

PRELIMINARY STEP

- The computer used for executing the steps should conform with the system data provided in the file **os-release.txt**
- Extract the archive **HDVDB.tar**
- Place **ACTAT_PERMS**, **newpos.sh**, **newpos.awk**, **dist.txt**, **sl1.sh**, **sl2.awk** and the directory **gx** at the same location as the untarred **HDVDB.tar**
- The script **sl1.sh** uses **RNAfold** – verify ViennaRNA is installed on computer and **RNAfold** is located at a directory included in the machine's **\$PATH**
- Place **sorted_candidates_genotypes_3_7_und.xlsx** and **candidates_genotypes_1_4.xlsx** at a location of your convenience on a Windows computer. This spreadsheet is provided for easy reference to the Results section. Please note that motif positions specified therein are **line positions** rather than **genome positions**.

At the same Linux location of the files from this step, run:

STEP 1

./newpos.sh

STEP 2

cat dist.txt | while read L; do ./sl1.sh \$L; done > candidates.txt &

STEP 3

- For genotype 3:

while IFS=\$'\t' read -r amber id _ sl1p _; do awk -v ID="\$id" -v AW="\$amber" -v SL1P="\$sl1p" -v SL2="CCGCAG" -f sl2.awk 1_xx\$id; done < ./gx/g3_44 | sort -nk2

- For genotype 7:

while IFS=\$'\t' read -r amber id _ sl1p _; do awk -v ID="\$id" -v AW="\$amber" -v SL1P="\$sl1p" -v SL2="CCGAGG" -f sl2.awk 1_xx\$id; done < ./gx/g7_44 | sort -nk2

- For undefined genotype:

while IFS=\$'\t' read -r amber id _ sl1p _; do awk -v ID="\$id" -v AW="\$amber" -v SL1P="\$sl1p" -v SL2="CCGCAG" -f sl2.awk 1_xx\$id; done < ./gx/gund_44 | sort -nk2

```
while IFS=$'\t' read -r amber id _ sl1p _; do awk -v ID="$id" -v AW="$amber" -v  
SL1P="$sl1p" -v SL2="CCGAGG" -f sl2.awk 1_xx$id; done < ./gx/gund_44 | sort -nk2
```

Note: the provided files **g3_44**, **g7_44** and **gund_44** are produced by running the script **find_44.sh**