GENIE 14.0-public release notes

Sage Bionetworks

September 14, 2023

Release notes

These release notes will encompass the changes made in the 14 series releases and general center information

- 15944 samples have been added since the last public release.
- **Bug fixes**
  - Created cases_sv.txt, so the number of structural variants in cBioPortal is more accurate.
  - Updated to the latest version of GENIE Genome Nexus and resolved a vcf-to-maf converter tool bug to address issues related to incorrect mutation annotations
- Updated the data guide to add missing GENIE panel information, a Data Harmonization and QC section, and minor formatting changes.
- **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.
- **CRUK**
  - Confirmed genomic data for **CRUK-TS** samples contain no artifacts.
- **MSK**
  - Confirmed genomic data for **MSK-IMPACT-HEME-400** 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.

Data Concerns/Issues

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

- 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.
- Sites/panels with mismatching information for the number of genes in `assay_information.txt` vs `genomic_information.txt`.
  - **COLU**, **DFCI**, **DUKE**, **GRCC**, **MSK**, **PROV**, **SCI**, **UCHI**, **UHN**, **VICC**, **WAKE**, **YALE**
- Not all variants have variant counts (t_depth, t_alt_count, t_ref_count).
- Duplicated variants listed from **VICC** are expected due to the site 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