GENIE 10.0-public release notes

Sage Bionetworks

June 22, 2021

Release notes

- 10249 samples added
- Variants counts filled in where possible: t_depth = t_alt_count + t_ref_count
- Stop germline filter from filtering out known somatic sites.
- Improvements to Genome Nexus annotation
- **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-HEMEP** and **CHOP-STNGS** samples contain no artifacts.
- **CRUK**
  - Confirmed genomic data for **CRUK-TS** samples contain no artifacts.
- **DFCI**
  - Expected discrepancies in expected and submitted gene count because of regions that are targeted for calling structural rearrangements.
- **MSK**
  - Confirmed genomic data for **MSK-IMPACT-HEME-399** samples contain no artifacts.
  - Removed 2172 unmatched heme samples
- **UCHI**
  - Confirmed genomic data for **UCHI-ONCOHEME55-V1** contains no artifacts.
- **UCSF**
  - Confirmed genomic data for **UCSF-NIMV4** samples contain no artifacts.
- **UHN**
  - 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.
- **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`.
  - **GRCC**, **SCI**, **UCHI**, **WAKE**,
- Not all variants have variant counts (t_depth, t_alt_count, t_ref_count)
- Genome Nexus related issues

– Duplicated variants
– Non-somatic mutations (Reference_Allele == Tumor_Seq_Allele1 == Tumor_Seq_Allele2)
– SNV variants annotated as DNP or ONP