# GENIE 15.0-public release notes

Sage Bionetworks

December 27, 2023

## Release notes

These release notes will encompass the changes made in the 15.0-public releases and general center information.

- 14749 samples have been added since the last public release.
- Added cross-file validation for data uploads.
  - A center's clinical file will be invalid if there is a SEQ_ASSAY_ID present with no associated bed file
  - A center's clinical file will be invalid if there is a SEQ_ASSAY_ID present with no associated panel in the assay information file
  - A center's maf file will be invalid if there are samples that exist in the maf file that do not exist in the clinical file
- Adding allele validation
  - Support standalone "N" allele values in vcf/maf validation
  - Allow only a combination of these allele values: "A", "T", "C", "G", "N"
- Updated the GENIE instance of Genome Nexus:
  - genomeNexus version: 1.0.2
  - database: version: 3.6.2
  - vep version: NA
  - cache version: NA,
  - annotationSourcesInfo:
    - name: VEP
      - version: grch37
      - type: mirrored
      - description: VEP determines the effect of your variants(SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions.,
      - url: https://grch37.ensembl.org/info/docs/tools/vep/index.html
    - name: Cancer Hotspots
      - version: v2,
      - type: mirrored
      - description: A resource for statistically significant mutations in cancer
      - url: https://www.cancerhotspots.org
    - name: 3D Hotspots
      - version: v2
      - type: mirrored,
      - description: A resource for statistically significant mutations clustering in 3d protein structures in cancer
      - url: https://www.3dhotspots.org/
    - name: HGNC
      - version: 22-10-01
      - type: mirrored
      - description: The resource for approved human gene nomenclature. Genome Nexus uses

HGNC gene symbols in annotation
      - url: http://ftp.ebi.ac.uk/pub/databases/genenames/hgnc/archive/monthly/tsv/
  * name: reVUE,
      - version: d8a7d01bac02671e74b4522bacfca6e82f360046
      - type: mirrored
      - description: A Repository for Variants with Unexpected Effects (VUE) in Cancer
      - url: https://www.cancerrevue.org/
  * name: My Variant Info
      - version: Includes many annotation sources, see https://docs.myvariant.info/en/latest/doc/data.html
      - type: external
      - description: MyVariant.info provides simple-to-use REST web services to query/retrieve variant annotation data, aggregated from many popular data resources.
      - url: https://myvariant.info
  * name: Mutation Assessor
      - version: v3
      - type: mirrored
      - description: Mutation Assessor predicts the functional impact of amino-acid substitutions in proteins, such as mutations discovered in cancer or missense polymorphisms.
      - url: http://mutationassessor.org/r3/

## Data Concerns/Issues

These are the known data issues for this release. **Note**: There could be more undiscovered issues.

- The number of expected genes in certain gene panels may be incorrect due to sites submitting invalid gene symbols, coordinates, or incorrect assay information.
- **CRUK**
  - Confirmed genomic data for **CRUK-TS** samples contain no artifacts.
- **CHOP**
  - Fusion panel includes genes that are not on the other panels/pipelines that generate the SNV and indel calls. No BED file for fusion panel.
  - Confirmed genomic data for **CHOP-STNGS** samples contain no artifacts.
- **MSK**
  - Confirmed genomic data for MSK HEME panel samples contain no artifacts.
- **DFCI**
  - Expected discrepancies in expected and submitted gene count because of regions that are targeted for calling structural rearrangements.
- **UHN**
  - Confirmed genomic data for **UHN-54-V1** samples contain no artifacts.
- **YALE**
  - Intentionally reports only amplifications in copy number data due to internal policy.
  - **YALE-OCP-V2** panel's expected gene count is 134 because one intron is included. Gene panels are created with only exons hence the difference in actual vs expected gene counts.
- Foundation Medicine genomic regions discrepancy
  - **DUKE/WAKE** use the Foundation Medicine T5A, T7, DX1, R2D2 bait sets, but some of the bed files uploaded don't seem to match the expected gene count per panel.
- Not all variants have variant counts (t_depth, t_alt_count, t_ref_count).
- Duplicated variants listed from **VICC** are expected due to providing Tempus samples.
- Genome Nexus related issues:
  - Duplicated variants
  - Non-somatic mutations (Reference_Allele == Tumor_Seq_Allele1 == Tumor_Seq_Allele2)
  - SNV variants annotated as DNP or ONP