CeGaT workflow

Bioinformatics Development

Sandy

2025.02.21, 02.24

Overview

- STAR-based workflows
  - Side-by-side comparison for the 4 tools
  - STAR commands (Arriba, STAR-Fusion, rMATS)
  - CeGaT – intragene_events.tsv
- CeGaT workflow
  - Workflow steps (run.sh)
    - STAR-Fusion output preprocessing
    - FFPM correction and target variant filtering
    - FFPM correction and target variant filtering
- Appendices
  - STAR-Fusion commands comparison (CeGaT, STAR-Fusion)
  - intragene_events.tsv break point conversion (hg 19 -> GRCh38)
  - intragene_events.tsv (GRCh38)

ACT GENOMICS ™

STAR-based workflows

- Side-by-side comparison for the 4 tools
- STAR commands (Arriba, STAR-Fusion, rMATS)
- CeGaT – intragene_events.tsv

ACT GENOMICS ™

Side-by-side comparison for the 4 tools

- v0.28.0 (v4), Fusion v4 pipeline
- Arriba (2.4.0)
- STAR-Fusion
- CeGaT (customized STAR-Fusion)

Detailed comparison table can be found in the tool comparison table:
https://actg.atlassian.net/browse/ABIE-907

The exon annotator provided by **Arriba ("annotate_exon_numbers.sh") may be incorrect**.
=>
Since it does not utilize preferred transcripts (i.e. **it may randomly pick one transcript to annotate**).

Feature \ Tool	v0.28.0 (v4)	Arriba (2.4.0)	STAR-Fusion	CeGaT (customized STAR-Fusion)
Assay type (amplicon / hybrid-capture)	amplicon	RNA-seq	RNA-seq	hybrid-capture (provided by the Twist Alliance CeGaT RNA Fusion Panel)
Internal control	+	-	-	-
Supporting reads (span-read)	-	+	+	+
Supporting reads (split-read)	+	+	+	+
Break point detection (genomic)	+	+	+	+
Exon-level break point annotation	+	+(utility => no preferred transcript available)	-	-
Protein translation	+	+	+	
Target protein alignment	+	-	-	-
Consensus read	+ (utility)	-	+	-
Amplicon-based variant types (% of IVTALL variants within amplicon-based assay data) => Recall %	96%	85%	23%	NA (only supports hg19)
Hybrid capture-based variant types (% of IVTALL variants within hybrid capture data)	**not available**	**not available**	**not available**	**not available**
Support variants (fusion)	+	+	+	+
Support variants (AR-V7)	+	-	-	-
Support variants (KDD)	+	-	-	-
Consensus read	+ (outdated)	-	-	-
QC matrices	+ (outdated)	-	-	-

ACT GENOMICS ™

Feature \ Tool	v0.28.0 (v4), Fusion v4 pipeline	Arriba (2.4.0)	STAR-Fusion	CeGaT (customized STAR-Fusion)
Docs	within the pipeline repo	Home - Arriba	Home	https://cegat.com/fusions/
Assay type	amplicon-based	RNA seq	RNA seq	hybrid-capture
Reference genome/transcriptome	Grch38 **MANE v0.95** (+ GENCODE-r38 (NRG1 NTRK3, ERG (first 3 exons), AR (editing) => preferred transcripts Refseq + GENCODE => transcript isoform elimination	Grch38 (GRCh38_**RefSeq_hg38**) (other versions available)	Grch38 (GRCh38_**gencode**_v44) (other versions: https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/)	hg19
Aligner (sequence alignment tool)	bwa-mem	STAR	STAR	STAR
Tools for fusion detection	Fusion v4 pipeline (v0.28.0) https://bitbucket.org/actgenomics/torrent_fusion_pipeline_nextflow/src/master/	Arriba (2.4.0)	STAR-Fusion	STAR-Fusion
Tools for de novo fusion construction	Consensus read utility (https://bitbucket.org/actgenomics/tool_fusion_consensus_read/src/main/ => for ***v0.24.0*** pipeline (***outdated***))	NA	Trinity de novo transcriptome assembly (include de novo reconstruction: "**--denovo_reconstruct**") (STAR-Fusion submodule)	NA (may need to specify "--denovo_reconstruct")
Tools for read inspection and validation	Utility repo (https://bitbucket.org/actgenomics/torrent_fusion_pipeline_utilities/src/master/)	Utility scripts: (not included in the standard workflow) 1. extract_fusion-supporting_alignments.sh 2. convert_fusions_to_vcf.sh 3. run_arriba_on_prealigned_bam.sh 4. quantify_virus_expression.sh 5. **annotate_exon_numbers.sh => to annotate fusion.tsv => exon-level breakpoint** (Remark: it only annotates breakpoints with annotated NM ID within the .tsv file)	FusionInspector (validate mode: "**--FusionInspector validate**") (STAR-Fusion submodule)	FusionInspector (inspect mode: "**--FusionInspector inspect**") (STAR-Fusion submodule)
Support reads	Functional count, Total read count (filtered), Total read count Decision ("+" for report)	coverage1, coverage2, confidence (several built-in filters applied)	FFPM (fusion fragments per million total reads)	FFPM (fusion fragments per million total reads)
break point resolution	exon-level	exon-level (via "annotate_exon_numbers.sh") Remark: the annotated exons may differ from v4 (since no preferred transcript for annotation) => use provided break point "breakpoint1", "breakpoint2" (Arriba 2.4.0) ~ "5' gene coordinate", "3' gene coordinate" (v0.28.0 v4)	gene-level	gene-level
Variant report files	/Report/{sample name}**_fusioncalling.boundary.QC.txt**	/path_to_output/**fusions.tsv** /path_to_output/fusions.discarded.tsv (discarded variants)	/path_to_output/**star-fusion.fusion_predictions.tsv** /path_to_output/FusionInspector-validate/ (output folder for "**--FusionInspector validate**") /path_to_output/FusionInspector-validate/finspector.mm2_trinity_GG.fusions.fasta (output for "**--denovo_reconstruct**")	/path_to_output/**fusions.tsv** /path_to_output/**intragene_events.tsv**
Internal control files (sample QC)	/Report/{sample name}_fusioncalling.Sample.QC.json	NA	NA	NA
Performance # of detected IVTALL variants / # of IVTALL variants (=81) See sheet "IVTALL-1 comparison"	96%	85%	23%	NA (the genomic location for hg 19 and Grch 38 are not compatiable)
Pros	Most of the required columns are built-in One can adjust LoB/LoD, detection threshold based on assay design	Faster runtime many built-in filters applied for background noise correction	De novo fusion construction is supported Built-in Trinity module via "**--FusionInspector validate**"	Some of the predefined splicing variants
Cons	No spanning read support	Lacking of the required columns for report purpose Need to fine-tune the arguments to rescue some clinical relavant variants	Lacking of the required columns for report purpose Limited detected variants	Lacking of the required columns for report purpose Limited detected variants
Decision based on RM (hybrid-capture) Wait for sequencing data	**TBC**	**TBC**	**TBC**	**TBC**

ACT GENOMICS ™

Detailed comparison table can be found in the tool comparison table: https://actg.atlassian.net/browse/ABIE-907

STAR commands (Arriba, STAR-Fusion, rMATS)

=> Same aligners but different argument sets in used
=> Favor different fusion/splicing variants

Arriba, STAR-Fusion, rMATS (STAR-based splicing variant caller) [ABIE-947]

Parameter	**Arriba**	rMATS	**STAR-Fusion**
Genome Index	--genomeDir /path/to/STAR_index	--genomeDir /path/to/STAR_index	--genomeDir ${star_index_dir}
Input Reads	--readFilesIn read1.fastq.gz read2.fastq.gz	--readFilesIn <R1.fastq.gz> <R2.fastq.gz>	--readFilesIn ${left_fq_filename} ${right_fq_filename}
Read Command	--readFilesCommand zcat	--readFilesCommand zcat	--readFilesCommand "gunzip -c"
Threads	--runThreadN 8	--runThreadN 4	Not explicitly provided
Output Format	--outStd BAM_Unsorted	--outSAMtype BAM SortedByCoordinate	--outReadsUnmapped None
SAMtype (Output Alignment)	--outSAMtype BAM Unsorted	--outSAMtype BAM SortedByCoordinate	Not specified for BAM, but intronMotif field is included (--outSAMstrandField intronMotif)
Output Unmapped	--outSAMunmapped Within	Not specified	--outSAMunmapped Within
Two-Pass Mode	Not specified	--twopassMode Basic	--twopassMode Basic
Chimeric Segment Minimum Length	--chimSegmentMin 10	Not specified	--chimSegmentMin 12
Chimeric Junction Overhang Minimum	--chimJunctionOverhangMin 10	Not specified	--chimJunctionOverhangMin 8
Chimeric Out Junction Format	Not specified	Not specified	--chimOutJunctionFormat 1
Maximum Multimapping	--chimMultimapNmax 50	Not specified	--chimMultimapNmax 20
Multimap Score Range	Not specified	Not specified	--chimMultimapScoreRange 3
Align Mate Gap Maximum	Not specified	Not specified	--alignMatesGapMax 100000
Align Intron Max	Not specified	--alignIntronMax 299999	--alignIntronMax 100000
Soft Clipping on Spliced Mates	--alignSplicedMateMapLminOverLmate 0.5	Not specified	--alignSplicedMateMapLminOverLmate 0
Read Overlap	--peOverlapNbasesMin 10	Not specified	--peOverlapNbasesMin 12
Splice Junction Stitch Mismatch	--alignSJstitchMismatchNmax 5 -1 5 5	Not specified	--alignSJstitchMismatchNmax 5 -1 5 5
Alignment Flush	Not specified	Not specified	--alignInsertionFlush Right
Multimap Non-GTAG Score	--chimScoreJunctionNonGTAG 0	Not specified	--chimScoreJunctionNonGTAG -4
Output RG Line	Not specified	Not specified	--outSAMattrRGline ID:GRPundef

GENOMICS ™

Ref. issue:
https://actg.atlassian.net/browse/ABIE-963

CeGaT – intragene_events.tsv

#fusion_name	gene	chr	start	stop	FFPM	canonical_support	chimeric_support
EGFR_VIII	EGFR	chr7	55087059	55223522	0.47	0	1
MET_ex14_skipping	MET	chr7	116411709	116414934	274.55	573	0

- CeGaT pipeline output
  output file:
  /mnt/RD_Develop/sandyteng/FusionCaptureTools/testresult/cegat_starfusion/intragene_events.tsv

- 16 predefined intragene_events.tsv
  built-in file found within container:
  /intragene_events.tsv

#fusion_name	chr	Start (1- based)	stop	gene	junction_cigar
EGFR_VIII	chr7	55087059	55223522	EGFR	136464N
EGFR_VII	chr7	55229325	55238867	EGFR	9543N
EGFR_del_ex25-26	chr7	55268107	55270209	EGFR	2103N
EGFR_del_ex25-27	chr7	55268107	55272948	EGFR	4842N
MET_ex14_skipping	chr7	116411709	116414934	MET	3226N
EGFR_del_ex2-3	chr7	55087059	55214298	EGFR	127240N
EGFR_del_ex2-4	chr7	55087059	55218986	EGFR	131928N
EGFR_del_ex2-14	chr7	55087059	55232972	EGFR	145914N
EGFR_del_ex2-22	chr7	55087059	55266409	EGFR	179351N
EGFR_del_ex5-6	chr7	55214434	55221703	EGFR	7270N
EGFR_del_ex6-7	chr7	55219056	55223522	EGFR	4467N
EGFR_del_ex9	chr7	55223640	55224451	EGFR	812N
EGFR_del_ex9-10	chr7	55223640	55225355	EGFR	1716N
EGFR_del_ex10	chr7	55224353	55225355	EGFR	1003N
EGFR_del_ex12	chr7	55225447	55229191	EGFR	3745N
EGFR_del_ex26-27	chr7	55269049	55272948	EGFR	3900N

Remark:
The built-in genomic coordinates for the predefined intra-gene events are based on hg19.
=>
We may convert them to GRCh 38 coordinates via https://genome.ucsc.edu/cgi-bin/hgLiftOver.

ACT GENOMICS ™

CeGaT workflow

- Workflow steps (run.sh)
  - **STAR-Fusion output preprocessing**
  - **FFPM correction and target variant filtering**
  - **FFPM correction and target variant filtering**

ACT GENOMICS ™

CeGaT – fusion detection workflow

Ref. issue:
https://actg.atlassian.net/browse/ABIE-963

- Running StarFusion 1.8.1
  - Obtain STAR and STAR-Fusion output files
    - STAR output: Chimeric.out.junction, Aligned.out.bam
    - star-fusion.fusion_predictions.abridged.coding_effect.tsv)
- Sorting and indexing alignments
  - Aligned.out.bam preprocessing (samtools fixmate => sort => index) # $SORTEDBAM
- Counting mapped reads
  - Total reads (TOTAL), deduplicated reads (DEDUP), duplicated rate (DUP) calculation
    - Total reads: value obtained from Log.final.out (STAR output, “Number of input reads”)
    - Duplicated: value calculated from $SORTEDBAM (samtools markdup -r => samtools view -F 256 => # of deduplicated reads = the number of unique lines (except for lines started with ‘@’))
- Preparing Fusion Results
  - Use raw STAR-Fusion output as a fusion report template
    - Obtain “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and
    - the file stored within /FusionInspector-inspect/ folder (“finspector.consolidated.cSorted.bam”, “finspector.consolidated.cSorted.bam.bai”, “finspector.fusion_inspector_web.html”)
- Correcting FFPM values
  - Correct the FFPM column within the “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and keep only the rows with FFPM_corrected >= 0.1
    - FFPM_corrected = FFPM * TOTAL / DEDUP
- Filtering to targets (defined in /target.bed)
  - Target bed filter to remove off-target variants reported by STAR-Fusion (variants recorded within “star-fusion.fusion_predictions.abridged.coding_effect.tsv”)
    - The filter identifies fusion events whose left or right breakpoints overlap with any target region (from /targets.bed) extended by 1,000 bp on both sides.
- Calling intragene events (=> intragene_events.tsv)
  - Chimeric support reads extraction
    - Chimeric.out.junction (one of the STAR-Fusion output files) → obtain columns 1-14 → chimeric support reads for each predefined fusion
  - Canonical support reads extraction
    - (Aligned.sorted.bam) => STAR
    - fusions_evidence_mapped.bam (one of the STAR-Fusion output files (generated by “STAR”), except it has been sorted using “samtools fixmate” + “samtools sort”) → canonical support reads for each predefined fusion
- Generating evidence BAM
  - Add the support reads of the detected intragene events to “finspector.consolidated.cSorted.bam” => output it as “fusions_evidence_mapped.bam”

ACT GENOMICS ™

Ref. issue:
https://actg.atlassian.net/browse/ABIE-963

CeGaT – STAR-Fusion output preprocessing

- Running StarFusion 1.8.1
  - Obtain STAR and STAR-Fusion output files
    - STAR output: Chimeric.out.junction, Aligned.out.bam
    - star-fusion.fusion_predictions.abridged.coding_effect.tsv)
- Sorting and indexing alignments
  - Aligned.out.bam preprocessing (samtools fixmate => sort => index) # $SORTEDBAM
- Counting mapped reads
  - Total reads (TOTAL), deduplicated reads (DEDUP), duplicated rate (DUP) calculation
    - Total reads: value obtained from Log.final.out (STAR output, “Number of input reads”)
    - Duplicated: value calculated from $SORTEDBAM (samtools markdup -r => samtools view -F 256 => # of deduplicated reads = the number of unique lines (except for lines started with ‘@’))
- Preparing Fusion Results
  - Use raw STAR-Fusion output as a fusion report template
    - Obtain “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and
    - the file stored within /FusionInspector-inspect/ folder (“finspector.consolidated.cSorted.bam”, “finspector.consolidated.cSorted.bam.bai”, “finspector.fusion_inspector_web.html”)
- Correcting FFPM values
  - Correct the FFPM column within the “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and keep only the rows with FFPM_corrected >= 0.1
    - FFPM_corrected = FFPM * TOTAL / DEDUP
- Filtering to targets (defined in /target.bed)
  - Target bed filter to remove off-target variants reported by STAR-Fusion (variants recorded within “star-fusion.fusion_predictions.abridged.coding_effect.tsv”)
    - The filter identifies fusion events whose left or right breakpoints overlap with any target region (from /targets.bed) extended by 1,000 bp on both sides.
- Calling intragene events (=> intragene_events.tsv)
  - Chimeric support reads extraction
    - Chimeric.out.junction (one of the STAR-Fusion output files) → obtain columns 1-14 → chimeric support reads for each predefined fusion
  - Canonical support reads extraction
    - (Aligned.sorted.bam) => STAR
    - fusions_evidence_mapped.bam (one of the STAR-Fusion output files (generated by “STAR”), except it has been sorted using “samtools fixmate” + “samtools sort”) → canonical support reads for each predefined fusion
- Generating evidence BAM
  - Add the support reads of the detected intragene events to “finspector.consolidated.cSorted.bam” => output it as “fusions_evidence_mapped.bam”

ACT GENOMICS ™

Ref. issue:
https://actg.atlassian.net/browse/ABIE-963

CeGaT – FFPM correction and target variant filtering

- Running StarFusion 1.8.1
  - Obtain STAR and STAR-Fusion output files
    - STAR output: Chimeric.out.junction, Aligned.out.bam
    - star-fusion.fusion_predictions.abridged.coding_effect.tsv)
- Sorting and indexing alignments
  - Aligned.out.bam preprocessing (samtools fixmate => sort => index) # $SORTEDBAM
- Counting mapped reads
  - Total reads (TOTAL), deduplicated reads (DEDUP), duplicated rate (DUP) calculation
    - Total reads: value obtained from Log.final.out (STAR output, “Number of input reads”)
    - Duplicated: value calculated from $SORTEDBAM (samtools markdup -r => samtools view -F 256 => # of deduplicated reads = the number of unique lines (except for lines started with ‘@’))
- Preparing Fusion Results
  - Use raw STAR-Fusion output as a fusion report template
    - Obtain “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and
    - the file stored within /FusionInspector-inspect/ folder (“finspector.consolidated.cSorted.bam”, “finspector.consolidated.cSorted.bam.bai”, “finspector.fusion_inspector_web.html”)
- Correcting FFPM values
  - Correct the FFPM column within the “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and keep only the rows with FFPM_corrected >= 0.1
    - FFPM_corrected = FFPM * TOTAL / DEDUP
- Filtering to targets (defined in /target.bed)
  - Target bed filter to remove off-target variants reported by STAR-Fusion (variants recorded within “star-fusion.fusion_predictions.abridged.coding_effect.tsv”)
    - The filter identifies fusion events whose left or right breakpoints overlap with any target region (from /targets.bed) extended by 1,000 bp on both sides.
- Calling intragene events (=> intragene_events.tsv)
  - Chimeric support reads extraction
    - Chimeric.out.junction (one of the STAR-Fusion output files) → obtain columns 1-14 → chimeric support reads for each predefined fusion
  - Canonical support reads extraction
    - (Aligned.sorted.bam) => STAR
    - fusions_evidence_mapped.bam (one of the STAR-Fusion output files (generated by “STAR”), except it has been sorted using “samtools fixmate” + “samtools sort”) → canonical support reads for each predefined fusion
- Generating evidence BAM
  - Add the support reads of the detected intragene events to “finspector.consolidated.cSorted.bam” => output it as “fusions_evidence_mapped.bam”

ACT GENOMICS ™

Ref. issue:
https://actg.atlassian.net/browse/ABIE-963

CeGaT – FFPM correction and target variant filtering

- Running StarFusion 1.8.1
  - Obtain STAR and STAR-Fusion output files
    - STAR output: Chimeric.out.junction, Aligned.out.bam
    - star-fusion.fusion_predictions.abridged.coding_effect.tsv)
- Sorting and indexing alignments
  - Aligned.out.bam preprocessing (samtools fixmate => sort => index) # $SORTEDBAM
- Counting mapped reads
  - Total reads (TOTAL), deduplicated reads (DEDUP), duplicated rate (DUP) calculation
    - Total reads: value obtained from Log.final.out (STAR output, “Number of input reads”)
    - Duplicated: value calculated from $SORTEDBAM (samtools markdup -r => samtools view -F 256 => # of deduplicated reads = the number of unique lines (except for lines started with ‘@’))
- Preparing Fusion Results
  - Use raw STAR-Fusion output as a fusion report template
    - Obtain “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and
    - the file stored within /FusionInspector-inspect/ folder (“finspector.consolidated.cSorted.bam”, “finspector.consolidated.cSorted.bam.bai”, “finspector.fusion_inspector_web.html”)
- Correcting FFPM values
  - Correct the FFPM column within the “star-fusion.fusion_predictions.abridged.coding_effect.tsv” and keep only the rows with FFPM_corrected >= 0.1
    - FFPM_corrected = FFPM * TOTAL / DEDUP
- Filtering to targets (defined in /target.bed)
  - Target bed filter to remove off-target variants reported by STAR-Fusion (variants recorded within “star-fusion.fusion_predictions.abridged.coding_effect.tsv”)
    - The filter identifies fusion events whose left or right breakpoints overlap with any target region (from /targets.bed) extended by 1,000 bp on both sides.
- Calling intragene events (=> intragene_events.tsv)
  - Chimeric support reads extraction
    - Chimeric.out.junction (one of the STAR-Fusion output files) → obtain columns 1-14 → chimeric support reads for each predefined fusion
  - Canonical support reads extraction
    - (Aligned.sorted.bam) => STAR
    - fusions_evidence_mapped.bam (one of the STAR-Fusion output files (generated by “STAR”), except it has been sorted using “samtools fixmate” + “samtools sort”) → canonical support reads for each predefined fusion
- Generating evidence BAM
  - Add the support reads of the detected intragene events to “finspector.consolidated.cSorted.bam” => output it as “fusions_evidence_mapped.bam”

ACT GENOMICS ™

Appendices

- STAR-Fusion commands comparison (CeGaT, STAR-Fusion)
- intragene_events.tsv break point conversion (hg 19 -> GRCh38)
- intragene_events.tsv (GRCh38)

ACT GENOMICS ™

STAR-Fusion commands comparison (CeGaT, STAR-Fusion)

- Comparison table (Remark: the 0.1 filter is applied after STAR-Fusion in CeGaT workflow)

	STAR-Fusion 1.8.1 (CeGaT workflow)	**STAR-Fusion 1.14.0** **(**trinityctat/starfusion:latest, 2024.12.12) (STAR-Fusion workflow)
--extract_fusion_reads	retrieves the fusion supporting reads from the fastq files	Not specified
--FusionInspector	“inspect“ include FusionInspector, options: • 'inspect' - considers only StarFusion-identified fusion reads in context of identified fusions (fast) • 'validate' - examines all reads, recovers evidence, computes fusion allele fractions (slow)	“validate“ same
--no_single_fusion_per_breakpoint	exclude filtering of potentially superfluous fusions involving different gene annotations overlapping the same fusion breakpoint.	Not specified (default: superfluous fusions filter is enabled)
--min_FFPM	minimum FFPM (fusion fragments per million rna-seq frags) (default: 0.1)	Not specified (default: 0.1)
--tmpdir	file for temporary files (default: /tmp)	Not specified

Ref. Issues:
https://actg.atlassian.net/browse/ABIE-907
https://actg.atlassian.net/browse/ABIE-963

ACT GENOMICS ™

intragene_events.tsv break point conversion (hg 19 -> GRCh38)

- Genome coordinate liftover
  https://genome.ucsc.edu/cgi-bin/hgLiftOver

							A	B			B-A
#fusion_name	chr	start (1-based) the end of 5' exon	stop (1-based) the start of 3' exon	gene	junction_cigar	chr	start (0-based) the end of 5' exon	stop (0-based) the start of 3' exon	input coordinates provided by liftover website (1-based)	base (0/1)	# of deletion bases
EGFR_VIII	chr7	55087059	55223522	EGFR	136464N	chr7	55019365	55155829	chr7:55087059-55223522	1	136464
EGFR_VII	chr7	55229325	55238867	EGFR	9543N	chr7	55161631	55171174	chr7:55229325-55238867	1	9543
EGFR_del_ex25-26	chr7	55268107	55270209	EGFR	2103N	chr7	55200413	55202516	chr7:55268107-55270209	1	2103
EGFR_del_ex25-27	chr7	55268107	55272948	EGFR	4842N	chr7	55200413	55205255	chr7:55268107-55272948	1	4842
MET_ex14_skipping	chr7	116411709	116414934	MET	3226N	chr7	116771654	116774880	chr7:116411709-116414934	1	3226
EGFR_del_ex2-3	chr7	55087059	55214298	EGFR	127240N	chr7	55019365	55146605	chr7:55087059-55214298	1	127240
EGFR_del_ex2-4	chr7	55087059	55218986	EGFR	131928N	chr7	55019365	55151293	chr7:55087059-55218986	1	131928
EGFR_del_ex2-14	chr7	55087059	55232972	EGFR	145914N	chr7	55019365	55165279	chr7:55087059-55232972	1	145914
EGFR_del_ex2-22	chr7	55087059	55266409	EGFR	179351N	chr7	55019365	55198716	chr7:55087059-55266409	1	179351
EGFR_del_ex5-6	chr7	55214434	55221703	EGFR	7270N	chr7	55146740	55154010	chr7:55214434-55221703	1	7270
EGFR_del_ex6-7	chr7	55219056	55223522	EGFR	4467N	chr7	55151362	55155829	chr7:55219056-55223522	1	4467
EGFR_del_ex9	chr7	55223640	55224451	EGFR	812N	chr7	55155946	55156758	chr7:55223640-55224451	1	812
EGFR_del_ex9-10	chr7	55223640	55225355	EGFR	1716N	chr7	55155946	55157662	chr7:55223640-55225355	1	1716
EGFR_del_ex10	chr7	55224353	55225355	EGFR	1003N	chr7	55156659	55157662	chr7:55224353-55225355	1	1003
EGFR_del_ex12	chr7	55225447	55229191	EGFR	3745N	chr7	55157753	55161498	chr7:55225447-55229191	1	3745
EGFR_del_ex26-27	chr7	55269049	55272948	EGFR	3900N	chr7	55201355	55205255	chr7:55269049-55272948	1	3900

intragene_events.tsv (1-based, hg19)

intragene_events.GRCh38.bed (0-based, GRCh38)

Ref. files:
- /mnt/RD_Develop/sandyteng/FusionCaptureTools/cegat_starfusion/intragene_events.tsv (1-based, hg19)
- /mnt/RD_Develop/sandyteng/FusionCaptureTools/cegat_starfusion/intragene_events.hg19.bed (0-based, hg19)
- /mnt/RD_Develop/sandyteng/FusionCaptureTools/cegat_starfusion/intragene_events.GRCh38.bed (0-based, GRCh38)

Ref. issue:
https://actg.atlassian.net/browse/ABIE-963

ACT GENOMICS ™

intragene_events.tsv (GRCh38)

- Local directory
  - /mnt/RD_Develop/sandyteng/FusionCaptureTools/cegat_starfusion/intragene_events.GRCh38.tsv
- Script for genomic coordinates conversion
  - /mnt/RD_Develop/sandyteng/ACTFusionV5/code/intragene_events_liftover.py
- Break point searching
  - STAR-Fusion results:
    - Aligned.out.bam, Chimeric.out.junction
      - EGFR_VIII:
        - 1 chimeric read
      - MET_ex14_skipping:
        - 682 canonical reads

#fusion_name	chr	start	stop	gene	junction_cigar
EGFR_VIII	chr7	55019366	55155829	EGFR	136464N
EGFR_VII	chr7	55161632	55171174	EGFR	9543N
EGFR_del_ex25-26	chr7	55200414	55202516	EGFR	2103N
EGFR_del_ex25-27	chr7	55200414	55205255	EGFR	4842N
MET_ex14_skipping	chr7	116771655	116774880	MET	3226N
EGFR_del_ex2-3	chr7	55019366	55146605	EGFR	127240N
EGFR_del_ex2-4	chr7	55019366	55151293	EGFR	131928N
EGFR_del_ex2-14	chr7	55019366	55165279	EGFR	145914N
EGFR_del_ex2-22	chr7	55019366	55198716	EGFR	179351N
EGFR_del_ex5-6	chr7	55146741	55154010	EGFR	7270N
EGFR_del_ex6-7	chr7	55151363	55155829	EGFR	4467N
EGFR_del_ex9	chr7	55155947	55156758	EGFR	812N
EGFR_del_ex9-10	chr7	55155947	55157662	EGFR	1716N
EGFR_del_ex10	chr7	55156660	55157662	EGFR	1003N
EGFR_del_ex12	chr7	55157754	55161498	EGFR	3745N
EGFR_del_ex26-27	chr7	55201356	55205255	EGFR	3900N

Ref. issue:
https://actg.atlassian.net/browse/ABIE-963

ACT GENOMICS ™