



# Upgrade history

## 2021-02-15 (data files updated)

- data: add gbags.hbop and gbags.cancer. They can be used as \$GBA in slurm.all\_steps and slurm.annotate.
- data: add eQTL (obtained from eQTL-catalog) and pQTL (obtained from 6 publications on plasma pQTL).
- data: add kg.aim.\* which are genotypes of ancestry informative markers (AIMs) obtained from the 1000 Genomes Project in PLINK format.
- slurm.all\_steps: add KEEP\_AIM. If KEEP\_AIM=kg.aim, the script will use the provided AIMs to infer race (AFR, AMR, EAS, EUR, SAS).
- slurm.all\_steps: supports EPACTS. See parameter\_section\_2/2 of the script.
- slurm.annotate: save outputs to a folder and then compress the folder into a file.
- slurm.annotate: now create output file sorted and indexed by gene symbol. Set VAG="--sort-bp" to turn off this feature.
- slurm.annotate and slurm.all\_steps: if PVN=no, skip running PROVEAN.
- vAnnGene: for intergenic variants, gene is Intergenic\_CHR\_POS, not NA.
- VQC: sex checking is based on race-specific allele frequencies; otherwise, use ancestry uninformative markers only.
- doc: add a file do.annotate.sh

## 2020-08-15

- vAAA vAAA2: --linear does Firth's penalized linear regression rather than linear regression.
- vAAA vAAA2: --my-R uses "OUTCOME\_VARIABLE" to represent outcome variable and "+COVARIATE\_NAMES" for covariates.
- vAnnGene: removed a bug that lead to program ends when calculating MaxEntScan near N.

## 2020-08-10

- vAnnGene: Use MaxEntScan to predict de novo splicing gain.
- vAnnGene: add options --sah-nl, --sam-nl, --sah-dg, --sam-dg, --sah-lof, --sam-lof. All default to "yes" except --sah-lof and --sam-lof.
- vAnnGene: add option --cat-par, default to Promoter,Enhancer.
- vAnnGene: --etg defaults to @GDB\_PETModule\_human,@GDB\_JEME\_EL; remove @GDB\_RED,@GDB\_JEME\_FL,@GDB\_FOCS\_E,@GDB\_FOCS\_P.
- vAAA vAAA2: --keep-ms default to false.
- all programs: add --lof-flag

## 2020-07-28

- vAnnGene: Use MaxEntScan to predict native splice loss. Removed scSNV.
- vSEG vSEG2: Check whether liability class is out of boundary. Previously it will lead to a segmentation fault.
- vSEG vSEG2: Removed PopAF from the default --af.
- HGVSreader: do not read dbSNP by default.
- PedPro: read CPF format.
- PedPro: add --gvol

## 2020-07-08

- vAnnGene: add g. nomenclature for insertion / deletion that has alternative alignments
- vAnnGene: write OriginalIndex only when aligned\_index is different from original\_index
- slurm.all\_steps: removed a bug in reporting incidental findings
- slurm.all\_steps: removed a bug in pedigree analysis

## 2020-06-11

- vAnnGene: rank consequence by FuncType.Flag, Guilt-By-Association, and Distance to transcript start site if Upstream.
- vAnnGene: add --except-flag.
- removed a bug in the 2020-06-09 version.

## 2020-06-09

- data: add RED FOCS JEME PETModule
- vQC: --rep uses the current file, not the replication file, as denominator to calculate call rate.
- vFIN: output gene set names from small set to big set (in terms of number of genes in each set)
- vAAA, vAAA2: add --keep-ms and default to yes.
- victor\_by\_chr: remove imputed missing genotype from BEAGLE (previously do so only for SHAPEIT).
- victor\_by\_chr: remove merging mnp from phased data.
- slurm.all\_steps: VAG="--biol=\$GBA --no-split" by default. This is important for non-coding regions.
- slurm.all\_steps: add "Analysis Location"
- slurm.all\_steps: ignore \_is\_a\_lwp
- slurm.all\_steps: changed filenames \$OUT.1st \$OUT.2nd to \$OUT.qc1st \$OUT.qc2nd.
- slurm.all\_steps: robust to BEAGLE without pedigree file.
- slurm.all\_steps: add VFT
- slurm.all\_steps: \*.ann.del.gz do not contain genotypes anymore.
- vAnnGene: add --biol option
- vAnnGene: add --etg option
- vAnnGene: --up changed default from 250 to 30000
- vAnnGene: sort lines by gene even with --no-split. In slurm.all\_steps, VAG="--no-split" will not analyze a variant twice for overlapping genes.
- vAnnGene: Output SpliceUT# instead of SpliceUTR# (# is 3/5), SpliceCDS instead of SpliceSite.
- vAnnGene: write the most severe outcome first even with --no-split=no.
- vAnnGene: --filter changed to --filt-func. --no-filter default changed to yes.
- all: penetrance is dominant by default
- all: test whether data folder is valid.
- add: vSEG2 program

## 2019-10-29

- vFIN: --gba-show default changed from MAX to MEDIAN.
- all: test whether data folder is valid.
- slurm.all\_steps and slurm.annotate: test whether users are running analysis with /path/to/VICTOR/.
- slurm.all\_steps and slurm.annotate: detection of genome name does not depends on terminal window size anymore.

## 2019-09-25

- vAnnGene: read ClinVar INS of TBS or SVA
- vAnnGene: remove a bug that annotate SpliceSite as Intronic(LoF)

- vSEG: shrink liability classes
- pedpro: add --join-gtp
- pedpro: column name uses 7 characters instead of 4
- all: obtain genome name from environmental variable VICTOR\_GENOME before obtaining from full path of the current directory
- vConvertTest: BOADICEA format now allows multiple pedigrees
- vConvertTest: read\_rr and create liability class by population and cohort

#### 2019-07-06

- vAnnGene: add (AlnFlag) in AltAnn.
- vConvertTest: allows affection status greater than 2.
- vConvertTest: pedigree file does not need to have the PedID column.
- vConvertTest: removed a bug: if aff=0, may output liab<0.
- vConvertTest: add --prepend-pid option and default to yes, so the website is robust to multiple PedID in a file with overlapping IndID.

#### 2019-05-15

- vAnnGene: improved alignment of InDels at splice sites.
- vAAA,vAAA2: when comparing to ExAC, does not filter ExAC variant that has no AC.
- vAAA,vAAA2: --xct-cov-pc changed to --xct-cov-pr. It applies to ExAC only, not study samples. And its default value is 0.1, not 0.9.
- vAAA,vAAA2: --xct study\_coverage\_file allows 3 formats and filename allows using the character @ to represent chr#. Add --xct-chr.
- vAnnGene: add the --vkr option.
- slurm.all\_steps: gene set analysis use GeneSet\_MSigDB.gmt instead of GeneSet\_MSigDB\_It50.gmt
- victor\_by\_chr: automatically uses --add-seg=no for vFIN if \$AAA contains --weight.
- libfbj\_geneipi: add --filt-AD but default to 0.
- vFIN: --vc also output MaxAF.
- slurm.all\_steps: write genome-wide significant results to slurm.all\_steps.run\_#.sig.
- vFILT: can read victor.chr\*.qc.ann.del.gz files.

#### 2019-04-03

- slurm.annotate: remove a bug in merging result files.

#### 2019-03-19

- Sample File: quantitative trait locus now allows any number, not just positive values.
- vGrp: output gene set name as genes with variants for analysis.

#### 2019-01-25

- vQC: --rep FILE2 allows using @ to represent a set of VCFs separated by chromosomes instead of one VCF.
- vFIN: add --gba-show.
- vAnnGene: now the order of --swap and --flip matters.
- vAnnGene: add --no-cgat to remove CG or AT SNPs no matter whether the REF is wrong.

#### 2019-01-04

- slurm.all\_steps: improved pedigree file error and warning outputs.
- vAnnGene: add the --swap and --flip options to correct REF errors.
- vAAA,vAAA2: --write works when there are only cases or controls. Previously, it reports an error when there is no case samples.

#### 2018-12-06

- slurm.all\_steps: use pedpro to fix kinship error produced by KING.
- slurm.all\_steps: use pedpro to check problems in the input Pedigree File.
- slurm.all\_steps: remove a bug that still filter variants by hard filters when VQSLOD is not calculated but par.txt says no.
- Pedigree File: now always require a header row.
- pedpro: --ind-wt default for all sequenced samples rather than just sequenced cases.
- vAAA,vAAA2: remove a bug that --logistic does not adjust for covariates.
- vAAA,vAAA2: add --my-R. Now users can run analysis with their own R codes.
- vAAA,vAAA2: add --flr-R. Now users can run Firth's logistic regression with their own R codes.
- vQC: output pedigree errors only when both phi and kinship are greater than 0.1. Add an option --phi-kin to control this feature.

#### 2018-11-22

- vQC: removed a bug that stops the program when there're pedigree structure errors.
- slurm.all\_steps: robust to wrong kinship calculation by the KING program (it rarely happens).
- vAnnGene: allows an alternative VCF format #CHROM,BP1-BP2|BP1,ID,"FromPOS",ALT. No option needed.

#### 2018-11-21

- vSEG: removed a bug that only exists in the MacOS version.
- vAAA,vAAA2: when combining p-values calculated by fisher's exact test in different cohorts, weight by the effective sample size in each cohort
- victor\_by\_char: map files for ShapeIt is named \$MAP/genetic\_map\_chr\$CHR.txt instead of \$MAP/genetic\_map\_chr"\$CHR"\_combined\_b37.txt
- slurm.all\_steps: PED\_QC/POP\_QC changed parameters that are more robust for small studies.
- configure file: supports --no-miss --filt-GP --filt-domGP --filt-recGP
- vAAA,vFIN,vGLR: --neg-biol=yes by default
- Add a new program: pedpro
- Add a new script: slurm.annotate

#### 2018-10-11

- vAnnGene: add annotation of SpliceUTR3/5. Now SpliceSite only means SpliceSite within coding regions.

#### 2018-10-09

- slurm.all\_steps: add sub-analysis
- vAAA,vSEG,vGrp,vFILT,vConvertVCF: add option --gene-wise-del to use gene-specific thresholds in variant filtering by deleteriousness
- data\_updates: add BayesDel\_GST containing gene-specific thresholds for BayesDel. It's the default parameter for --gene-wise-del.
- vQC,vAAA: removed a bug (cannot recognize \_iPop in the Sample File).

#### 2018-09-27

- vQC: removed a bug that leads to program crash when reading a VCF without genotypes or a VCF generated by IMPUTE2.
- vAAA,vAAA2,vSEG: do not filter variant by VQSLOD or hard filter anymore.

#### 2018-09-14

- slurm.all\_steps: PED\_QC and SPL\_PC exclude the MHC region.

- slurm.all\_steps: allows setting sample-wise QC cutoff values.
- slurm.all\_steps: does not allow SPL\_PC when \$SPL already has a "pop" column.
- slurm.all\_steps: automatically sets the parallelism option if NCPU<NCHR and PRL is not set by the user.
- slurm.all\_steps: use parallelism for vQC if NCPU>NCHR.
- vAAA,vSEG,vGrp,vConvertVCF: add --no-mhc and --mhc-only.

#### 2018-09-10

- vQC: added --filt-impute2, --filt-GP, --filt-domGP, --filt-recGP. Now it can do QC on VCFs created by IMPUTE2.
- slurm.all\_steps: IMPUTE workflow and filenames changed. Also added a slurm.impute2mrg.
- vAAA,vAAA2: use TMPDIR to decide where to put temporary files, instead of hard-coded as /tmp.
- vAAA,vAAA2: --logistic allow --one-sided test
- genotype: no more reading and filtering of GQ/DP/GP if the cutoff value is 0.
- vQC,vAAA,vAAA2,vSEG,vSVA: read FORMAT at each line.

#### 2018-09-06

- vAAA,vAAA2: automatically chooses inferred population or principle components for adjustment depending on analysis type.
- vAAA,vAAA2: add penalized logistic regression.
- vFILT: can use --filt-del --lof-only.
- slurm.all\_steps: vConvertVCF and PLINK use --id-delim : so that the pedigree file allow duplicated IndID between pedigrees.
- slurm.all\_steps: move result files to a slurm.all\_steps.run\_#.results folder.
- slurm.all\_steps: gene set analysis employs parallelism for a fast computation.

#### 2018-07-29

- victor\_by\_chr: remove a bug that does not handle the situation of an empty input file while PHA is not "no".
- all programs: report LatestUpdates version.

#### 2018-07-26

- victor\_by\_chr: now robust to an empty VCF for phasing. This happens in small studies.
- Add a program vFILT.

#### 2018-07-09

- slurm.all\_steps: determines the number of threads for PROVEAN by the amount of memory instead of number of CPUs, otherwise it may crash.

#### 2018-07-06

- Some small changes to make the pipeline works for small studies.

#### 2018-06-25

- Add a new program vPROV. slurm.all\_steps now automatically annotates PROVEAN if it is installed.
- slurm.all\_steps: change the MaxAF filter to 0.05 in reporting incidental findings.

#### 2018-06-12 (data files updated)

- vQC: add the --cohort option to do missing rate QC in each cohort separately.
- vAAA, vAAA2: --min-cohort defaults to 1.
- vAAA, vAAA2: when comparing to ExAC, checks for the validity of input files including .ann.del.gz and .cov.gz.
- vAAA, vAAA2: add --xct-pop option to select a population in ExAC (NFE, AFR, AMR, EAS, SAS, MALE, FEMALE, etc.) for comparison.
- vAAA, vAAA2: .cov.gz file format changes. The 3rd column is now the proportion of samples instead of the number of samples covered.
- vBED2: removed a bug that output chromosome number (23, 24, 25) instead of chromosome names (M, X, Y).
- vQC: add --hwe-controls and is default to false.
- slurm.all\_steps: can do phasing without parallelism.
- slurm.all\_steps: add parameters VAR\_MISS and SPL\_MISS for variant-wise missing rate and sample-wise missing rate QC, respectively.
- other minor improvements

#### 2018-06-01 (data files updated)

- Bug fix in both data and programs

#### 2018-05-29 (data files updated)

- Data files: Add GRCh38 and GRCh37, in addition to hg19. See Manual for how to tell the programs which assembly to use.
- all programs: VICTOR\_STDERR\_MUTEX changed to STDERR\_MUTEX
- vAAA, vAAA2: removed the --recessive option for Fisher's exact test. Now FET detects the model from penetrance.
- vQC: "The number of RVs per sample is significantly different between cases and controls" is not an error but a warning.
- vAnnGene: add the --prefer option to choose the preferred gene among overlapping genes for annotation.
- vDEL: "use the highest BayesDel score for LoF variants" applies only to genes selected by ExAC pLi+pRec.
- vAAA,vAAA2,vAAA\_GLR,vSEG,vGrp: add --lof-tol so that --lof-only analysis excludes LoF-tolerated genes.
- slurm.2\_ann: changed name to slurm.step2.
- slurm.all\_steps: use --hard-filter=yes if VQSLOD is not calculated.
- vConvertVCF, vGrp, vMAF, vQC: --filt-mac will be ignored if total sample size is too small.

#### 2018-04-13

- Add the support of EPOCHS. Include two new scripts, slurm.epacts and run\_epacts.
- slurm.all\_steps: default MZ twin cutoff is changed to 0.35355 from 0.40612
- slurm.all\_steps: varRank does not filter variants by BayesDel.
- vQC: --join-sample-qc doesn't need to be the last program option.
- vGrp: remove a bug that produce an error when --pergen is not set.
- vAAA, vAAA2: add Wilcoxon rank sum test --ranksum.
- vAAA, vAAA2: --var-wt default to true.
- vAAA, vAAA2: --min-set changed to --min-cohort.
- vAAA, vAAA2: removed a bug that make FET not computed in some situations.
- vAAA, vAAA2, vSEG, vGrp: do not filter variants by quality (FILTER, VQSLOD, missing rate).

#### 2018-03-29

- vAnnGene: remove a bug that results in wrong Grantham score or no Grantham annotation.
- vAnnGene: for missense changes of the same position as a ClinVar variant but to a different AA, compare to missense ClinVar variant only.
- vAnnGene: add the Blossum62 score annotation.

#### 2018-03-27

- vAnnDel: the default --ann is empty. Otherwise, slurm.all\_steps will not work properly if AN1/An2/An3 is not set.

- vAAA, vAAA2: outputs an error message and quits the program instead of crashing when the input Genotype File is truncated.
- slurm.all\_steps: stops running when an error occurs.

#### 2018-03-22 (data files updated)

- vAnnGene: annotate whole intron deletion as INTRONIC instead of SPLICING.
- vFIN: output is more informative when reading gene-set analysis results.
- slurm.all\_steps: default Het/Hom cutoff changed from 2.1 to 4.
- slurm.all\_steps: add a step ExAcnT.
- slurm.all\_steps: add an option QCJ.
- slurm.all\_steps: AN1 AN2 AN3 are no longer Annotation Files, but parameters for vAnnDel.
- slurm.all\_steps: CHF default changed to chr1c\_noMT.
- vQC: sex imputation from Y is more conservative. Added options --y-call-numb-m --y-call-rate-m --y-total-var-f --y-call-rate-f.
- hg19: removed a bug in g\_symbol.gz that will make gnGBA crash.
- hg19: updated ClinVar to 20180225.

#### 2018-03-12 (data files updated)

- vAnnGene: INFO add the annotation of Grantham score, Exon number, and whether the variant is inside the last exon when --no-split=no.
- vAAA, vAAA2: BayesDel filter will not be applied if a variant is LoF or is classified as pathogenic by ClinVar.
- hg19: updated APPRIS.

#### 2018-03-07 (data files updated)

- vAAA: add --show-gt and --show-if.
- vSVA: set --test to no/damaging/enhancing, no more "all".
- vSVA: add --pDup-damage.
- vSVA: detect and read the new XHMM format.
- vAAA, vAAA2: --sv is back! This time it reads the output of "vSVA --detail" instead of "vSVA --out-lof".
- vAAA, vAAA2: --collapse, weight for each strata is not simply sample size. If controls are >4 times of cases, sample size is 5\*NumberOfCases.
- vQC: impute sex from X uses FdrAF only when there are 1000+ chromosomes, otherwise uses MaxAF. So, it's robust to small samples.
- vAnnGene: default --vks use ClinVar1reports.
- vAnnGene: calculate Grantham score.
- vAnnGene: removed a bug: now no (knClinSig=1.h) or (knClinSig=1.r) if (knClinSig=1.\*) or (knClinSig=0.\*) already exists.
- hg19: add panel\_haploinsufficiency and panel\_recessive. Updated panel\_incidental.

#### 2018-02-14 (data files updated)

- all programs: --nt decide whether there're enough number of cores at the end of the command line.
- slurm.all\_steps: add an option PRL to control number of threads in analysis.
- slurm.all\_steps: add gene set analysis.
- slurm.all\_steps: add analysis after removing carriers of variants in known genes.
- slurm.all\_steps: add incidental findings.
- Add four programs: vAAA2, vGrp, vBED2, vMerge. The first 3 are used for gene set analysis. vMerge merge gene lists to obtain top candidates.
- vAAA: output header change to AAAlbf / AAAlp / AAAlpvl.
- vAAA, vSEG, vGrp: --filt-del --filt-MaxAF --filt-FdrAF --filt-SplAF does not apply to LoF or ClinVar pathogenic/likely pathogenic variants.
- gnGBA: default analysis is --gs.
- vAAA: temporary removed the support of --sv. I am rewriting this part. Please look for it in the future updates!
- hg19: updated ClinVar.
- hg19: add InterPro.
- hg19: include several GeneSet files.
- hg19: add panel\_incidental.
- vAnnGene: remove a bug that keeps the old OriginalIndex when adding a new one.
- other small improvements.

#### 2018-01-18

- slurm.all\_steps: \$OUT.vcf.gz and \$OUT.qc.vcf.gz compressed by bgzip and tabix-indexed.
- slurm.all\_steps: prune variants and more aggressive variant QC before pedigree structure QC.
- gnGBA: minimum number of genes reduced from 2 to 1.
- vPC: automatically determine number of clusters from PCs, adjust for cluster prediction instead of the raw PCs.
- vAAA: --collapse Fisher's exact test now adjust for covariates (combine p-values by Fisher's method weighted by sample size).
- vAAA: default analysis method is now --FET (the mid-p version of Fisher's exact test by Irwin's rule).
- vAAA: OR calculated by Mantel-Haenszel instead of crude OR.
- vAAA: --filt-miss-rate=1 and --filt-miss-pv=0 by default.
- vAAA: --HLR can do stratified analysis.
- vSEG: --filt-miss-rate=1 by default.
- vConvertVCF: --fam output two files.
- vQC: --sep and --dup can be used together now.
- vQC: --filt-uo-dn=0 --filt-obs-pv=0 --filt-cov-DP=0 --filt-cov-pc=0 --filt-min-DP=0 --filt-max-DP=0 by default.
- vQC: add --filt-af-diff. But 0 by default.
- vAnnGene: debug: now OriginalIndex for the INFO field is defined in the meta data of the output VCF.

#### 2017-11-27

- slurm.all\_steps: fix a bug that halt the script when \$SPL is empty and \$QC1="--do-nothing". This happens when doing BayesDel annotation.

#### 2017-10-31

- vQC: allow case-insensitive string-type covariates in a Sample File.
- vAAA: allow case-insensitive string-type covariates in a Sample File.
- slurm.all\_steps: automatically adjust for string or numeric covariates including PCs in the test of #RV/spl difference between cases and controls.
- slurm.all\_steps: automatically save sample file and pedigree file for logging.
- slurm.all\_steps: automatically remove related individuals.
- slurm.all\_steps: abort if there is a pedigree structure error.
- vSVA: removed the option --test.
- vBED: add --split.

#### 2017-10-24

- slurm.all\_steps: fix a bug that makes STEP\_3 fails.
- slurm.all\_steps: automatically save the par.txt file as part of the analysis log.

#### 2017-10-16

- slurm.steps123: changed name to slurm.all\_steps
- slurm.all\_steps: missing rate cutoff changed to 0.005.
- slurm.all\_steps: does principle component analysis and determines how many principle components to be adjusted for in analysis.
- slurm.all\_steps: improved logging (it saves the script; it uses the --array option to name log files; it names log files after the script filename).
- slurm.all\_steps: automatically determines the VQSLOD cutoff.
- vAAA: --collapse also calculate a p-value by Fisher's exact test. Outputs are sorted by p-values.
- vAAA: --HLR and --SSUw ignores covariates.
- vAnnGene: clearer output of ClinVar annotation. It differentiate "uncertain" and "ambiguous".
- vQC: --join-sample-qc tests for the difference in #RV/Sample between cases & controls if --spl is set.
- vQC: add --min-bed.
- vQC: remove a bug that lead to the output vcf file not validated (the meta data for VICTOR\_QC should have "Number=1").
- vFIN: add --no-zero to omit the genes or variants that have no non-reference genotypes observed in data.
- vFIN: supports the vAAA outputs produced by --detail or --collapse.
- victor\_by\_chr: \$OUT\_PFX.qc.vcf.gz is indexed.
- Adds a script victor.sbatch for job submission.

#### 2017-09-15

- vBED: does not calculate intersection anymore if --pc=0.
- vAnnGene: --vks works for StopGain too.
- vAnnGene: --vks excludes conflicting records by matching variants at the protein sequence level.
- vConvertTest: does not exit with a fatal error if proband is not set.
- vConvertTest: PID IID can be anywhere in the input pedigree file.
- vAnnGene: removed a bug that reports knClinSig=ambiguous even if the variant has no match in ClinVar.
- HGVS\_to\_genomic: prints a header line. Adds the --no-header option.
- vConvertTest: reports a fatal error if there's no carrier in the input pedigree file.
- HGVS\_to\_genomic: adds the function to select a SNV by observation in ClinVar if a protein-level HGVS has multiple genomic interpretations.

#### 2017-09-06

- vAAA: --collapse calls haplotype=2 if both 1 and 2 have equal probability.
- vSEG: proband labels are required for the IID column only, not IID+DAD+MOM.
- vSEG: adds --allele-freq.
- vSEG: if the penetrance file contains only one liability class model, use that for chrX too.
- vConvertTest: supports standard linkage file, including pre/post makeped.
- vConvertTest: supports pedigree file obtained from the BOADICEA web tool.
- vConvertTest: treats successive spaces as one delimiter and trims leading white spaces for a non-BOADICEA Pedigree File.
- HGVS\_to\_genomic: adds the default values for -x --min --step.
- HGVS\_to\_genomic: removed a bug that leads to "cannot find MaxAF database" file even it's there.
- vAnnGene: skips a line and continues if REF=ALT (previously exit with a fatal error).
- vAnnGene: labels (knClinSig=ambiguous) if ambiguous (previously no label).
- vSEG: automatically chooses a proband.

#### 2017-08-31

- vAnnGene: removed a bug that outputs fs for synonymous InDel.
- vAnnBase: removed a bug introduced in the last version, which may add MaxAF to BayesDel twice for some InDel variants.
- slurm.steps123: does not write "--rm-ind" to par.txt.
- vFIN: does nothing if the input is vAAA --collapse output.
- vAAA: --collapse also outputs odds ratio with Yate's correction. Now --collapse outputs damaged haplotype counts rather than genotype counts.

#### 2017-08-24

- MaxAF: includes gnomAD, UK10K, GoNL and 1kJpn. Only variants with 5+ MAC are included, except those in gnomAD\_WES.
- BayesDel: fixed a bug that leads to many missing values in the database file
- BayesDel: this version does not include MaxAF. Accordingly, vDEL adds an option --add-af so that you can add MaxAF to BayesDel.
- All programs: array options accept EmptyString. This affects vMAF --pop.
- vQC: --join-sample-qc outputs to <prefix>.sex\_problem and <prefix>.sample\_qc. So the --prefix option is important.
- ClinVar: updated to clinvar\_20170801
- InDels.provean.tsv.gz: updated with more variants
- vAAA: adds an option --var-wt, which is default to "no"
- vConvertVCF: add an option --keep-unk
- vFIN: debug: --vc not robust to variants in multiple overlapping genes

#### 2017-07-26

- victor\_by\_chr: log problems discovered by vAnnGene in step 1
- vMAF: add --out-ms, and def to false
- vMAF: --pop removed \_OTH
- vAnnGene: is now robust to an uncommon VCF format like 1,1097411,AC,C
- vAnnGene: add --do-nothing
- vAnnGene: log duplicated variants
- vAnnGene: removed a bug that annotate p.S1S as missense variant
- vAnnGene: removed a bug that makes output not sorted by POS
- vAnnGene: removed a bug that makes the program crash when the transcription start site is at the first basepair of an exon
- vDEL: removed a debug that didn't update BayesDel based on PROVEAN if BayesDel is already annotated
- vDEL: improved speed
- The nsfp33a.gz file in the FAQ page has been updated (previously lines were not sorted properly).
- Improved the work flow in the Example 8 of the VICTOR Tutorial, annotation deleteriousness and functional consequence.

#### 2017-07-17 (data files updated)

- vAnnDel: improved speed.
- vAnnDel vAnnMAF vQC HGVS\_to\_genomic: now robust to tabix returning a result with a wrong bp (a bug in tabix).
- vSEG: treats chrX PAR as autosomal.
- vAnnDel: added the option --add-ms
- data/hg19: fixed the sorting of MaxAF\_gUGN, snp147 and ClinVar2reports. Added @GDB\_cdd (conserved domain databases).

#### 2017-07-10

- All programs: improved stderr outputs. They do not mix together even with Unix pipe and multi-processing.
- vAAA and vSEG: do not apply allele frequency filters when --vc is set.
- vAnnGene: removed a bug introduced on 2017-07-04, which makes program stop if BayesDel is NaN and --vks is set.

#### 2017-07-04

- vAAA: added an option --weight.
- vAAA: --write now always add a column of individual weight at the end.
- vAnnGene: improved log to stderr and --log output.
- vAnnGene: --no-split do not always write to INFO.
- vAnnGene: improved HGVS nomenclature (H178\_179insPHP should be H178\_H179insPHP).
- vAnnDomain: added --wr.
- vAAA and vSEG: --LoF-only also filters NMD.
- vDEL: removed a bug that --check-ms does not work when PROVEAN is not annotated at all.
- vMAF: removed \_ASJ\_FIN because MaxAF should not be calculated from founder populations.
- vQC: removed options --cadd-cutoff and --fmnc-cutoff.
- vQC: added INFO fields AN\_Founder AC\_Founder AF\_Founder Hom\_Founder, which can be read by vMAF --pop=\_Founder.
- vConvertTest: can read liability class from a pedigree file.
- vAnnDel: added an option --remove.
- slurm.step123: small improvements.

#### 2017-05-25

- vAAA: added options --out-pv and --out-mlp.
- vAnnGene: added options --gene-file, --tx-file, --genes, --txs, --no-filter. Accordingly, the Example 8 in Tutorial changed too.
- victor\_by\_chr: replaced the flag QC\_ONLY with STEP1\_ONLY. But QC\_ONLY is still valid for backward compatibility.

#### 2017-05-22

- Added a program HGVS\_to\_genomic to convert HGVS nomenclature to genomic coordinates #CHROM,POS,REF,ALT.
- vAnnGene: added the option (--vks) to annotate variant of known significance (VKS). A variant will have a label knClinSig=1.a if it matches a known pathogenic VKS; or knClinSig=1.b if the variant leads to the same amino acid substitution as a VKS; or knClinSig=1.c for a missense change of the same amino acid position as a VKS but for different substitution with a higher deleteriousness score than the VKS. Variants that matches a known benign variant have a similar naming scheme, but they will be filtered out by default. This filtering can be turned off by the option --filt-benign.
- The MaxAF database added GoNL.
- vSPLIT and gnGBA removed a bug that doesn't read the par.txt.
- Pedigree File: a proband ID can end with [p] or (p) or [proband] or (proband), or simply "proband" (case-insensitive).
- vAnnGene the default value for --sv changed to yes.
- vSEG do not calculate and use allele frequency from data anymore. If MaxAF is not annotated, use the default MAF.
- vQC added an option --do-nothing.
- vQC changed the default value of --filt-obs-maf to 0.01.
- vSEG added more check points and messages.
- vAAA added linear regression. And the default analysis method is now linear regression, not HLR.
- All programs do not check web version automatically. They have a new option --version to invoke version checking.
- Added a SLURM script template to do step 2 and 3 on one computer node with multi-processing.
- Removed a bug in vDEL

#### 2017-04-20

##### Major upgrades

- vSEG added the option --mut-male and --mut-female to take into account de novo mutation rate in co-segregation analysis. Some genes have a high de novo mutation rate among individuals affected with a certain disease. An example is the *TP53* gene for Li-Fraumeni Syndrome, which has a 7%-20% germline de novo mutation rate. This upgrade is suitable for variant classification for the genes like *TP53*.
- vDEL calculates BayesDel from PROVEAN for InFrame InDels, and from Fathmm-MKL\_noncoding + CADD + MaxAF for non-coding InDels and SNVs. BayesDel for coding SNVs is still calculated from an ensemble of individual scores like before, but the MaxAF database is updated (see below) and the individual scores now use dbNSFP v3.3 (academy), which included FATHMM, GERP+ + \_RS, LRT, MutationAssessor, MutationTaster, Polyphen2\_HDIV, Polyphen2\_HVAR, PROVEAN, SIFT, SiPhy\_29way, VEST3, fitCons, fathmm-MKL\_coding, phastCons100way, phastCons20way, phyloP100way, and phyloP20way.
- The package provides the pre-computed BayesDel scores for all possible SNVs in the entire genome. The size of this database is only 8.7 GB and the annotation is super fast due to the new vAnnBase program that is also shipped with the package. This new program can also annotate InDels using the SNV scores in affected regions. Interestingly, the performance of this strategy is better than CADD for all InDels and is very close to PROVEAN for In-Frame InDels. After integrating MaxAF, this method outperforms PROVEAN and CADD for all types of InDels.
- The MaxAF database is calculated from the UK10K whole genome cohorts ALSPAC and TWINSUK, the 1000 Genomes Project, and the gnomAD. Several quality control filters (HWE  $p > 0.000001$ , VQSLOD  $> -5.368$  for SNV and  $> -4.208$  for InDel, AN  $> 200$ ) were applied before computing allele frequency. Another improvement of the MaxAF data is that both the 5'-most and 3'-most alignment of InDels are recorded. This makes the database robust to data that is not always left-normalized. See below for the reason why some data is not left-normalized intentionally.
- The gene database is updated to Ensembl release 87, and refGene.txt version Mar 27, 2017.
- vAnnGene rewrites the POS/REF/ALT fields based on the functional prediction of a variant. Previously, outputs are always left-normalized, a popular method to make the genomic coordinates of InDels comparable between datasets. For example, an indel could be annotated as a damaging variant by destroying a splice site if 5'-aligned or as a benign variant within an intron if 3'-aligned. In the previous version, vAnnGene annotates with the most probable functional consequence, but the output POS/REF/ALT fields are always left-normalized for easier variant matching, disregarding the functional consequence. This created a problem in subsequent deleteriousness score calculation, since most in silico programs (such as CADD) do not consider an alternative alignment for InDels. In this new version of vAnnGene, the output genomic coordinate of a variant corresponds to the functional consequence annotation. Some variant may be 5'-aligned, while others may be 3'-aligned. Although this makes variant matching less straightforward, the correct deleteriousness prediction is far more important. In addition, variant matching could be easily solved by creating a variant database with both alignments whenever possible. This is why the MaxAF database provided by VICTOR contains both left- and right-aligned InDels (see above).
- vAnnGene changes the format of --pas input files. Now it allows different location types (DNA, RNA, CDS, AA, UTR3, DOWN).
- vAnnGene now supports nonsense-mediated decay (NMD) annotation. vDEL treats NMD as LoF.
- vAnnGene labels SpliceAltering for variants predicted by dbNSFP::dbcsSNV instead of SpliceSite(LoF).
- vAnnGene SpliceSite is not LoF anymore if the variant is not an LoF in an alternative transcript of the same gene.

##### Minor upgrades

- vAnnGene splits overlapping genes into multiple lines even with the --add-info option. A new option (--no-split) is added to address the needs for not splitting variant by genes in the output.
- vAnnGene has improved output. Write HGVS even for SpliceAltering/miRNAbinding/TF\_binding variants. The source for SpliceAltering, miRNAbinding and TF\_binding are written to Func\_Type instead of Func\_Detail.
- vAnnGene now is robust to variants like 1,1273412,GTAGGCAGG,GC (equivalent to 1,1273413,TAGGCAGG,C) or 1,11256034,TGTGA,CGTGA (equivalent to 1,11256034,T,C).
- vAnnGene updates the miRNA-binding site database to TargetScan 7.1 default\_predictions.
- vAnnGene option --filter now supports "coding" and "noncoding".
- vAnnGene add an option --one-lett for one-letter amino acid code in HGVS, and --ter to specify code for termination.
- vAnnGene add an option --log to record variants that have been removed by --filter.
- vAnnDel allows an empty parameter for the --ann option, which will make the program does nothing.
- vFIN output is sorted by the results. So the top genes will be listed up front.
- vFIN removed a bug that doesn't add biological relevance score to the overall result for Ensembl genes.
- vMAF tests for HWE before calculating the allele frequency from a cohort.
- vMAF adds \_ASJ to the population list, which is helpful for reading gnomAD data.
- vMAF requires AN  $> 200$  for allele frequency calculation. This can be changed by the --min-an option.
- vSVA adds the support of CLAMMs output file format.

- vQC adds an option --check-ms to find the variants that need a PROVEAN or CADD annotation and write the info to a file.
- vQC do hard filtering if VQSLOD does not exist. This feature can be turned off by --HardFiltIfNoVQSLOD.
- vQC p-value cutoff for several QC filters changed to 0.000001.
- vQC's default value for vQC --filt-cov-dp is changed to 10.
- vQC removed a bug that makes the program crash when doing HWE test with an extremely large sample size.
- vQC removed a bug that cause --join-sample-qc failed to read the first line of an input file.
- The package's default cutoff for MaxAF variant filter is changed to 0.01.
- The package's default setting for removing a variant without a VQSLOD score is "no". Instead, vQC performs hard filtering.
- Some input filenames can contain @GDB as a template to be replaced with refGene or ensGene according to --gdb.
- Add a new program vAnnDomain to annotate protein domains.
- All programs: -h will show the final parameters after parsing par.txt and program options.
- Script do\_gp\_by\_chr became victor\_by\_chr. It logs parameters to stdout, and adds a step to fill in missing in silico scores.
- Script victor\_by\_chr add a parameter AF1 to annotate user-defined allele frequency in addition to the provided MaxAF\_gUG.

#### **2016-12-24**

- vDEL calculates BayesDel from CADD for non-coding regions.

#### **2016-06-03**

- First release of VICTOR.

#### **2016-05-06**

- First release of VANNER.

#### **2015-01-10**

- First release of PERCH.