Analysis result

Patient name	Jane Doe	Analysis date	2022-09-26 10:54:32
Patient ID	19790525-1234	Model	Acute Lymphoblastic Leukemia
Sample file	case_22.bam	Model version	1.0.918
Reference Genome	hg19	SW version	0.9.10
Gene fusion file(s)	star-fusion.fusion_predictions.abridged.coding_effe	Technology	RNA Seq.

Method description RNA-seq based analysis for detection of gene fusions and gene expression classification for Acute Lymphoblastic Leukemia.

Conclusion The subtype based on gene expression signature for patient Jane Doe is KMT2A(MLL)-rearranged with confidence level 1.0. The following gene fusions have been detected: KMT2A::MLLT1.

Classification

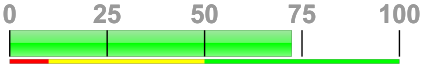
Group	Probability
KMT2A(MLL)-rearranged	1.00
High hyperdiploidy	0.00
BCR::ABL1 or BCR::ABL1-like	0.00
ETV6::RUNX1 or ETV6::RUNX1-like	0.00
DUX4-rearranged	0.00
TCF3::PBX1	0.00

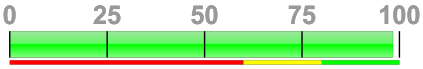
Gene fusion(s) of significance

Gene fusion	Breakpoint reads	Spanning fragments
KMT2A::MLLT1	6	3

Sequencing quality

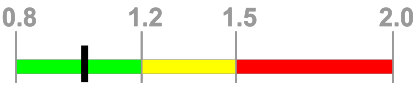
Number of fragments mapped	11342627
Number of fragments mapped to features	8208376
Mapped reads mapping to features (%)	72.4
BAM read as stranded?	*No*
Reads fractions	
Ambiguous	0.02 (0.00)
Forward	0.49 (0.50)
Reverse	0.49 (0.50)
Combined quality (%)	98.4
Paired reads insert size	mean = 2.66, std = 78.01





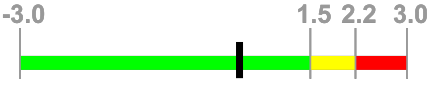
Classifier model characteristics

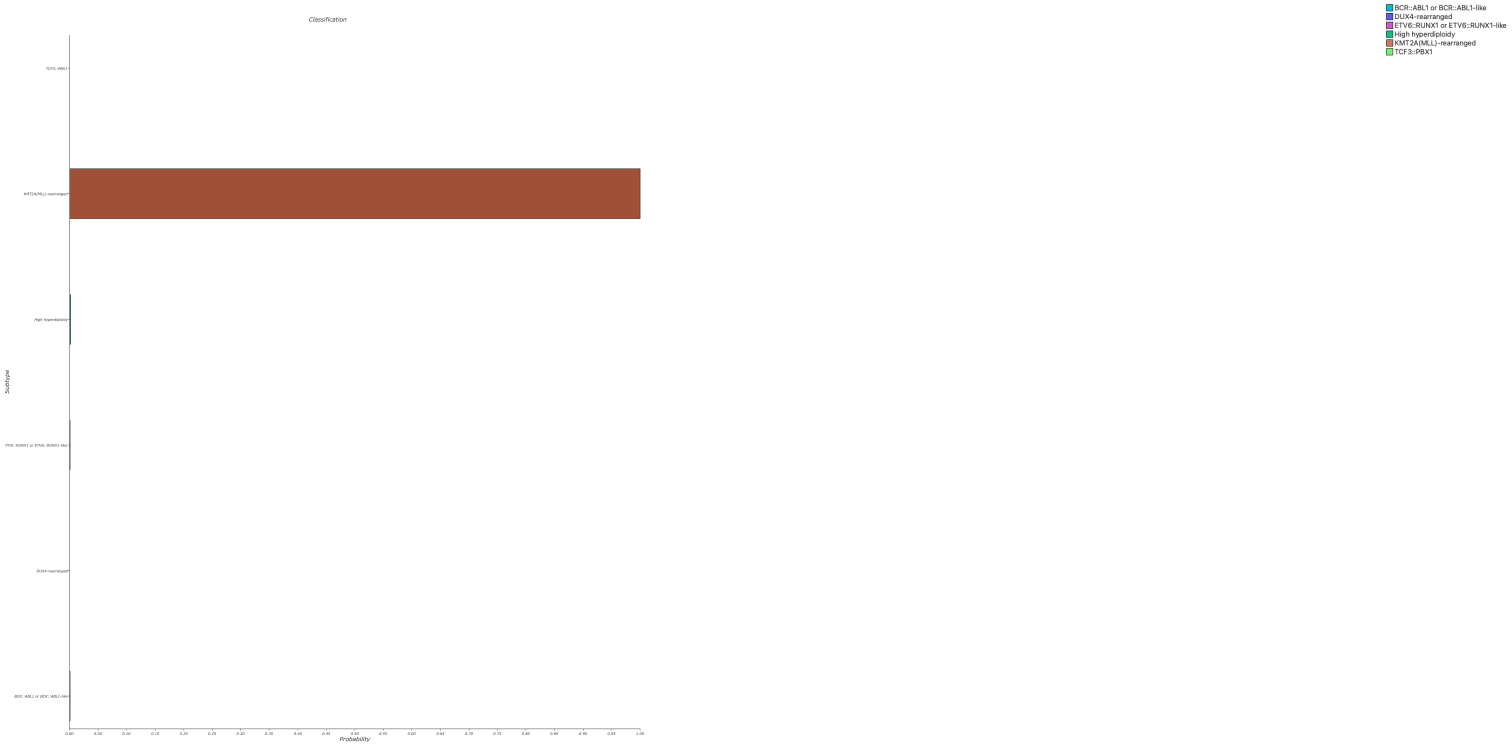
Local outlier factor	1.02



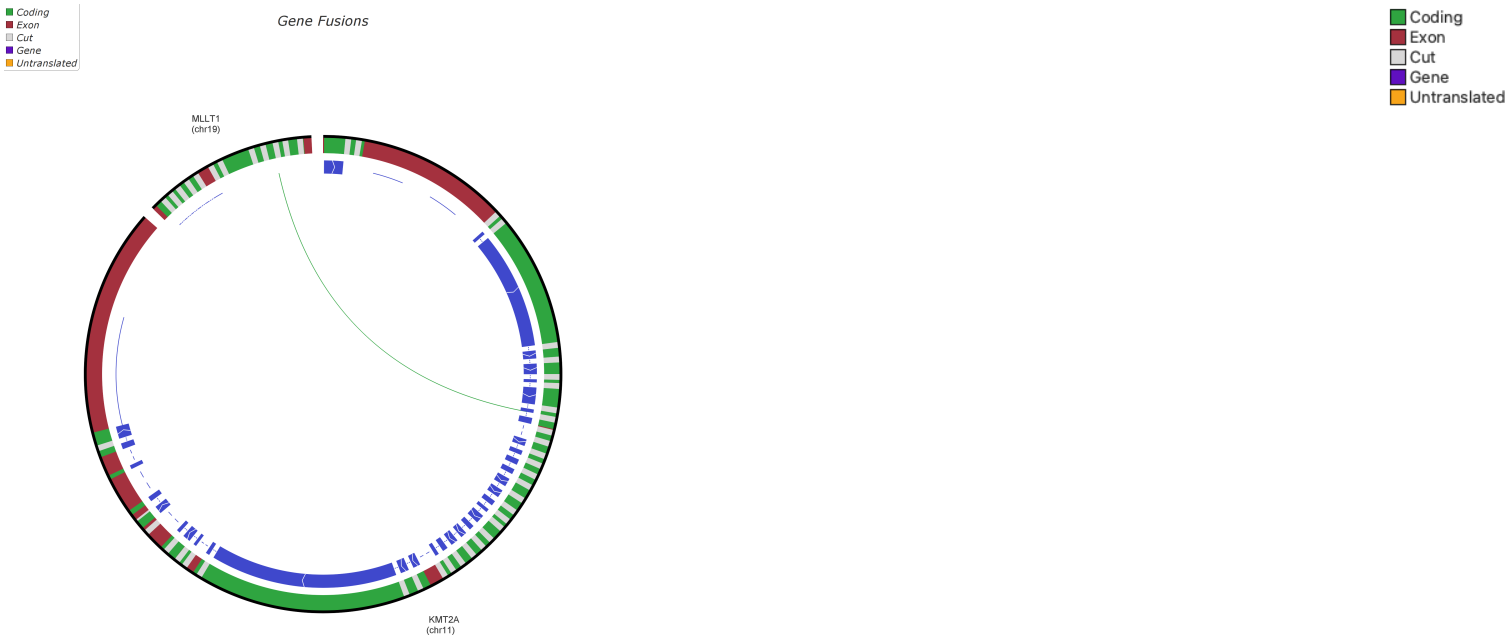
Sample assessment

Predicted normal cell content	0.403




Classification
TCF3::PBX1
KMT2A(MLL)-rearranged
High hyperdiploidy
ETV6::RUNX1 or ETV6::RUNX1-like
DUX4-rearranged
BCR::ABL1 or BCR::ABL1-like
Subtype
0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Probability
BCR::ABL1 or BCR::ABL1-like
DUX4-rearranged
ETV6::RUNX1 or ETV6::RUNX1-like
High hyperdiploidy
KMT2A(MLL)-rearranged
TCF3::PBX1


Classified: KMT2A(MLL)-rearranged
BCR::ABL1 or BCR::ABL1-like
DUX4-rearranged
ETV6::RUNX1 or ETV6::RUNX1-like
High hyperdiploidy
KMT2A(MLL)-rearranged
TCF3::PBX1
Unassigned
2 (16 %)
Classified: KMT2A(MLL)-rearranged
1 (49 %


Coding
Exon
Cut
Gene
Untranslated
Gene Fusions
MLLT1
(chr19)
KMT2A
(chr11)
Coding
Exon
Cut
Gene
Untranslated

Filters used in gene fusion detection

Existence in public database(s) : Mitelman or TumorFusions

All gene fusions detected in sample

Gene fusion	Breakpoint reads	Spanning fragments
AC131097.3::AC093642.5	1	3
ACIN1::TRDC	1	2
CIRBP::TMPRSS9	2	0
KMT2A::MLLT1	6	3
MBNL1::HMGB1	2	0
MBNL1::RSRC1	1	1
MLLT1::ANGPTL4	1	1
MLLT1::RPS28	2	0
RP11-96H19.1::RP11-446N19.1	2	1
RP11-367G6.3::FAM65B	2	5
RP11-494M8.4::OVCH2	3	0
RPS29::AOAH	2	0
TCF3::IQCB1	2	1
TIMM23::PARGP1	13	1
TRDV2::TRAC	4	3

Appendix: Quality measurements in this report

Number of fragments mapped

The number of fragments (read pairs for paired-end sequencing) in the sample that was mapped to the reference genome. A fragment is mapped if there is at least one region in the reference genome with a sequence similar to the fragment.

Number of fragments mapped to features

The number of fragments that are unambiguously assigned to a feature. The features are formed by taking the union of all exons of all transcripts of a gene and a fragment is assigned to a feature if it is completely contained in the feature. Fragments assigned to multiple features are not counted.

Mapped reads mapping to features (%)

The "Number of fragments mapped to features" divided by "Number of fragments mapped", i.e. the proportion of fragments providing gene expression information.

Stranded

Whether the classifier expects the sample to be sequenced using a stranded RNA-seq protocol or not. This entity can have any of the following values: "No", "Yes" and "Reverse".

Reads fractions

How reads are assigned to strands is a bit complicated, but depends on strand of the feature, the strand of the read and whether the read is the first or second read in the pair. A basic explanation follows: If stranded="Yes", almost all read pairs should be assigned to the Forward strand, if stranded="Reverse", almost all read pairs should be assigned to the Reverse strand and if stranded="No" roughly half should be assigned to each. Read pairs are considered ambiguous if they cannot be assigned to a strand (because of overlapping genes on different strands). Only read pairs within features are included when computing the read fractions.

Combined quality

How close the actual reads fractions are to the expected result. A value close to 100 is good and means that the sample was sequenced with the expected strandedness protocol.

Paired reads insert size

The insert size is the distance between the end of the first read and the beginning of the second read in a fragment. The value can be negative if the two reads overlap. Since the nucleotides between the first and second reads are unknown, the insert size must be estimated from the distance in the reference genome sequence. Introns are excluded when estimating the insert size and outlier fragments (fragments with an estimated insert size below -250 or above 250) are ignored.

Local Outlier Factor

The Local Outlier Factor is a way to assess whether a sample is an outlier or not. Local densities are computed for the sample as well as its neighbors. The density is higher if there are more samples close by. By comparing the density of the sample to the average densities of its neighbors, we get the Local Outlier Factor. Values close to 1 means that the sample has a density comparable to its neighbors - the sample is not an outlier by this measure. A value below 1 is rare, but also indicates that the sample is not an outlier. Large values indicate that the sample is different from the neighboring samples. The local outlier factor does not give any information on why the sample is different, the reasons behind this may be biological or technical.

Appendix: References for the Acute Lymphoblastic Leukemia (ALL) model

The ALL model contains a classifier that has been built using a boosted decision tree approach as implemented in the XGBoost software package. This classifier has been trained on curated gene expression data from known cases of ALL. In the training process, relevant variables were chosen partly based on specific knowledge about the ALL diagnosis, partly based on a number of U-tests performed in the process.

Technical requirements on sample files

- BAM files should be aligned to reference genome hg19 using STAR. Use STAR version 2.7.9a and the CTAT resource library "GRCh37_gencode_v19_CTAT_lib_Mar012021.plug-n-play" for this.
- Sample fusion files should be in supported file format (.tsv, .txt) as created with any of the following supported fusion callers:
    - STAR-Fusion
    - Fusion-Catcher
    - Arriba
- The ALL model was developed with data aligned to reference genome hg19 using STAR and the following commands:

```
STAR --genomeDir path/to/
GRCh37_gencode_v19_CTAT_lib_Mar012021.plug-n-play/
ctat_genome_lib_build_dir/ref_genome.fa.star.idx --readFilesIn
mysample_R1_001.fastq.gz mysample_R2_001.fastq.gz --runThreadN
40 --outReadsUnmapped None --twopassMode Basic --
readFilesCommand zcat --outSAMstrandField intronMotif --
outSAMunmapped Within --chimSegmentMin 12 --
chimJunctionOverhangMin 8 --chimOutJunctionFormat 1 --
alignSJDBoverhangMin 10 --alignMatesGapMax 100000 --
alignIntronMax 100000 --alignSJstitchMismatchNmax 5 -1 5 5 --
outSAMattrRGline ID:GRPundef --chimMultimapScoreRange 3 --
chimScoreJunctionNonGTAG -4 --chimMultimapNmax 20 --
chimNonchimScoreDropMin 10 --peOverlapNbasesMin 12 --
peOverlapMMp 0.1 --alignInsertionFlush Right --
alignSplicedMateMapLminOverLmate 0 --alignSplicedMateMapLmin
30 --outFilterMultimapNmax 200 --outSAMtype BAM Unsorted
```

, where --genomeDir, --readFilesIn, --runThreadN and possibly --readFilesCommand should be adjusted depending on file locations, file formats and number of available threads.

- Sample fusion files were produced with STAR-Fusion using reference genome hg19 and this command:

```
STAR-Fusion --genome_lib_dir path/to/
GRCh37_gencode_v19_CTAT_lib_Mar012021.plug-n-play/
ctat_genome_lib_build_dir/ --CPU 40 -J path/to/
Chimeric.out.junction --output_dir star_fusion
```

, where -J should refer to the file created by STAR and --CPU should be adjusted to the number of available threads.

Method references

Below follows references for some specific concepts that are central to the function of Qlucore Acute Lymphoblastic Leukemia (ALL) model.

TMM Normalization

M.D. Robinson, A. Oshlack, A scaling normalization method for differential expressionanalysis of RNA-seq data, Genome Biology (2010).

SAM/BAM File Import

Li H.*, Handsaker B.*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.

Crossvalidation of classifier

G James, D Witten, T Hastie, R Tibshirani: An introduction to statistical learning. Springer-Verlag 2013.

STAR

Dobin, Alexander, et al. "STAR: ultrafast universal RNA-seq aligner." Bioinformatics 29.1 (2013): 15-21.

STAR-Fusion

Haas, B.J., Dobin, A., Li, B. et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol 20, 213 (2019) doi:10.1186/s13059-019-1842-9

Fusion-Catcher

Nicorici D, Satalan M, Edgren H, et al. FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv, Nov. 2014, doi:10.1101/011650

Arriba

Uhrig S, Ellermann J, Walther T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Research*. Published in Advance January 13, 2021, doi:10.1101/gr.257246.119

Mitelman

Mitelman F, Johansson B and Mertens F (Eds.). "Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2019)". http://cgap.nci.nih.gov/Chromosomes/Mitelman

TumorFusions

Hu, Xin, et al. "TumorFusions: an integrative resource for cancer-associated transcript fusions." Nucleic acids research 46.D1 (2017): D1144-D1149.

Disclaimer

The contents of this document are subject to revision without notice due to continuous progress in methodology, design, and manufacturing. Qlucore shall have no liability for any error or damages of any kind resulting from the use of this document. Qlucore Insights and its "Acute Lymphoblastic Leukemia" model are only intended for research purposes.