GENIE 8.1-public release notes

Sage Bionetworks

October 29, 2020

Release notes

- 16198 samples added
- New file formats
  - `data_guide.pdf`
  - `assay_information.txt`
- **CHOP**
  - All data removed due to known genomic artifacts.
- **COLU**
  - Fixed genomic data issues.
- **DFCI**
  - 8 samples retracted due to patients withdrawing consent.
- **MSK**
  - Corrected oncotree code for some samples
- **NKI**
  - Updated all sample and patient ids. Sample and patient IDs from past releases are no longer valid and do not link/correlate to the current sample and patient IDs in the 8.2 release
  - Updated all `RACE`/`ETHNICITY` values from 'Unknown' to 'Not Collected'.
  - 84 samples retracted due to patients withdrawing consent.
- **PHS**
  - Fixed genomic data issues.
- **UCSF**
  - Fixed genomic data issues.
- **VICC**
  - 1 sample retracted due to patients withdrawing consent.
- **WAKE**
  - missing variants due to a Foundation Medicine conversion issue (Work in Progress)
  - 313 samples retracted due to patients withdrawing consent.
- **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**/**VICC**/**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 possible genomic artifacts.
  - **CRUK-TS**, **JHU-500STP**, **UCHI-ONCOHEME55-V1**, **UHN-54-V1**, **VICC-01-MYELOID**, **WAKE**, **DUKE**, **COLU**
- Sites/panels with mismatching information for the number of genes in `assay_information.txt` vs `genomic_information.txt`.
  - **COLU**, **CRUK**, **DFCI**, **GRCC**, **JHU**, **MSK**, **NKI**, **PHS**, **SCI**, **UHN**
- Sites with missing panels in `assay_information.txt`
  - **UCHI**