20190821_data_release_notes

Overview of changes in the release of Aug 21, 2019

Please also see our RSS feed for information about accessing the data.

Overview of submissions: 2019

Date	Total Submissions
Jan 01, 2019	759562
Feb 07, 2019	778673
Mar 01, 2019	782638
Apr 01, 2019	787656
May 01, 2019	795045
Jun 01, 2019	811551
Jul 01, 2019	819827
Aug 01, 2019	825177

Content

These release notes describe changes made to ClinVar's annotation of HGVS expressions, molecular consequence, single-letter protein changes, and preferred names.

These changes were first released in the 8/21/19 weekly release. The release on 9/5/19 will be the first monthly, archived release with these changes.

Summary

ClinVar now uses NCBI's Variation services and SPDI notation to annotate variants with HGVS expressions, molecular consequence, single-letter protein changes, and preferred names.

More than one million HGVS expressions have been added to the database. Previously, ClinVar reported a subset of HGVS expressions per variant. ClinVar now reports HGVS expressions on all NM/NR/NP/NG accessions for the gene; on NC accessions for GRCh37 and GRCh38; and on LRGs. Users who download data from ClinVar now have a more complete set of HGVS descriptions for variants. Users who search the ClinVar website now get results for more queries with HGVS expressions.

The variant type for some variants has been corrected; for example, an insertion that should be reported as a duplication, or deletion or duplication that should be reported as a microsatellite. If you filter searches based on variant type, consider searching for additional variant types.

Changes to note

- We have added mapped locations on genomic sequence for >400 variants.
- >1,000,000 new HGVS expressions are reported, on <28,000 RefSeqs that were not previously reported in ClinVar.
- >11,000 HGVS expressions that had been inappropriately deleted were restored.
- >700 HGVS expressions reported as insertions were corrected to be reported as duplications.
- >5000 variants that were previously reported as deletions and >1400 variants previously reported as insertions are now reported as microsatellites.
- Molecular consequence is calculated for all HGVS expressions on transcripts (NMs and NRs).
- ClinVar's XML products report only the left-shifted genomic location. In other words, start/stop and vcf_pos all report the left-shifted location.
- ClinVar's XML products no longer report referenceAllele and alternateAllele corresponding to an HGVS expression. Instead, only referenceAlleleVCF and alternateAlleleVCF are being reported.
- ClinVar's XML product based on the RCV accessions (ClinVarFullRelease) will no longer report XRefs for any HGVS expression. ClinVar now treats these as properties of the variant (Measure), rather than data that needs attribution.
- Variant length is calculated differently for insertions and deletions, such that the length of the variant is reported as whichever is longer, the length of the reference allele or the length of the alternate allele. For example, a change of AAA to TT, has a length of 3, and a change of AA to TTT also has a length of 3.

Known issues

There are a few outstanding issues that are known to the ClinVar team; we will fix these issues shortly.

- HGVS and molecular consequence are not calculated for approximately 5300 variants.
- Some HGVS expressions for intronic variants downstream of the stop codon are reported incorrectly, without the * indicating that the location is downstream of the stop codon
- Some HGVS expressions for variants in introns, where the gene is on the opposite strand, may be incorrect.

These changes were advertised in the last few monthly release notes as:

coming soon: HGVS expressions updated to standard format	We calculate HGVS expressions for variants that are reported to ClinVar. Within the next several weeks, we will update many of the HGVS expressions that we calculate, to reflect changes to the HGVS standard, including do not report the duplicated or deleted sequence for a duplication or deletion. Note that HGVS expressions provided by submitters will not be updated; they are retained as submitted.
coming soon - genomic location for variants with defined endpoints will be reported only as "VCF-style"	Genomic location for variants with defined endpoints currently reported in the ClinVar XML in two ways: • start and stop, which are right-shifted similar to the HGVS standard • vcf_start and vcf_stop, which are left-shifted similar to VCF format We plan to simplify reporting of genomic location and only report the location as the left-shifted position, for consistency with VCF format. This change is not yet scheduled. If you have any concerns, please write to us as clinvar@ncbi.nlm.nih.gov.

Overview of submissions: 2018

Date	Total Submissions
Jan 01, 2018	579543
Feb 01, 2018	582113
Mar 01, 2018	593651
Apr 01, 2018	610005
May 01, 2018	645149
Jun 01, 2018	676018
Jul 01, 2018	676575
Aug 01, 2018	685942
Sep 01, 2018	701880
Oct 01, 2018	708726
Nov 01, 2018	715516
Dec 01, 2018	749203

Overview of submissions: 2017

Date	Total Submissions
Jan 01, 2017	396005
Feb 01, 2017	405182
Mar 01, 2017	406220
Apr 01, 2017	446265
May 01, 2017	482941
Jun 01, 2017	486420
Jul 01, 2017	488658
Aug 01, 2017	492592
Sep 01, 2017	504299
Oct 01, 2017	512373
Nov 01, 2017	517157
Dec 01, 2017	519359

Overview of submissions: 2016

Date	Total Submissions
Jan 01, 2016	172867
Feb 01, 2016	176710
Mar 01, 2016	178032
Apr 01, 2016	180549
May 01, 2016	181155
Jun 01, 2016	192617
Jul 01, 2016	204415
Aug 01, 2016	209842
Sep 01, 2016	210200
Oct 01, 2016	213499
Nov 01, 2016	236420
Dec 01, 2016	240042

Overview of submissions: 2015

Date	Total Submissions
Jan 01, 2015	149013
Feb 01, 2015	156999
Mar 01, 2015	162455
Apr 01, 2015	171408
May 01, 2015	172044
Jun 01, 2015	173236
Jul 01, 2015	184506
Aug 01, 2015	154686
Sep 01, 2015	158580
Oct 01, 2015	160538
Nov 01, 2015	170931
Dec 01, 2015	172006

Overview of submissions: 2014

Date	Total Submissions
Jan 01, 2014	68204
Feb 01, 2014	73492
Mar 01, 2014	83343
Apr 01, 2014	111501
May 01, 2014	112349
Jun 01, 2014	117209
Jul 01, 2014	127132

Aug 01, 2014	127557
Sep 1, 2014	143114
Oct 1, 2014	143601
Nov 1, 2014	144117
Dec 1, 2014	148008

Overview of Submissions: 2013

Date	Total Submissions
Apr 05, 2013	30333
May 01, 2013	30386
Jun 01, 2013	39047
Jul 01, 2013	39170
Aug 01, 2013	45901
Sep 01, 2013	50263
Oct 01, 2013	52047
Nov 01, 2013	64750
Dec 01, 2013	64881