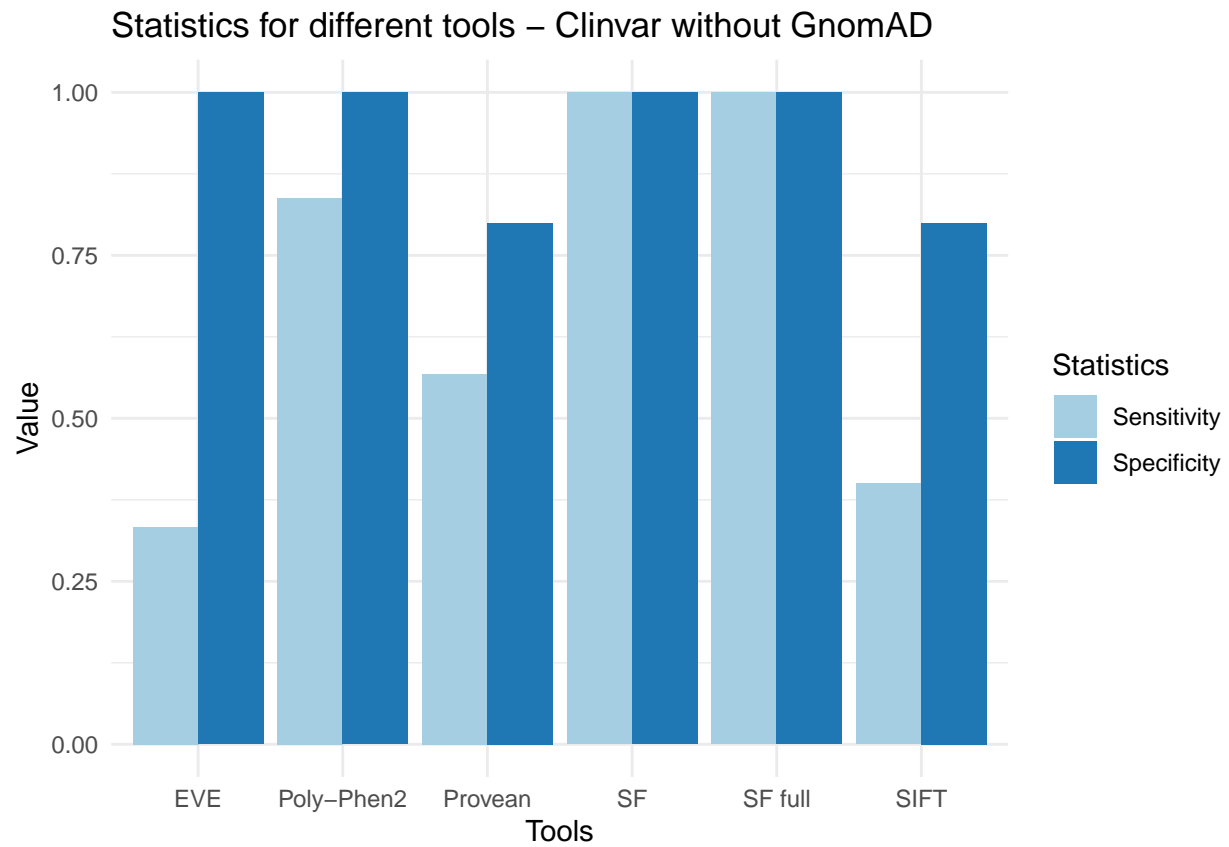
Only one Gnomad variant left - NP_006209.2 134 P S



Calculating Statistics - Sensitivity, Specificity, Accuracy

Statistics - Sensitivity, Specificity

| Statistics | Tools | Value |
|-------------|------------|-----------|
| Sensitivity | Provean | 0.5675676 |
| Specificity | Provean | 0.8000000 |
| Sensitivity | Poly-Phen2 | 0.8378378 |
| Specificity | Poly-Phen2 | 1.0000000 |
| Sensitivity | SIFT | 0.4000000 |
| Specificity | SIFT | 0.8000000 |
| Sensitivity | EVE | 0.3333333 |
| Specificity | EVE | 1.0000000 |
| Sensitivity | SF | 1.0000000 |
| Specificity | SF | 1.0000000 |
| Sensitivity | SF full | 1.0000000 |
| Specificity | SF full | 1.0000000 |



Statistics - Accuracy

| Accuracy | Tools |
|-----------|------------|
| 0.5952381 | Provean |
| 0.8571429 | Poly-Phen2 |
| 0.4500000 | SIFT |
| 1.0000000 | SF |
| 0.4782609 | EVE |
| 1.0000000 | SF full |

