GENIE 9.0-public release notes

Sage Bionetworks

January 20, 2021

Release notes

- 17017 samples added
- New version of oncotree used: 2019_12_01
- Variants annotated with Genome Nexus
  - New germline filter: If any variant has a max gnomAD subpopulation AF over 0.0005, the variant is filtered out.
- **CHOP**
  - Fixed variant data (removed known artifacts)
  - Fixed gene count in assay information
  - 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-HEMEP** and **CHOP-STNGS** samples contain no artifacts.
- **COLU**
  - Fixed gene count in assay information
- **CRUK**
  - Confirmed genomic data for **CRUK-TS** samples contain no artifacts.
  - Fixed gene count in assay information
- **DFCI**
  - Expected discrepancies in expected and submitted gene count because of regions that are targeted for calling structural rearrangements.
- **JHU**
  - Fixed gene count in assay information
  - Removed BRAF p.V600M variants
- **PHS**
  - Added **PHS-TST170-V1** panel
  - Fixed gene count in assay information
- **MSK**
  - Renamed panel from **MSK-IMPACT-HEME-400** to **MSK-IMPACT-HEME-399**
  - Confirmed genomic data for **MSK-IMPACT-HEME-399** samples contain no artifacts.
- **NKI**
  - Fixed gene count in assay information
- **UCHI**
  - Confirmed genomic data contains no artifacts
  - Removed BRAF p.V600M variants
  - Uploaded `assay_information.yaml`
- **UCSF**
  - Confirmed genomic data for **UCSF-NIMV4** samples contain no artifacts.
- **UHN**
  - Fixed variant data (removed as many known artifacts as possible)
  - Confirmed genomic data for **UHN-54-V1** samples contain no artifacts.
  - Cannot guarantee that there are no artifacts for the genomic data of **UHN-555** samples, but have done the best in removing most.

  - Fixed gene count in assay information
- **VICC**
  - Reprocessed variant data
- **WAKE**
  - Fixed variant data (previously missing some variants)
- **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/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 mismatching information for the number of genes in `assay_information.txt` vs `genomic_information.txt`.
  - **GRCC**, **SCI**, **WAKE**
- Not all variants have variant counts (t_depth, t_alt_count, t_ref_count)