Author: Professor Thoas Fioretos
Revision: E
Date: 2024-06-14

White paper:

Genetic classification of acute lymphoblastic leukemia (ALL) – RNA-sequencing as an emerging clinical diagnostic tool

BACKGROUND

Acute lymphoblastic leukemia (ALL) is a highly heterogenous disease characterized by different genetic alterations of critical importance for a correct clinical diagnosis, disease stratification, disease follow up and treatment selection. ALL is the most common acute leukemia in children and accounts for approximately 30% of all cases of acute leukemia in adults. The incidence of ALL is approximately 1.7 cases per 100 000 with 6,150 cases diagnosed each year in the US (1).

With the introduction of next-generation sequencing (NGS) it has become possible to detect most genetic alterations in ALL both at the DNA- and RNA-level (2, 3). Many of these alterations provide important diagnostic and prognostic information, with some also forming the basis for so called targeted therapies. According to the most recent World Health Organization (WHO) classification in 2022 (4), genetically defined entities of ALL include a total of 13 subtypes (Table 1). Collectively, these subtypes represent about 90-95% of both childhood and adult cases of ALL, although the frequency of each subtype differs substantially between the age groups (2, 3). The recent International Consensus Classification (ICC) is more detailed and includes 22 genetic subtypes and 5 provisional entities of ALL (5)

Whole-genome sequencing (WGS) and RNA-sequencing are now increasingly used to identify genetic alterations of clinical importance in research settings. In particular, RNA-sequencing, which is more rapid and cost-effective than WGS, has enabled the detection of several new subtypes of pediatric ALL, characterized by specific gene fusions and distinct gene expression profiles (2, 3, 6). In addition, genetic alterations at the DNA-level, specific copy number alterations (CNVs) have been shown to infer prognostic information in ALL (7).

GENETIC METHODS IN CURRENT CLINICAL DIAGNOSTICS

Currently several different genetic methods are used in clinical diagnostics of ALL: chromosome banding analysis, which allows numerical and structural chromosomal alterations to be detected under the microscope; fluorescent in situ hybridization (FISH) enabling structural chromosomal alterations, e.g. balanced chromosomal translocations and gene fusions, to be indirectly visualized by fluorescent probes; DNA-arrays allowing CNVs to be identified, as well as various directed molecular genetic assays identifying gene fusions, single nucleotide variants (SNVs), and CNVs. Hence, more than four different methods are currently used in a diagnostic setting of ALL, with FISH necessitating several different probe sets to be used, a process mandating several different competences, instruments and collectively time consuming.

Whole-genome sequencing (WGS) and RNA-sequencing have now started to emerge in a clinical diagnostics setting at larger centers for various disease indications in a clinical diagnostic setting. Given their high resolution enabling the detection of most clinically relevant genetic alterations, the decreasing sequencing costs associated with such tests, and their amenability to automatization, these methods are likely to become more widely adopted, particularly in the diagnostic workup of ALL .

RNA-SEQUENCING AS AN EMERGING CLINICAL DIAGNOSTIC TOOL IN ALL

RNA-sequencing allows the detection and measurement of two key features of critical importance for ALL diagnostics: 1) chimeric gene fusions and 2) global gene expression signatures that can be used for subtype classification. Recent larger RNA-sequencing studies of pediatric ALL have identified several new subtypes, characterized by gene fusions, including *DUX4*-, *ZNF384*-, and *MEF2D*-rearranged subtypes, as well as a new group with an *ETV6/RUNX1*-like gene expression signature (2, 3 ,6, 8) (Table 1). Such studies have also identified a varying number of additional gene expression subtypes, which closely segregate with the underlying disease causing (driver) genetic alteration. In two studies, 14 and 18 gene expression subtypes, respectively, were identified in large series of pediatric ALL, each including >1000 cases (9, 10). Thus, by performing RNA-sequencing it is possible to use classifier models to predict which subtype a newly diagnosed case of ALL belongs to.

BIOINFORMATIC STRATEGIES TO ANALYZE RNA-SEQUENCING DATA

Multiple strategies have been described to analyze RNA-sequencing data (11, 12). In brief, the bioinformatic processing starts with a large number of *reads*, short snippets of genetic code read by the sequencing instrument and subsequently stored as text files. The reads are then *aligned* to a reference genome using different algorithms, and any reads that fall uniquely within the transcripts of a gene are counted as a measurement of the expression level of that gene. (11-13). A normalization step accounts for differences in total expression between samples. Notably, sequences from a gene fusion transcript will result in reads that tie distant parts of the reference genome together, which forms the basis for gene fusion detection (14, 15). Technical variation is minimized by using well-defined automatic workflows that are the same for both diagnostic samples and samples used for model training.

QLUCORE'S SOLUTIONS ENABLE INTUITIVE VISUALIZATION AND REPORTING OF RNA-SEQUENCING DATA

Qlucore's software solutions allow the user to import RNA-sequencing data that have been aligned and processed by different gene fusion detection algorithms (Figure 1), e.g. STAR-fusion (14) and FusionCatcher (15). Following import into Qlucore, the software enables the detection of two key features:
1) classification of the tested sample into known subtypes of ALL based on known gene expression subtypes
2) identification and visualization of gene fusions

Importantly, these two key features can support each other in rendering a correct clinical diagnosis.

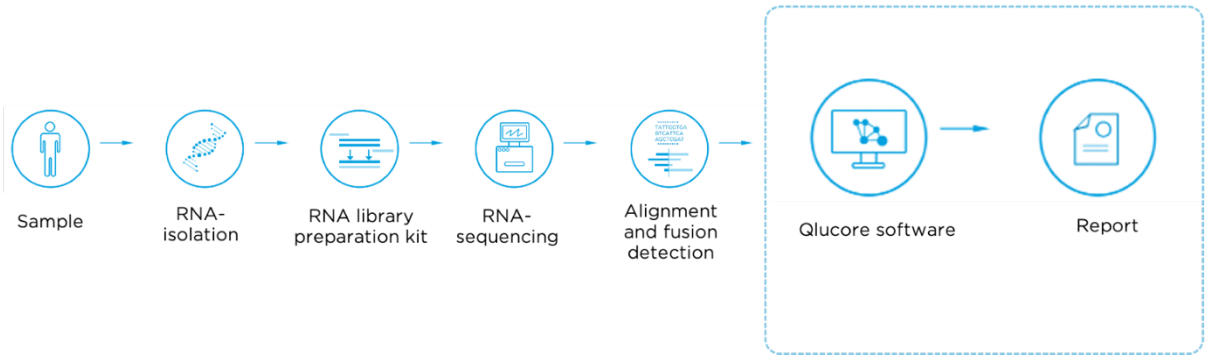

Figure 1

The ALL subtype of a diagnostic sample can be predicted from gene expression levels, using a highly tailored classifier built by Qlucore from a training data set of 212 pediatric ALL samples, combining machine learning with detailed clinical knowledge of the subgroups. The Qlucore Insights classifier has been validated using external, independent data sets and shown to have high sensitivity and specificity. Quality scores for both the classification and data quality in general, facilitates data interpretation and results.

To identify a gene fusion, the user obtains a sortable list that can be filtered in various ways. For example, the user can easily filter the identified gene fusions on the number of reads spanning the fusion breakpoint, if the gene fusion has been reported in available data bases (16, 17), and if it is predicted to be in-frame.

The identified gene fusions can also be visualized in a genomic context through so called circos plots (Figure 2) or in a genome browser where the fusion sequence can be studied and exported to allow additional analyses, e.g., to design primers for reverse transcriptase PCR to validate the presence of a rare fusion.

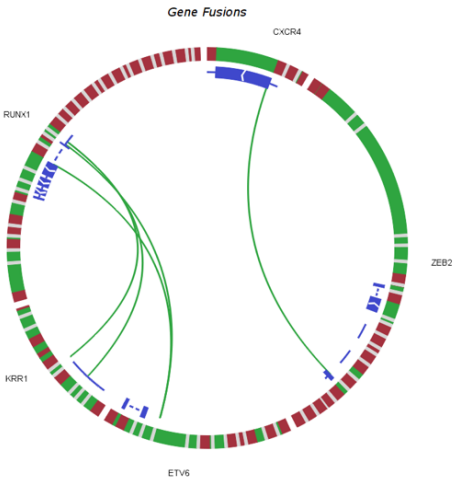

Figure 2

Currently, Qlucore Insights allow classification of a sample into one of the following clinically relevant subtypes of ALL:

1) High hyperdiploidy
2) *ETV6::RUNX1* or *ETV6::RUNX1*-like
3) *KMT2A*(*MLL*)-rearranged
4) *TCF3::PBX1*
5) *BCR::ABL1* or *BCR::ABL1*-like
6) *DUX4*-rearranged.

The classification is complemented with tailored visualizations to facilitate interpretation and communication (Figure 3).

Notably, because all but one of these gene expression subtypes of ALL also harbor gene fusions, sometimes multiple fusions as observed in *KMT2A*(*MLL*)-rearranged and *BCR/ABL1*-like cases, additional support from the gene fusion analysis can be obtained. Hence, by the gene fusion analysis it is, for example, possible to distinguish a *BCR/ABL1*-positive case from a *BCR/ABL*-like and to identify the fusion partner gene of *KMT2A*(*MLL*)-rearranged cases. The analysis of gene fusion can also be performed separately to arrive at a diagnosis in cases where the software does not provide the ability to classify a sample into a specific gene expression subtype.

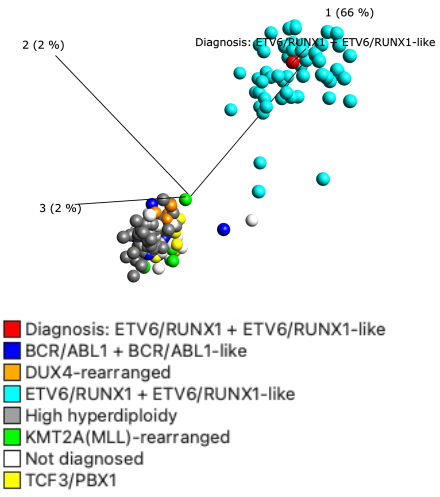

Figure 3

Following analysis in Qlucore Insights, a report is automatically generated summarizing the results. Several features are available to enter standardized texts, e.g., about the methods used and quality metrics obtained when generating the data. The user can then add a summary text to the report, if needed, and export the report as a pdf document.

Qlucore Insights is for research use and Qlucore Diagnostics is planned to be regulatory approved for clinical use. The target date for IVDR approval is Feb 2025.

CONCLUSIONS

ALL is characterized by several genetic alterations of great clinical diagnostic importance, as recognized by the current WHO and ICC classifications, and undoubtedly constitutes a role model for the current paradigm of precision diagnostics and precision medicine. The great majority of ALLs are characterized by gene fusions or display a distinct gene expression signature that can be used for classification purposes. A number of gene fusions in ALL can be targeted with specific drugs with improved outcomes for patients or used for risk stratification, making it critical to detect such alterations at the diagnostic workup. In this context, RNA-sequencing is becoming an increasingly important diagnostic tool as it can detect gene fusions in an unbiased manner and aid the subtype classification based on distinct gene expression signatures. Here, software meeting clinical diagnostic needs for robust and quality-controlled data analysis, visualization, and reporting will be instrumental for realizing the promises of RNA-sequencing based diagnostics of ALL and other malignant disorders

REFERENCES

1. Cancer Stat Facts: Leukemia — Acute Lymphocytic Leukemia (ALL): https://seer.cancer.gov/statfacts/html/alyl.html

2. Lilljebjörn H, Fioretos T. New oncogenic subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood*. 2017;130(12):1395-1401

3. Kimura S, Mullighan CG. Molecular markers in ALL: Clinical implications. *Best Pract Res Clin Haematol*. 2020;33(3):101193.

4. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*. 2022;36:1720-1748

5. Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood*. 2022;140(11):1200-1228

6. Roberts KG, Mullighan CG. The Biology of B-Progenitor Acute Lymphoblastic Leukemia. *Cold Spring Harb Perspect Med*. 2020;10(7):a034835

7. Hamadeh L, Enshaei A, Schwab C, et al. Validation of the United Kingdom copy-number alteration classifier in 3239 children with B-cell precursor ALL. *Blood Adv*. 2019;3(2):148-157

8. Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nat Commun*. 2016;7:11790

9. Li JF, Dai YT, Lilljebjörn H, et al. Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proc Natl Acad Sci U S A*. 2018;115(50):E11711-E11720

10. Gu Z, Churchman ML, Roberts KG, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet*. 2019;51(2):296-307.

11. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet*. 2019;20(11):631-656

12. Sahraeian SME, Mohiyuddin M, Sebra R, et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat Commun*. 2017;8(1):59

13. Baruzzo G, Hayer KE, Kim EJ, Di Camillo B, FitzGerald GA, Grant GR. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods*. 2017;14(2):135-139

14. Haas, BJ, Dobin A, Stransky N, et al. STAR-Fusion: fast and accurate fusion transcript detection from RNA-Seq. *BioRxiv.* 2017:120295

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

16. Hu X, Wang Q, Tang M, et al. TumorFusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res*. 2018;46(D1):D1144-D1149

17. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2024). Mitelman F, Johansson B and Mertens F (Eds.), https://mitelmandatabase.isb-cgc.org

18. Alaggio R, Gujral S, Lee Wood B, et al. **B-lymphoblastic leukaemia/lymphoma**. In: WHO Classification of Tumours Editorial Board. Haematolymphoid tumours [Internet]. Lyon (France): International Agency for Research on Cancer; 2024. (WHO classification of tumours series, 5th ed.; vol. 11). Available from: https://tumourclassification.iarc.who.int/chapters/63.

Table 1: Overview of common genetic abnormalities associated with B-ALL/LBL, provisional entities, methods allowing their detection and prognostic relevance. Content in table based on table 4.09 (18).

Genetic Entity	Frequency	Genetic Testing	Prognostic Relevance
Aneuploidy			
a. Hypodiploid (24-43 chromosomes) * **B-lymphoblastic leukaemia/lymphoma with hypodiploidy**	~2% of childhood are Near hypodiploid. ~1% of childhood, 10% of adult are Low hypodiploid.	Karyotyping/FISH	Unfavourable for Near and Low hypodiploid.
b. High hyperdiploidy) (>50 chromosomes) **B-lymphoblastic leukaemia/lymphoma with high hyperdiploidy**	Up to 30% of childhood	Karyotyping/FISH/SNP arrays	Favourable
*BCR::ABL1*rearranged * **B-lymphoblastic leukaemia/lymphoma with BCR::ABL1 fusion**	2-5% in children, 6% in adolescents and young adults and >25% in adults	Karyotyping/FISH/ PCR/RNA Sequencing for detection of *BCR::ABL1* fusion. Real-time PCR and/or ddPCR for monitoring treatment responses.	Unfavourable, but incorporation of TKI targeted therapy has improved outcomes. */KZF1*rearranged unfavourable
KMT2A rearranged * **B-lymphoblastic leukaemia/lymphoma with KMT2A rearrangement**	70-85% of infant, 2% of childhood and adult	Karyotyping/FISH/Global or targeted RNA sequencing	Unfavourable
*TCF3*rearranged * **B-lymphoblastic leukaemia/lymphoma with TCF3::PBX1 fusion, B-lymphoblastic leukaemia/lymphoma with TCF3::HLF fusion**	*TCF3::PBX1*in 4-6%	Karyotyping/FISH/RT-PCR/Global or targeted RNA sequencing for uncommon partners	*TCF3::PBX1*has better outcomes with contemporary therapeutic regimens. *TCF3::HLF* has a universally dismal prognosis
IGH*::IL3*rearranged **B-lymphoblastic leukaemia/lymphoma with IGH::IL3 fusion**	Rare	Karyotyping/FISH	Intermediate
IG*::MYC*rearranged	Rare in children, ~2% of adult	Karyotyping/FISH.	Unfavourable

ETV6::RUNX1 rearranged * **B-lymphoblastic leukaemia/lymphoma with ETV6:: RUNX1 fusion**	Nearly 25% of cases aged 2–10 years, <3% of adults	FISH/PCR/RNA Sequencing for detection of *ETV6::RUNX1*. Real-time PCR and/or ddPCR for monitoring responses.	Favourable
iAMP21 **B-lymphoblastic leukaemia/lymphoma with iAMP21**	1-2%	FISH for *ETV6::RUNX1* used to identify iAMP21.	Unfavourable; outcomes may improve with intensive chemotherapy
BCR::ABL1-like (Ph-like) * **B-lymphoblastic leukaemia/lymphoma with BCR::ABL1-like features**	Frequency increases with age; 10-15% children, 20-25% of adult	Whole transcriptome or targeted RNA sequencing is the most reliable method for diagnosis. When unavailable, alternative methods. e.g., FISH, PCR, and TaqMan low-density arrays.	Unfavourable, associated with high MRD positivity rates. Targeting the underlying genetic abnormality may improve survival rates.
ETV6::RUNX1-like * **B-lymphoblastic leukaemia/lymphoma with ETV6::RUNX1-like features**	3% of childhood	Whole transcriptome or targeted RNA sequencing. Increased frequency of *ETV6*-like B-ALL may be seen in patients with germline *ETV6* mutation.	Unfavourable
DUX4 rearranged *	4-7% of childhood, slightly higher in the AYA population	Difficult to identify using FISH and PCR due to telomeric localization and multiple DUX4-copies present in the genome. Whole transcriptome or targeted RNA sequencing are more reliable.	Favourable
MEF2D rearranged	4% of childhood and 10 % of adult	*MEF2D::BCL9* is difficult to detect using FISH. Whole transcriptome or targeted RNA sequencing is more reliable.	Unfavourable

ZNF384 rearranged	5% of childhood and 10% of adults B-ALL; 48% MPAL, B/Myeloid	FISH, whole transcriptome or targeted RNA sequencing	*EP300::ZNF384* favourable, *TCF3::ZNF384* unfavourable.
PAX5^{alt}/PAX5 p.P80R	*PAX5*^{alt} in 7% and *PAX5* P80R in 3%	PCR followed by sequencing, gene panels, FISH, Whole transcriptome or targeted RNA sequencing.	Intermediate for *PAX5* alt in children and unfavorable in adults. *PAX5* P80R may be unfavourable in children
NUTM1 rearranged	<2%	Whole transcriptome or targeted RNA sequencing	Favourable

*) gene expression-based classifiers for these subtypes are available in Qlucore's diagnostic solutions.