BaseSpace Core Apps Software Release Notes

Isaac Enrichment v1.0.0.1

BWA Enrichment v1.0.0.1

Isaac Whole Genome Sequencing v2.0.0.1

BWA Whole Genome Sequencing v1.0.0.1

Tumor Normal v1.0.0.1

For BaseSpace

March 4, 2014

Introduction

These Release Notes detail key changes to software components for the BaseSpace Core Apps. The versions are compared to the package containing the versions of BWA Enrichment v1.0.0.0, WGS v1.0.0.0 and TUNE v1.0.0.0 released in January and Isaac v1.0.0.0 version. This release is the first release of the Isaac Enrichment v1.0.0.1 app.

For more information about these apps and how to use them, refer to the app User Guides, available from the details page of each app.

The software package includes:

- Isaac Enrichment v1.0.0.1
- BWA Enrichment v1.0.0.1
- Isaac Whole Genome Sequencing v2.0.0.1
- BWA Whole Genome Sequencing v1.0.0.1
- Tumor Normal v1.0.0.1

I. Isaac Enrichment v1.0.0.1

NEW FEATURES:

- Initial Release.

DEFECT REPAIRS:

- None.

KNOWN ISSUES:

- In some situations, there may be a discrepancy in the number of variants included in calculation of the “Total Passing” as compared to the “Variants by Sequence Context” and “Variants by Consequence” entries in the small variants summary table. It may be possible for the number in the “Total Passing” section to be less than the sum of all the variants in the “Variants by Sequence Context” section. This situation may occur because the “Total Passing” calculation includes only simple variants where the alternate alleles are all of the same type (either SNV, Insertion or Deletion). However, the other sections, “Variants by Sequence Context” and “Variants by Consequence,” also include variants at tri-allelic sites where the two alternate alleles are classified as different types (e.g. SNV and Insertion).

II. BWA Enrichment v1.0.0.1

NEW FEATURES:

- None.

DEFECT REPAIRS:

- Bug fix: In rare cases annotation fails and the “Variants by Sequence Context” and “Variants by Consequence” sections of the report display all zeros.

KNOWN ISSUES:

- In some situations, there may be a discrepancy in the number of variants included in calculation of the “Total Passing” as compared to the “Variants by Sequence Context” and “Variants by Consequence” entries in the small variants summary table. It may be possible for the number in the “Total Passing” section to be less than the sum of all the variants in the “Variants by Sequence Context” section. This issue may occur because the “Total Passing” calculation includes only simple variants where the alternate alleles are all of the same type (either SNV, Insertion or Deletion). However, the other sections, “Variants by Sequence Context” and “Variants by Consequence,” also include variants at tri-allelic sites where the two alternate alleles are classified as different types (e.g. SNV and Insertion).

  .

III. Isaac Whole Genome Sequencing v2.0.0.1

NEW FEATURES:

- Improved aligner speed and accuracy.
- Added structural variant (SV) and copy number variant (CNV) calling capabilities.
- Added support for multiple reference genomes.
- Improved reporting of analysis results including variant annotation using the Illumina Annotation Service.
- The variant filtering applied to NextSeq data has been improved. This feature improves the sensitivity and accuracy of SNV calling with NextSeq data, without increasing the false positive rate. The quality of SNV calls with data from other sequencing instruments is unaffected. Specifically, variants are filtered if more than 40% of the associated base calls have a Q-score less than 17.
- Introduced Illumina-curated gap regions for the hg19 reference build. This change only affects customers who enable SV/CNV calling when launching the app.

DEFECT REPAIRS:

- None

KNOWN ISSUES:

- Structural variant identifiers used in VCF files are not unique. This feature does not adhere to VCF 4.1 format specifications but does not impact the usability of the structural variant information in the VCF file. Some VCF validator tools return a warning for CNV.vcf files: "The header tag 'contig' not present for CHROM=chr1. (Not required but highly recommended.)”
- In some situations, there may be a discrepancy in the number of variants included in calculation of the “Total Passing” as compared to the “Variants by Sequence Context” and “Variants by Consequence” entries in the small variants summary table. It may be possible for the number in the “Total Passing” section to be less than the sum of all the variants in the “Variants by Sequence Context” section. This issue may occur because the “Total Passing” calculation includes only simple variants where the alternate alleles are all of the same type (either SNV, Insertion or Deletion). However, the other sections, “Variants by Sequence Context” and “Variants by

Consequence," also include variants at tri-allelic sites where the two alternate alleles are classified as different types (e.g. SNV and Insertion).

IV. BWA Whole Genome Sequencing v1.0.0.1

NEW FEATURES:

- None

DEFECT REPAIRS:

- Bug fix: In rare cases annotation fails and the "Variants by Sequence Context" and "Variants by Consequence" sections of the report would display all zeros.
- Bug fix: In some cases SV calling on NextSeq data would fail producing no SV.vcf file.
- Bug fix: In some cases, the number of SVs reported did not match the number of variants found in the SV.vcf file

KNOWN ISSUES:

- The BWA WGS app does not support analysis of extremely high coverage data sets. For data sets with estimated genomic coverage depth of 5,000,000x or higher, analysis with the BWA WGS may abort. For such data sets, use of the Isaac WGS app is recommended.
- VCF output files: In some instances, the variants reported in the VCF and gVCF files are not identical. Any discrepancies between these files include only non-passing variants or variants with unknown genotypes where there is an SNV inside a deletion. These events are rare and are not observed in most samples or data sets. For variants that involve SNVs inside a deletion, the variant call may only appear in the gVCF file and not the VCF file.
- VCF output files: In some cases, VCF validation tools report an inconsistency error due to the inclusion of more than one entry in the INFO tag. This issue may occur because the information in the AF field in the INFO tag may originate from the Illumina Annotation Service tool. The Illumina Annotation Service tool reports the Allele Frequency from all populations of 1000 Genomes project. However, if the information in the AF field in the INFO tag is either 0.5 or 1.0, then this indicates that this is the Allele Frequency calculated from the genotypes in the VCF file itself, and further investigation of the variant is recommended.
- Structural variant identifiers used in VCF files are not unique. This feature does not adhere to VCF 4.1 format specifications but does not impact the usability of the structural variant information in the VCF file. Some VCF validator tools may return a warning for CNV.vcf files: "The header tag 'contig' not present for CHROM=chr1. (Not required but highly recommended)".
- In some situations, there may be a discrepancy in the number of variants included in calculation of the "Total Passing" as compared to the "Variants by Sequence Context" and "Variants by Consequence" entries in the small variants summary table. It may be possible for the number in the "Total Passing" section to be less than the sum of all the variants in the "Variants by Sequence Context" section. This may occur because the "Total Passing" calculation includes only simple variants where the alternate alleles are all of the same type (either SNV, Insertion or Deletion).

However, the other sections, “Variants by Sequence Context” and “Variants by Consequence,” also include variants at tri-allelic sites where the two alternate alleles are classified as different types (e.g. SNV and Insertion).

V. Tumor Normal v1.0.0.1

NEW FEATURES:

- The variant filtering applied to NextSeq data has been improved. This feature improves the sensitivity and accuracy of germline SNV calling with NextSeq data, without increasing the false positive rate. The quality of germline SNV calls with data from other sequencing instruments is unaffected. Specifically, if more than 40% of the associated base calls have a Q-score less than 17, germline variants are filtered.
- Introduced Illumina-curated gap regions for the hg19 reference build. This change only affects customers who enable SV/CNV calling when launching the app.

DEFECT REPAIRS:

- An issue leading to the reporting of an incorrect number of structural variants in the Normal sample only has been fixed. The correct number of SVs are now reported for the Normal sample.

KNOWN ISSUES:

- The maximum small indel size reported in the somatic.Indels.vcf file is 10 bp. As a result, for somatic variants, indels with size greater than 10 appear in the somatic.SV.vcf file rather the somatic.Indels.vcf file. This reporting is in contrast to the normal sample germline variants. Here, indels less than or equal to 50 bases appear in the small variant VCF file while indels larger than 50 bases appear in the SV VCF file.
- The dbSNP concordance for all the variants from the data set reported in the resequencing statistics file does not match the number of dbSNP variants in the VCF file. The dbSNP concordance reported in the resequencing statistics file reflects SNVs, Insertions, and Deletions. The VCF file reports overall dbSNP concordance that includes all variants. If customers wish to calculate overall dbSNP concordance, they are directed to use the VCF file information.
- Structural variant identifiers used in VCF files are not unique. This feature does not adhere to VCF 4.1 format specifications but does not impact the usability of the structural variant information in the VCF file. Some VCF validator tools may return a warning for CNV.vcf files: "The header tag 'contig' not present for CHROM=chr1. (Not required but highly recommended)”.
- In some situations, there may be a discrepancy in the number of variants included in calculation of the “Total Passing” as compared to the "Variants by Sequence Context” and “Variants by Consequence” entries in the small variants summary table of the Normal Sample report. It may be possible for the number in the “Total Passing” section to be less than the sum of all the variants in the “Variants by Sequence Context” section. This issue may occur because the “Total Passing” calculation includes only simple variants where the alternate alleles are all of the same type

(either SNV, Insertion or Deletion). However, the other sections, “Variants by Sequence Context” and “Variants by Consequence,” also include variants at tri-allelic sites where the two alternate alleles are classified as different types (e.g. SNV and Insertion).