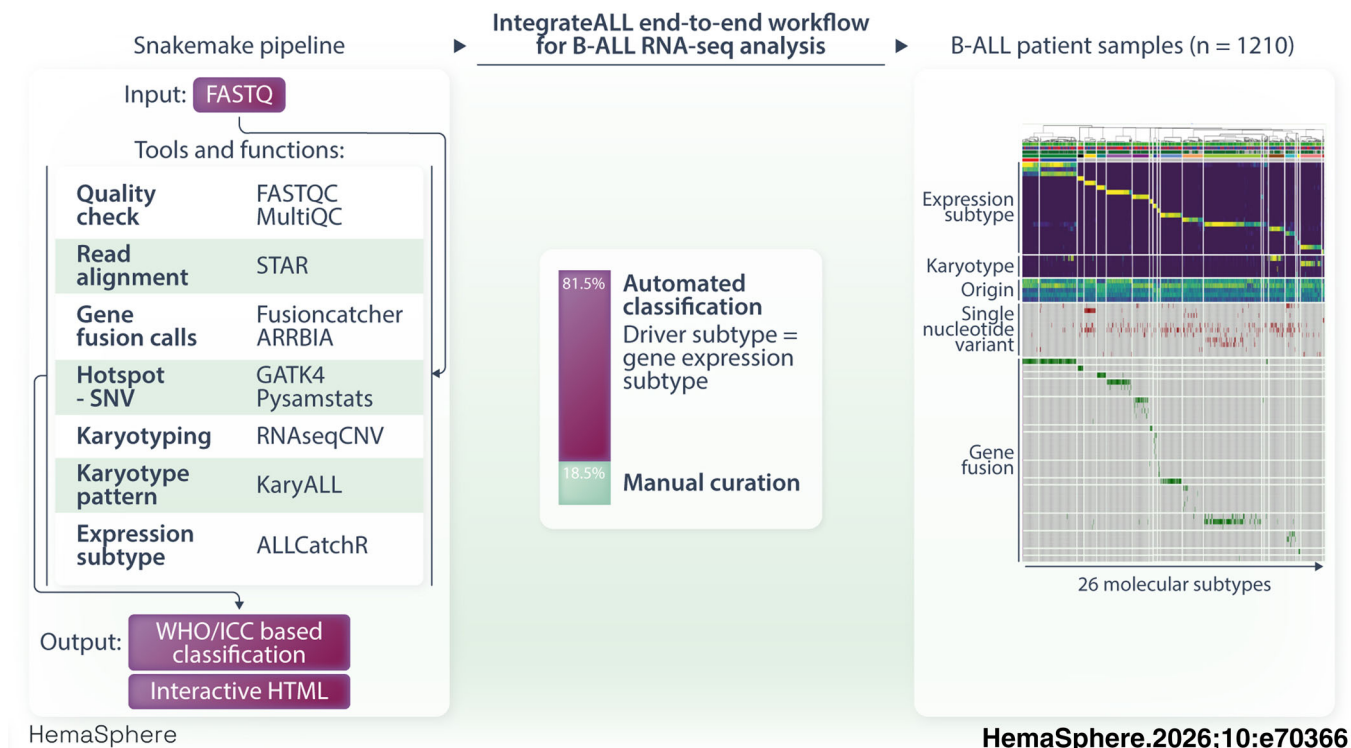



IntegrateALL: An end-to-end RNA-seq analysis pipeline for multilevel data extraction and interpretable subtype classification in B-precursor ALL

Nadine Wolgast^{1,2,3}  | Thomas Beder^{1,2,3}  | Mayukh Mondal^{3,4,5}  |
 Wencke Walter⁶  | Stephan Hutter⁶  | Sonja Bendig^{1,2,3}  |
 Jan Kässens^{1,2}  | Björn-Thore Hansen^{1,2}  | Katharina Iben^{1,2,3} |
 Sebastian Wolf⁷  | Anjali Cremer⁷  | Malwine Barz^{1,2,3}  |
 Martin Neumann^{1,2,3} | Nicola Göbke⁷  | Claudia Haferlach⁶  |
 Monika Brüggemann^{1,2,3}  | Claudia D. Baldus^{1,2,3}  | Alina M. Hartmann^{1,2,3,^}  |
 Lorenz Bastian^{1,2,3,^} 

Graphical Abstract



IntegrateALL: An end-to-end RNA-seq analysis pipeline for multilevel data extraction and interpretable subtype classification in B-precursor ALL

Nadine Wolgast^{1,2,3}  | Thomas Beder^{1,2,3}  | Mayukh Mondal^{3,4,5}  |
 Wencke Walter⁶  | Stephan Hutter⁶  | Sonja Bendig^{1,2,3}  |
 Jan Kässens^{1,2}  | Björn-Thore Hansen^{1,2}  | Katharina Iben^{1,2,3} |
 Sebastian Wolf⁷  | Anjali Cremer⁷  | Malwine Barz^{1,2,3}  |
 Martin Neumann^{1,2,3} | Nicola Göbke⁷  | Claudia Haferlach⁶  |
 Monika Brüggemann^{1,2,3}  | Claudia D. Baldus^{1,2,3}  | Alina M. Hartmann^{1,2,3,^}  |
 Lorenz Bastian^{1,2,3,^} 

Correspondence: Lorenz Bastian (lorenz.bastian@uksh.de)

Abstract

Transcriptome sequencing (RNA-seq) is emerging as a diagnostic standard for B-cell precursor acute lymphoblastic leukemia (B-ALL). Expression-based classifiers reach ~95% accuracy, but reproducible end-to-end solutions that also integrate transcript-derived genomic drivers and quantitative virtual karyotyping are lacking. We developed IntegrateALL, a Snakemake pipeline that standardizes RNA-seq analysis from FASTQ to rule-based subtype assignment across 26 WHO-HAEM5/ICC entities by integrating expression-based subtype prediction, gene fusion-/hotspot SNV calling, and virtual karyotyping. We introduce KaryALL, a machine learning classifier that uses normalized expression and minor-allele-frequency features (RNASeqCNV), to distinguish near-haploid, hypodiploid, and high-hyperdiploid B-ALL and chromosome-21 gains/iAMP21 (accuracy: 0.98/F1 score: 0.96 on 615 independent test samples). SNP-array concordance supported RNA-based karyotyping. Applied to 774 unselected B-ALL cases, IntegrateALL yielded unambiguous subtype assignments in 81.5%, based on concordance of gene expression class with a defining driver (75.3% of all cases) or, in selected cases, high-confidence expression-based classification alone (6.2%); the remainder (18.5%) were flagged for manual curation. Independent validation (three cohorts; $n = 436$, including pediatric cases) reproduced these distributions. Across all patients ($n = 1210$), 2.6% harbored two subtype-defining drivers, including hyperdiploidy in fusion-driven subtypes, where it was not expected, or subtype-defining SNVs (e.g., *PAX5* P80R/*IKZF1* N159Y) co-occurring with *BCR::ABL1*-positive/-like, *KMT2A*-, or *DUX4*-fusions. In most dual-driver cases, one subtype gene expression signature predominated, consistent with oncogenic hierarchies, but also with the possibility of technical artifacts, which should prompt individual orthogonal validations. IntegrateALL provides an adaptable fully reproducible workflow for molecular B-ALL characterization by systematically integrating genomic drivers and downstream gene regulation.

¹Medical Department II, Hematology and Oncology, University Medical Center Schleswig-Holstein, Campus Kiel, Kiel, Germany

²University Cancer Center Schleswig-Holstein, University Medical Center Schleswig-Holstein, Kiel, Germany

³Clinical Research Unit CATCH ALL (KFO 5010) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Kiel, Germany

⁴Institute for Clinical Molecular Biology, Christian Albrechts University Kiel, Kiel, Germany

⁵School of Biology, Indian Institute of Science Education and Research Thiruvananthapuram (IISER TVM), Thiruvananthapuram, Kerala, India

⁶MLL Munich Leukemia Laboratory, Munich, Germany

⁷Department of Medicine II, Hematology/Oncology, Goethe University Hospital, Frankfurt/M, Frankfurt/M, Germany

[^]Alina M. Hartmann and Lorenz Bastian share last authorship.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *HemaSphere* published by John Wiley & Sons Ltd on behalf of European Hematology Association.

INTRODUCTION

Diagnostic accuracy in hematology relies on systematic and reproducible workflows. The current World Health Organization Classification of Haematolymphoid Tumours (WHO-HAEM5)¹ and the International Consensus Classification (ICC)² of Myeloid Neoplasms and Acute Leukemias recognize 12 and 22 molecular subtypes of B precursor acute lymphoblastic leukemia (B-ALL), respectively, as distinct diagnostic entities. Additionally, the ICC defines five provisional subtypes. Diagnostic definitions are based on recurrent genomic aberrations such as specific aneuploidy patterns, gene fusions, single-nucleotide variants, and complementary gene expression signatures. These expression signatures typically reflect either the transcriptional regulation driven by subtype-specific genetic alterations or integrated regulatory networks downstream of multiple driver aberrations (e.g., *BCR::ABL1*-like ALL, *PAX5*alt ALL). Further subclusters with these subtypes reflect functional and clinical heterogeneity (e.g., multilineage vs. lymphoid *BCR::ABL1*-positive ALL and *JAK/STAT* vs. *ABL* class-driven *BCR::ABL1*-like ALL).

Existing subtype definitions, however, have mostly been characterized in specifically selected cohorts or age-specific contexts. Age-overriding systematic catalogues of driver aberrations and corresponding gene expression signatures remain lacking.

Transcriptome analysis by RNA-sequencing (RNA-Seq) allows comprehensive profiling of gene expression, together with inference of underlying genomic drivers. A variety of bioinformatic tools have been established for tasks such as gene expression count quantification,³ gene fusion calling,^{4,5} and identification of expressed single-nucleotide variants.^{6,7} We and others have developed approaches for systematic gene expression-based subtype classification.^{8–13} Accurate subtype classification requires integration of these multi-level data, accounting also for the rare possibility of multiple drivers in a single sample. Some existing classifiers^{11–13} already use outputs from multiple tools for interactive visualization and classification. However, comprehensive and transparent end-to-end workflows from raw FASTQ-files to rule-based classifications according to current diagnostic definitions are lacking, making subtype allocation resource-intensive and prone to errors during manual curation.

Certain molecular subtypes, such as near-haploid, low-hypodiploid, and hyperdiploid ALL, are defined by patterns of non-random chromosomal gains and losses. They represent a particular challenge for RNA-based classification due to the absence of a single defining driver lesion and highly similar gene expression patterns between near-haploid and hyperdiploid ALL.⁸ RNASeqCNV¹⁴ addresses this challenge by reconstructing virtual karyotypes from normalized gene expression count data and expressed variant allele frequencies. However, the definition of these aneuploidy-based subtypes currently relies on frequency distributions rather than clear quantitative criteria, and methods to objectively assess a sample's karyotype against these distributions are not available. Thus, karyotype interpretation remains, to some extent, subjective, limiting reproducibility.

To address these gaps and enable systematic and reproducible subtype allocation, we developed IntegrateALL, a free open-source Snakemake¹⁵ pipeline. IntegrateALL provides a standardized workflow from raw RNA-Seq FASTQ files through quality assessment, gene expression quantification, identification of gene fusions and hotspot SNVs, and reconstruction of virtual karyotypes. Classification according to current diagnostic criteria is implemented via a novel, rule-based system informed by our own^{8,16,17} and published data^{18–20} on driver aberrations and corresponding gene regulation. For an automated quantitative analysis of B-ALL karyotype patterns, we developed and implemented KaryALL, the first machine learning

classifier for RNA-Seq-based virtual karyotypes. Easy accessibility of all data layers in an interactive HTML output ensures transparency for the classification process and offers a user-friendly environment for manual curation of complex cases that do not fit standard diagnostic definitions. Validation on 1210 B-ALL samples from four cohorts across age groups (Supporting Information S2: Table 1) confirmed the scalability and robustness of our pipeline. Its open design ensures adaptability to evolving diagnostic criteria and systematic analytical workflows in B-ALL.

MATERIALS AND METHODS

To establish IntegrateALL as a systematic B-ALL RNA-Seq analysis workflow, we used Snakemake,¹⁵ a Python-based workflow management system, to integrate raw data quality check (FASTQC²¹ and MULTIQC²²), read alignment³ to GRCh38.83, gene fusion calling (ARRIBA⁵ and FusionCatcher⁴), and raw single-nucleotide variant calling (GATK, Genome Analysis Toolkit).⁷ High-confidence hotspot SNVs were filtered with pysamstats⁶ against a curated list of B-ALL relevant codon changes (*ZEB2*, *KRAS*, *NRAS*, *FLT3*, *CRLF2*, *IDH1*, *IDH2*, *JAK1*, *JAK2*, *TP53*, *PAX5*, and *IKZF1*; Supporting Information S2: Table 2). RNASeqCNV¹⁴ was used to infer virtual karyotypes from normalized counts and expressed variant allelic frequencies derived from GATK calls.⁷ The GATK workflow, executed with Snakemake wrappers, adheres to best practices for RNA-seq short variant discovery, including read group addition, duplicate marking, exon segmentation, base recalibration, and variant filtration. For a systematic classification of aneuploid B-ALL subtypes, we developed KaryALL, which uses chromosome-level expression features and minor-allele-frequency distributions from RNA-Seq CNV outputs. Models were built in Python using scikit-learn²³ (data preprocessing, RandomForestClassifier, and KNeighborsClassifier), imblearn²⁴ for class balancing via SMOTE,²⁵ and the XGBClassifier from the xgboost package.²⁶ For systematic gene expression-based subtype allocation, we included ALLCatchR,⁸ our machine learning-based classifier, which provides automated allocation to 21 B-ALL subtypes and reports proximity to normal lymphopoiesis, inferred immunophenotype, sex, and blast proportion. Outputs from individual tools were integrated using a newly established rule set reflecting genomic and transcriptomic definitions consistent with current diagnostic definitions.^{1,2}

We tested IntegrateALL on $n = 774$ B-ALL first diagnosis bone marrow/peripheral blood samples from adult patients, extending our previously published cohort of patients treated on protocols of the German Multicenter Study Group for Adult ALL (GMALL).¹⁷ Patients consented to the scientific use of diagnostic leftover materials (Kiel University Ethics committee vote D416/21). RNA-Seq was performed as described previously.^{16,17} Briefly, libraries were prepped using TruSeq or Illumina Stranded mRNA Prep (Illumina, San Diego, USA), and sequenced, aiming for 30 Mio paired-end reads (75–150 bp) on a NovaSeq or NextSeq sequencing system (Illumina, San Diego, USA). For orthogonal genomic karyotyping, DNA from diagnostic materials was used for SNParray profiling (Infinium Global Screening Array-24 v3.0, Illumina, San Diego, USA). Array raw data were processed in Genome Studio (Illumina, San Diego; USA, v2.0.5) to derive allelic depth and variants. For further validation of KaryALL and IntegrateALL, we used real-world diagnostic RNA-Seq data with a genomically validated ground truth²⁷ and published RNA-Seq data^{28–30} with a specific focus on aneuploid subtypes that were identified by agreement between ALLCatchR and ALLSorts.⁹ IntegrateALL is available as a free open-source pipeline under the MIT License (<https://github.com/NadineWolgast/IntegrateALL>) and includes a Dockerfile for containerized and reproducible deployment.

RESULTS

Development of a computational framework for systematic subtype allocation in B-ALL

RNA-Seq data contain multiple data levels, which are informative for molecular subtype classification. In B-ALL, subtypes are defined by consistency between gene expression signatures and genomic drivers inferred from RNA-Seq profiles. To systematically extract these features in an accessible manner, we developed IntegrateALL, a Snake-make pipeline (Figure 1), which starts from RNA-Seq FASTQ files to perform the following analyses: read quality assessment, gene expression-based subtype allocation, gene fusion calling, identification of hotspot single-nucleotide variants, and analysis of virtual karyotypes. Data are made accessible via an interactive HTML report and are used for integrative subtype classification using a rule set based on current diagnostic definitions.^{1,2}

KaryALL: A machine learning classifier for aneuploid B-ALL subtypes

Non-random chromosomal gains and losses define near-haploid (23–29 chromosomes), low-hypodiploid (33–39 chromosomes), and high-hyperdiploid (52–67 chromosomes)³¹ ALL. Near-haploid and high-hyperdiploid subtypes share highly similar gene expression patterns, representing a challenge to gene expression-based subtype classification. Intrachromosomal amplification of chromosome 21 (iAMP21) is a structural aberration sufficiently large to be detectable by virtual karyotyping. It defines a molecular B-ALL subtype with a gene expression signature highly similar to that of BCR::ABL1-like ALL. To facilitate systematic classification of these four molecular subtypes, we used RNASeqCNV¹⁴ outputs of chromosome-specific gene expression counts and minor allele frequency distributions (Figure 2A) to train and validate a machine learning classifier.

We extracted 186 features comprising chromosome-level CNV metrics (peak positions, median weighted log2fold changes, and range boundaries) and 24 discriminative chromosome 21 genomic positions (chr21:25,772,460-39,428,528) for enhanced iAMP21 detection. For training, we used an aggregated RNA-Seq data set of 395 samples from seven cohorts classified according to published ground truth and ALLCatchR⁸/ALLSorts⁹ predictions as near haploid ($n = 6$), low hypodiploid ($n = 28$), hyperdiploid ($n = 75$), iAMP21 ($n = 7$), or non-aneuploid (“other”; $n = 279$; Supporting Information S1: Figure 1). For comparison with genomic karyotypes, 33 samples ($n = 11$ hyperdiploid, $n = 12$ low-hypodiploid, and $n = 10$ diploid karyotypes) were subjected to SNParray profiling. RNASeqCNV calls were concordant in $n = 719/759$ chromosomes analyzed (Figure 2B; $P = 5.00E-4$), supporting the applicability of RNASeqCNV for virtual karyotyping. Unsupervised UMAP analysis separated most subtypes (Figure 2C), with the least distinct clustering for iAMP21. To improve iAMP21 performance, we defined 29 additional chromosome 21-specific count features and minor allele frequency values for incorporation into training. Analysis of rare non-iAMP21 cases with isolated chromosome 21 amplification and otherwise diploid karyotypes indicated that these could not be distinguished in RNASeqCNV outputs from iAMP21. Therefore, we considered these as a combined class: iAMP21/chr21-amplification. Features were standardized and class imbalance was addressed using SMOTE (synthetic minority over-sampling technique) during training. We established an ensemble soft-voting classifier from the three best-performing machine learning models (random forest, K-nearest neighbors (KNN), and XGBoost; Supporting Information S1: Figure 2). Hyperparameter optimization was conducted using a randomized search with fivefold stratified cross-validation. In leave-one-out cross-validation, individual models achieved accuracies of 0.95–0.97 and F1 scores of 0.82–0.93 (Supporting Information S1: Figure 3). Hyperparameter tuning was conducted to optimize performance, particularly for under-represented subtypes such as iAMP21 and near-haploid cases. The ensemble classifier achieved an accuracy of 0.98 [95% CI: 0.96–0.99]

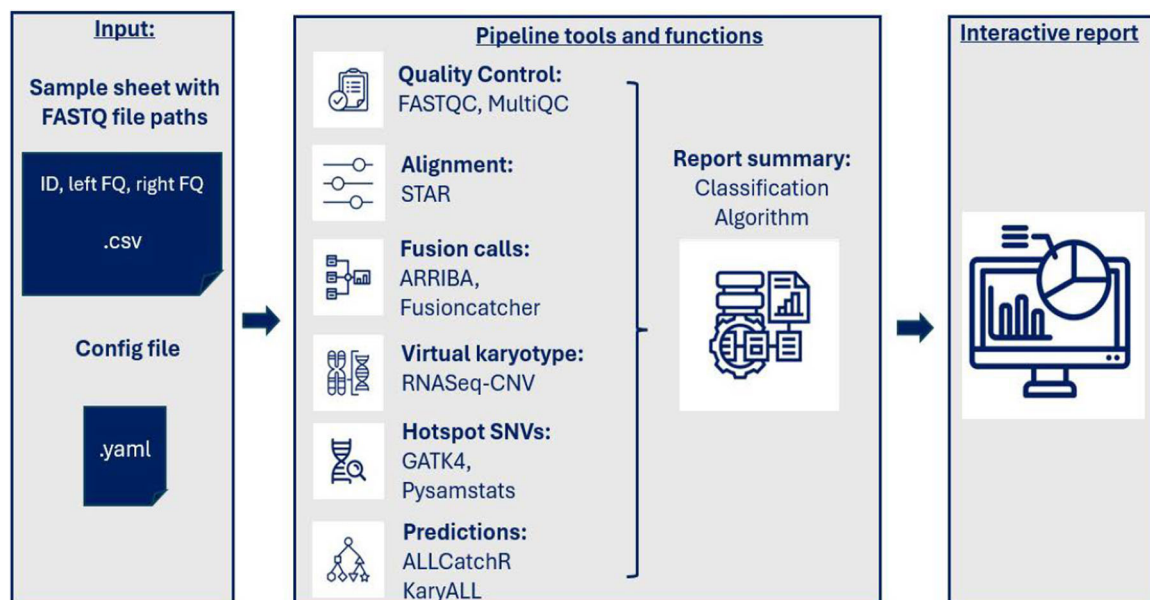


FIGURE 1 Schematic overview of the IntegrateALL pipeline. The workflow starts with an input sample sheet and raw RNA-seq FASTQ files. Subsequent steps integrate quality control (FASTQC,²¹ MULTIQC²²), read alignment (STAR³), gene fusion detection (ARRIBA,⁵ FusionCatcher⁴), SNV calling (GATK,⁷ pysamstats⁶), virtual karyotyping (KaryALL classification based on RNASeqCNV¹⁴ outputs), and gene expression-based subtype classification (ALLCatchR⁸). A configuration file enables resource optimization and parameter customization. An interactive summary output integrates results from all tools for manual review and displays the final classification (WHO-HAEM5¹/ICC²).

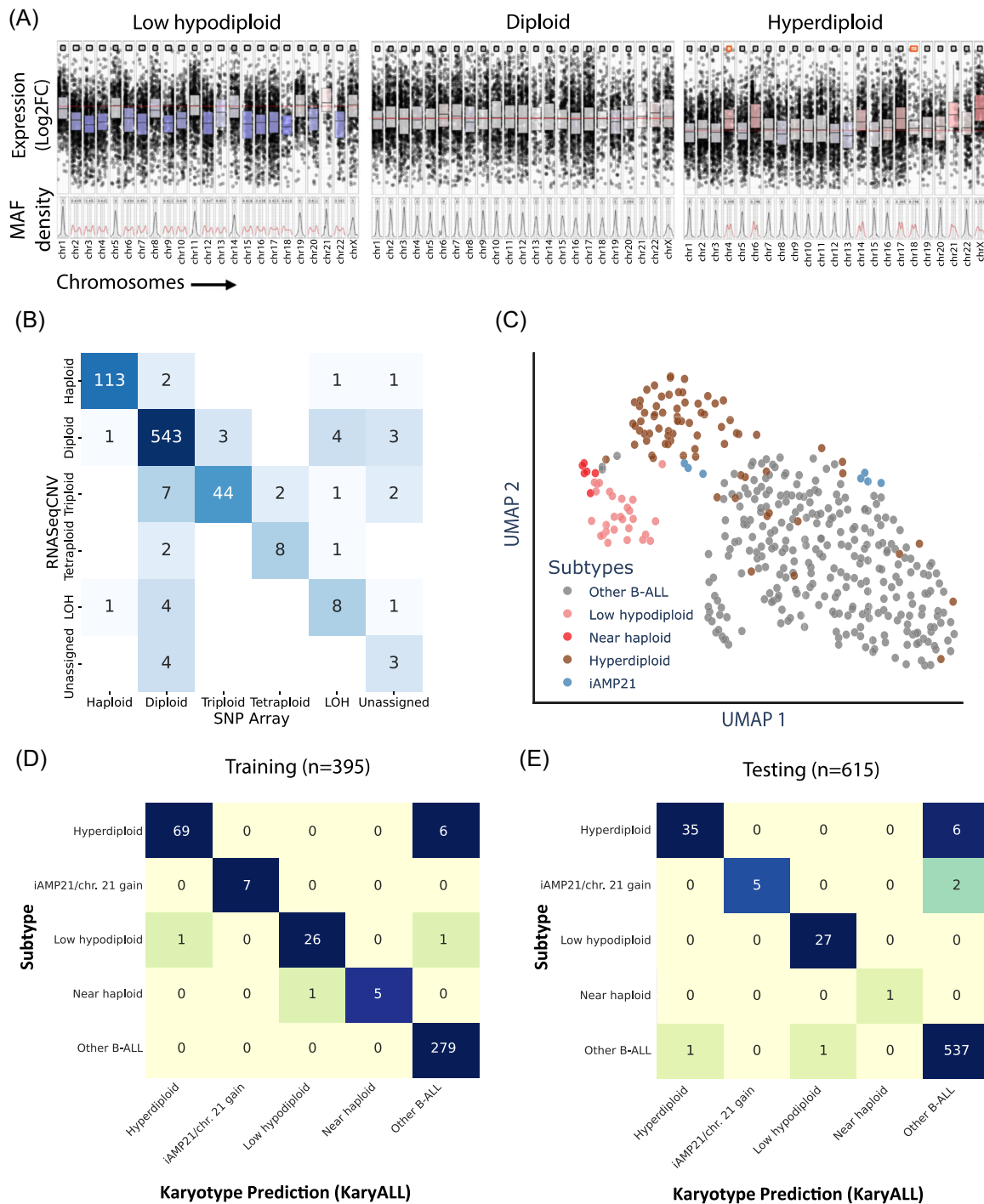


FIGURE 2 Training and validation of the KaryALL virtual karyotype classifier. (A) RNASeqCNV provides virtual karyotypes from RNA-seq data leveraging normalized counts and variant allele frequency (MAF) densities. Plots illustrate representative low-hypodiploid, diploid, and high-hyperdiploid cases. (B) Ploidy states and loss of heterozygosity (LOH) of individual chromosomes ($n = 33$ samples) were determined by manual curation of RNASeqCNV outputs and compared to SNParray outputs. (C) B-ALL samples ($n = 395$) with available SNParray and RNA-seq data were analyzed by uniform manifold approximation and projection for dimension reduction (UMAP) using RNASeqCNV-normalized expression and MAF density data. Colors indicate molecular subtypes according to genomic ground truth and manual curation of RNASeqCNV/SNParray outputs. This data set was used for training the machine learning classifier KaryALL to predict B-ALL karyotypes based on RNASeqCNV outputs. The corresponding comparison of RNASeqCNV and SNParray profiles and details on the KaryALL design are shown in Supporting Information S1: Figures 1–3. (D) Confusion matrix of KaryALL classification accuracy on the training data set (accuracy: 0.98/F1 score: 0.96). (E) Confusion matrix of KaryALL classification accuracy on test data from three independent cohorts (accuracy: 0.98/F1 score: 0.98).

and an F1 score of 0.96 using leave-one-out cross validation (Figure 2D). The average specificity and sensitivity for detection of aneuploid karyotypes or iAMP21/chr 21 amplification were 0.99 [95% CI: 0.97–1.00] and 0.92 [95% CI: 0.88–0.95], respectively. The

lowest performance was observed for near-haploid cases (sensitivity 1.00, specificity 0.83; $n = 6$), while the highest accuracy was achieved for iAMP21 chr 21 amplification (sensitivity and specificity both 1.00; $n = 7$). Robustness was confirmed in two independent cohorts

($n = 105$ Munich Leukemia Laboratory; $n = 206$ pediatric B-ALL samples²⁸) and our own cohort ($n = 304$), totaling 615 samples (Figure 2E). The overall weighted F1 score was 0.98 [95% CI: 0.97–0.99] and the accuracy was 0.98 as well, demonstrating the model's robustness and its applicability to broader data sets.

To our knowledge, KaryALL is the first systematic classifier for virtual B-ALL karyotypes and has therefore been implemented to complement gene expression-based subtype allocation with the corresponding genomic driver profile. Detailed training procedures, hyperparameters, and code are provided at <https://github.com/NadineWolgast/KaryALL>.

Classification for 26 molecular B-ALL molecular subtype definitions

Current hematologic disease classifications define 12¹ and 22² B-ALL molecular subtypes as well as five provisional entities. We formalized B-ALL subtype allocation within our pipeline by establishing a rule set for systematic integration of (i) our gene expression-based reference for 21 B-ALL subtypes,^{8,32} (ii) an expanded catalogue of 165 driver gene fusions driver candidates based on our own¹⁶ and published data sets,^{16,19,20} (iii) KaryALL karyotype classifications, and (iv) subtype-defining single-nucleotide variants (e.g., *PAX5* p.P80R or *IKZF1* p.N159Y; Figure 3, Supporting Information S1: Figure 4; Supporting Information S2: Table 3).

For an unambiguous subtype allocation, we require concordance between a subtype-defining genomic driver and the corresponding ALLCatchR gene expression subtype and absence of any other subtype-defining driver aberration (Figure 3). For such cases, any ALLCatchR confidence level was accepted. Because original definitions included gene-expression-only allocations for certain subtypes (e.g., *BCR::ABL1*-like NOS¹⁸ and *PAX5alt*¹⁹) and because gene fusions involving the IGH locus pose a specific challenge for identification by RNA-Seq, we allowed high-confidence expression-only calls for IGH-fusion driven subtypes (i.e., *DUX4* and *CEBP*) or *BCR::ABL1*-like and *PAX5alt*, provided that no alternative driver was present. All samples fitting these definitions were automatically classified according to $n = 10/11$ WHO-HAEM5 and $n = 26/27$ ICC molecular subtypes; *IGH::IL3* rearranged B-ALL was not represented due to its rarity and lack of gene expression data. The remaining cases were flagged for manual curation. An overall open rule set design permits flexible adaptation to incorporate novel subtype definitions.

IntegrateALL yields concordant expression and driver/karyotype calls in 80% of B-ALL RNA-Seq profiles

We applied IntegrateALL to $n = 774$ unselected adult B-ALL cases, comprising samples from our previously published dataset¹⁶ and additional cases from ongoing RNA-Seq-based routine diagnostics (Figure 4). Of all patients, 43.9% were female and 56.1% were male, with 14.3% and 85.7% pro-B versus pre-B/common inferred ALL immunophenotype, and with an overall median blast proportion of 72.8% according to ALLCatchR predictions. ALLCatchR provided high-confidence gene expression subtypes in $n = 560/774$ cases (72.3%) and candidate subtypes in $n = 181/774$ cases (23.4%); $n = 33/774$ patients (4.3%) remained unclassified by expression alone. KaryALL identified near-haploid ($n = 2$), low-hypodiploid ($n = 42$), high-hyperdiploid ($n = 46$), and iAMP21 ($n = 1$) karyotypes, confirmed by manual curation of RNASeqCNV outputs. Subtype-defining driver fusions were detected in 476 samples (54 unique drivers; Supporting Information S2: Table 4; Supporting Information S1: Figure 5).

Recurrent cooperating hotspot events affected 46 positions in 12 genes, including *NRAS* ($n = 121/774$; 15.6%), *KRAS* ($n = 70/774$; 9.0%), and *FLT3* hotspot SNVs ($n = 40/774$; 5.2%; Supporting Information S2: Table 2).

Application of our defined rule set (Figure 3) to our B-ALL cohort provided automated subtype allocations in $n = 631/774$ (81.5%) of cases (Figure 5A) based on concordance between gene expression subtype and gene fusions ($n = 481/631$; 76.2%; Figure 5A), aneuploid karyotypes ($n = 68/631$; 10.8%), hotspot SNVs ($n = 34/631$; 5.4%), or high-confidence gene expression alone ($n = 48/631$; 7.6%). These proportions varied between molecular subtypes (Figure 5A).

IntegrateALL identifies novel driver constellations in samples flagged for manual curation

Samples with ambiguous or incomplete matches to the predefined subtype rule set ($n = 143/774$; 18.5%) were flagged for manual curation (Figure 5B). Of these, 28 (19.6%) and 89 (62.2%) had high-confidence or candidate-confidence ALLCatchR subtype allocations, respectively; 26 (18.2%) were unclassified by ALLCatchR and remained so after review. Most flagged samples ($n = 88/143$; 61.5%) lacked a confirming genomic driver call despite an ALLCatchR prediction. This mainly involved subtypes whose original definitions allowed expression-only cases (i.e., *BCR::ABL1*-like ($n = 32$), *PAX5alt* ($n = 16$)), aneuploid subtypes where KaryALL did not recover a matching karyotype (i.e., hyperdiploid ($n = 12$) and low hypodiploid ($n = 12$)), or other molecular subtypes ($n = 16$). Here, IntegrateALL identifies cases that would either benefit from repeat sampling with higher blast infiltration and/or orthogonal genomic validation.

Discordance between gene expression subtype and driver calls occurred in 13/143 (9.1%) cases. These included samples with candidate or high-confidence gene expression subtypes in which drivers from a different subtype were present while the expected driver was absent ($n = 7$; e.g., *PAX5* P80R mutations in cases with *PAX5alt*, *CEBP*, or *BCR::ABL1* gene expression signatures). One *BCR::ABL1*-like gene expression case without driver fusion was resolved as *BCR::ABL1*-positive after identification of the gene fusion. KaryALL corrected two ALLCatchR hyperdiploid predictions by identifying near-haploid karyotypes, illustrating its value in separation of these subtypes that share similar gene expression signatures. Conversely, three low-hypodiploid cases by expression and manual karyotype review were misclassified as hyperdiploid or iAMP21 by KaryALL.

Importantly, in the remaining 16/143 (11.2%) cases (2.6% of the entire cohort), IntegrateALL identified secondary genomic drivers in addition to the subtype-defining alteration (Figure 6A,B, Supporting Information S2: Table 5). Examples included hyperdiploidy co-occurring with *KMT2A* fusions and *PAX5* P80R or *BCR::ABL1*-like gene fusions in *ZNF384*, *PAX5* P80R, and hyperdiploid cases. Two *BCR::ABL1*-like cases harbored driver fusions from different signaling trajectories (Figure 6B).

Notably, ALLCatchR predicted a dominant subtype in nearly all dual-driver samples. UMAP analysis based on 2802 LASSO-selected, subtype-defining genes⁸ grouped these cases together with one of the ALLCatchR subtypes (Figure 6A). To assess how closely individual double-driver samples align with their annotated subtypes, we used the 600 top variable expressed genes from all samples of the two involved subtypes for a subtype-centered principal component analysis (PCA) that compared the position of the double-driver sample in PCA space compared to the centroid (mean PC1/PC2 coordinates) of each subtype group. The Euclidean distance to each centroid was calculated to quantify subtype proximity (Figure 6B). This nearest-neighbor analysis and unsupervised clustering of sample-level

ALLCatchR	Confidence	Gene Fusion Call	Karyotype Classifier	SNV-Hotspot Call	ICC: B-ALL ...	WHO-HAEM5: B-lymphoblastic leukaemia/lymphoma with ...
Ph-pos, lymphoid	any	BCR :: ABL1	hyper-diploid	#	with t(9;22)(q34.1;q11.2) / BCR::ABL1 with lymphoid only involvement	BCR::ABL1 fusion
Ph-pos, multilineage					with t(9;22)(q34.1;q11.2) / BCR::ABL1 with multilineage involvement	
Ph-like	any	NUP214, ETV6, ZMIZ1, RANBP2, RCSD1, SNX2, GENPC, LSM14A, EXOSC2, MYO18B, NUP153, SFPQ, SNX1, FOXF1 :: ABL1	other	#	BCR::ABL1-like, ABL-1 class rearranged	BCR::ABL1-like features
		RCSD1, ZC3HAV1, PAG1 :: ABL2				
		SSBP2, MEF2D, TBL1XR1 :: CSF1R				
		GATAD2A, NCOR1 :: LYN				
		FIP1L1 :: PDGFRA				
		EBF1, TNIP1, ATF7IP, ETV6, ZEB2, ZMYND8, SNX29, SSBP2 :: PDGFRB				
		IGH, P2RY8 :: CRLF2				
		IGH, IGK, LAIR1, THADA :: EPOR				
		MYH9 :: IL2RB				
		IGH :: IL7R				
		PAX5, BCR, SSBP2, ATF7IP, ETV6, PCM1, PPF1BP1, TERF2, TPR, EBF1, OFD1, STRN3, USP25, ZNF274, RFX3, SMU1, WDR37, ZNF340 :: JAK2				
		MYB, ZNF340, SMARCA1 :: TYK2				
		KANK1 :: CBL				
		ZFAND3 :: DGKH				
		ZEB2 :: IKZF1				
ETV6 :: NTRK3						
TMEM2 :: PTK2B						
GOPC :: ROS1						
ETV6::RUNX1	any	ETV6 :: RUNX1			with t(12;21)(p13.2;q22.1)ETV6::RUNX1	ETV6::RUNX1 fusion
ETV6::RUNX1-like	any	AMPH, C7ORF72, CASC15, CD163, EBF1, ELMO1, ERC1, EXTL1, FAM136A, FOXO3, IKZF1, QSOX1, SLC30A7, SRRM1, TMTC1 :: ETV6			ETV6::RUNX1-like (provisional entity)	ETV6::RUNX1-like features
	high	RNF22, STIM2, ZBPB :: IKZF1				
Hyperdiploid	any	#	hyper-diploid		hyperdiploid	high hyperdiploidy
Low hypodiploid	any	#	low hypodiploid		low hypodiploid	Hypodiploidy
iAMP21	any	#		iAMP21	with iAMP21	iAMP21
KMT2A	any	KMT2A :: AFF1, DCPS, EPS15, IKZF1, MLLT1, MLLT10, MLLT3, TNS3, UBASH3B, USP2			with t(v;11q23.3) / KMT2A rearranged	KMT2A rearrangement
BCL2/MYC	any	IGH :: BCL2			with MYC rearrangement	
	high	BCL2 :: MYC				
CDX2/UBTF	any	UBTF :: ATXN7L3			with UBTF::ATXN7L3/PAN3, CDX2 (CDX2/UBTF)	
CEBP	high	#		ZEB2 H1038R	with mutated ZEB2 (p.H1038R) / IGH::CEBPE (provisional entity)	
IKZF1 N159Y	any	#		IKZF1 N159Y	with mutated IKZF1 N159Y	
DUX4	any	ETV6, IGH, PCGF5 :: DUX4			with DUX4 rearrangement	
KMT2A	any	MED12 :: HOXA9			KMT2A rearranged-like (provisional entity)	
	high	#				
MEF2D	any	MEF2D :: BCL9, DAZAP1, FOXJ2, HNRNPM, HNRNPUL1, PYGO2, SS18			with MEF2D rearrangement	other defined genetic abnormalities
Near haploid	any	#	near haploid		near haploid	
NUTM1	any	ACIN1, BRD9, CUX1, IKZF1, SLC12A6, ZNF618 :: NUTM1			with NUTM1 rearrangement	
PAX5 P80R	any	#		PAX5 P80R	with mutated PAX5 P80R	
PAX5alt	any	BMP2K :: LEF1	other	#	with PAX5 alteration (provisional entity)	
		IGH :: ID4				
		MBNL1 :: MEIS2, PAX5				
high	#					
ZNF384	any	SMARCA2, ARID1B, CLTC, CREBBP, DDX42, EP300, EWSR1, NIPBL, SMARCA2, TAF15, TCF3 :: ZNF384			with ZNF384(362) rearrangement	
	high	#			ZNF384 rearranged-like (provisional entity)	
HLF	any	TCF3, TCF4 :: HLF			with HLF rearrangement	TCF3::HLF fusion
TCF3::PBX1	any	TCF3 :: PBX1			with t(1;19)(q23.3;p13.3) / TCF3::PBX1	TCF3::PBX1 fusion

FIGURE 3 B-ALL subtype definitions. Molecular B-ALL subtype definitions used for subtype classification in IntegrateALL. This overview represents an integration of gene expression-based subtype (ALLCatchR), gene fusion calls, virtual karyotype classification (KaryALL), and recurrent hotspot single-nucleotide variants. The entire rule set and the classification logic are presented in Supporting Information S2: Table 3 and Supporting Information S1: Figure 4. Unambiguous matches to this rule set are automatically classified according to current classifications (ICC/WHO-HAEM5). “#” indicates classes where subtype-defining genomic aberrations are absent.

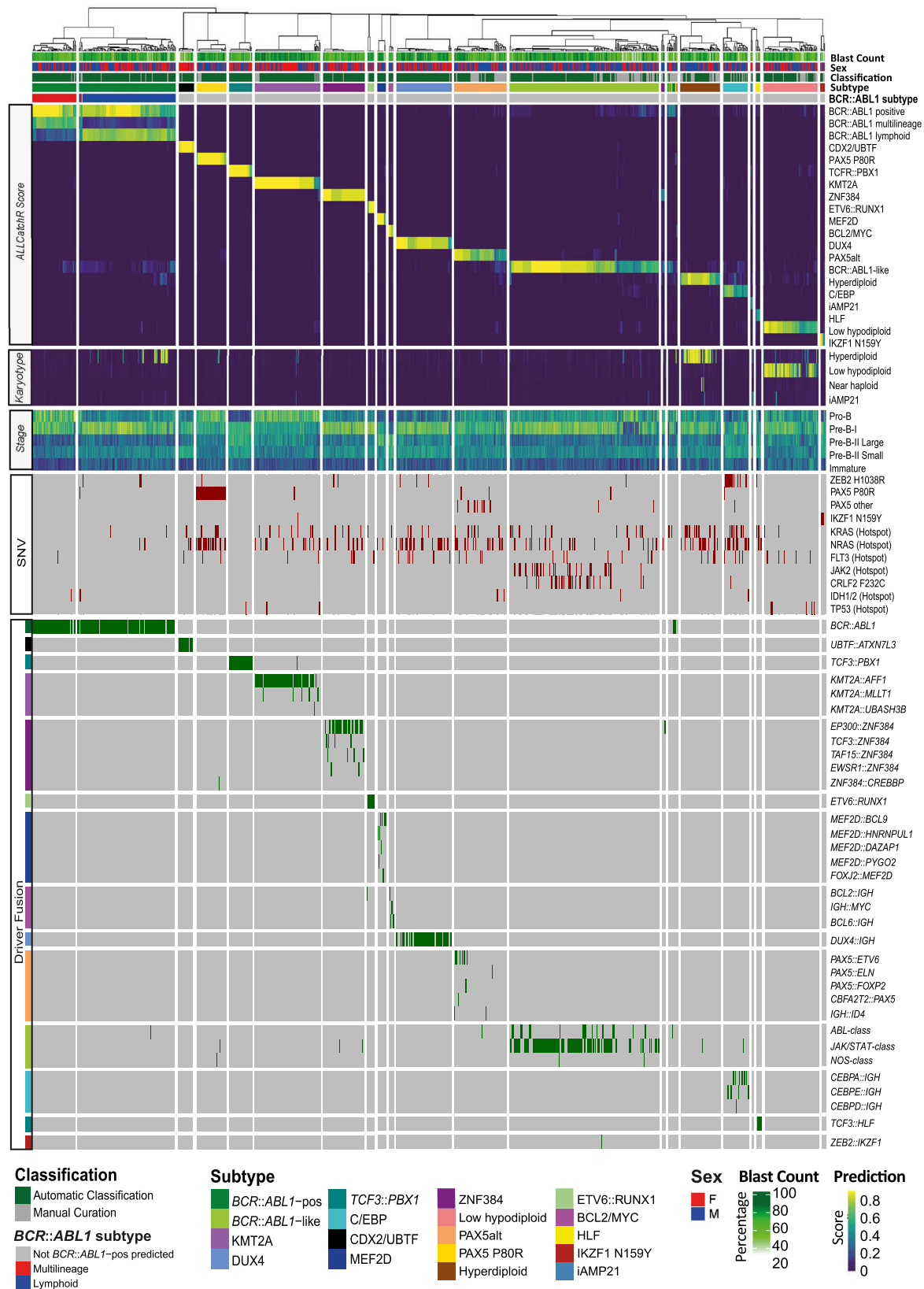


FIGURE 4 IntegrateALL classification of unselected adult B-ALL RNA-seq samples. A total of 774 B-ALL samples from our own sequencing were processed through IntegrateALL, yielding an overview of gene expression-based subtype prediction (ALLCatcher), virtual karyotype classification (KaryALL), proximity to normal B cell development (ALLCatcher), hotspot single-nucleotide variants (GATK workflow, Pysamstats), and driver fusion calls (FusionCatcher, ARRIBA), together with patients' sex and blast count predictions (ALLCatcher) and the IntegrateALL annotation for automated classification versus manual curation.

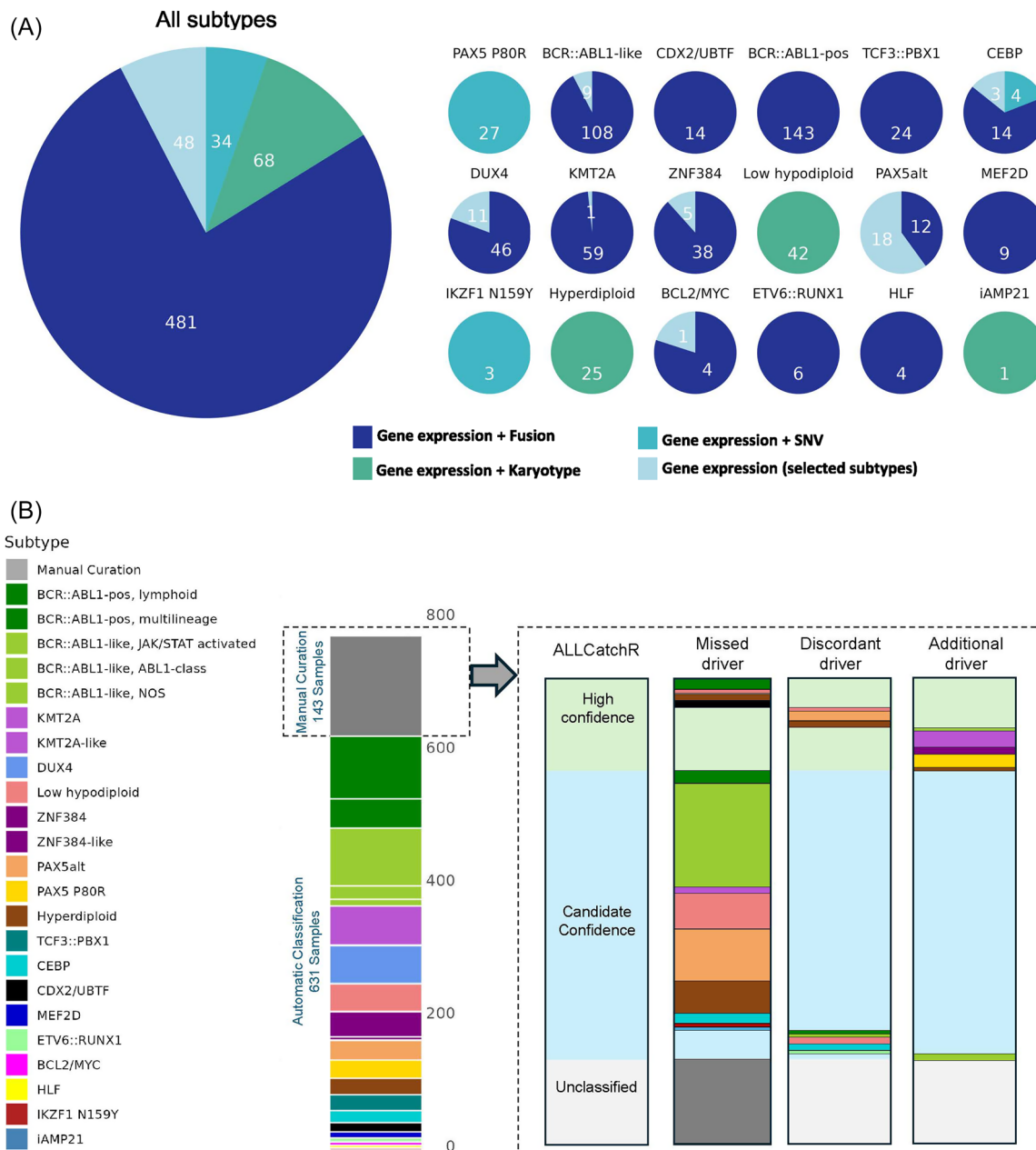


FIGURE 5 Overview of classification confidence and manual curation. **(A)** Summary of the basis for classification in cases receiving an unambiguous automatic IntegrateALL subtype assignment for the entire cohort ($n = 631/774$; 81.5%) and individual subtypes. **(B)** The remaining 143/774 cases (18.5%) were flagged for manual curation. The ALLCatchR confidence categories for gene expression-based subtype classification are shown together with the results of manual curation, which identified three categories: (i) *Missed driver*, with “candidate” or “unclassified” gene expression-based predictions lacking expected driver events; (ii) *Discordant driver*, representing cases with “candidate” or “high-confidence” gene expression-based predictions and driver calls from other subtypes; and (iii) *Additional driver*, where an additional genomic driver from a distinct molecular subtype was identified alongside a subtype definition that would otherwise meet the requirements of unambiguous automatic classification. Validations on three independent cohorts ($n = 436$) are shown in Supporting Information S1: Figure 7.

expression profiles (Figure 6C) using the top 600 variable expressed genes both confirmed dominance of a single driver program, consistent with either a dominant role of one of the drivers or a technical artifact (e.g., sample swapping) and underscoring the need for independent orthogonal validation. Where independent material was available, validations confirmed 13/13 tested primary drivers and 5/8 tested secondary drivers (Supporting Information S2: Table 5). Single-nucleotide variant calls as secondary drivers (*PAX5 P80R*, $n = 2$; *IKZF1 N159Y*, $n = 1$) were genomically confirmed in only one case with a

dominant *CRLF2*-driven *BCR::ABL1*-like gene expression signature, further emphasizing the importance of such validation in these special scenarios.

Blast fraction influences manual curation outcomes. ALLCatchR-inferred blast proportions in discordant (median: 75.2%; $P = 0.057$) or additional-driver cases (median: 77.3%, $P = 0.09$) were comparable to automatically classified samples (median: 73.9%). In contrast, unclassified cases (median: 54.7%; $P < 0.01$) and cases with missed drivers (median: 57.7%; $P < 0.01$) showed significantly lower blast

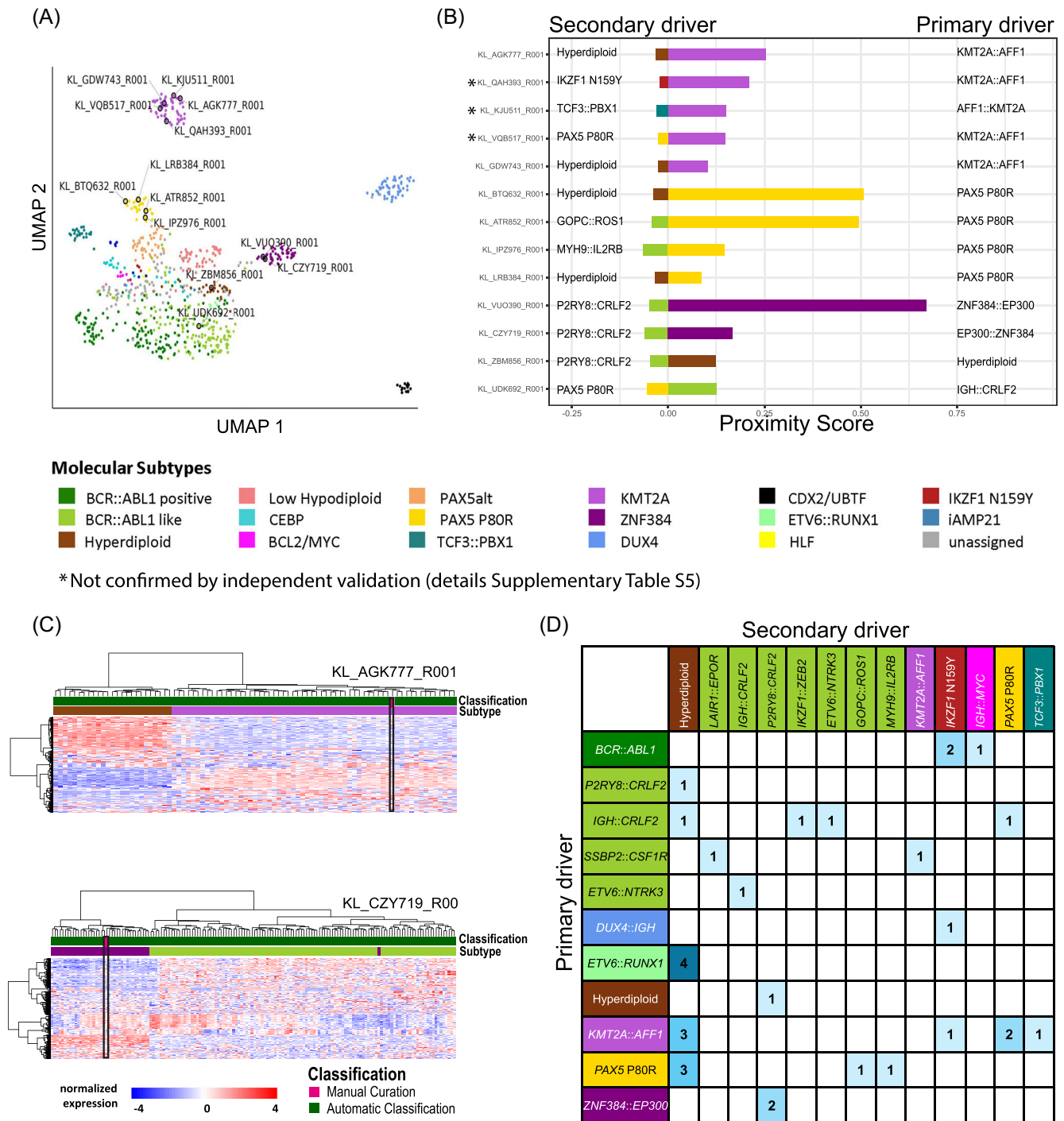


FIGURE 6 Characterization of double-driver cases. **(A)** UMAP projection of 774 samples based on 2802 LASSO-selected subtype-specific genes, highlighting double-driver cases ($n = 12$) by sample identifiers. **(B)** Gene expression-based proximity analyses of double-driver cases to both molecular subtypes were performed, analyzing each double-driver case together with all other samples of the two involved subtypes. Proximity scores were calculated using Euclidean distances of all samples of both involved subtypes to the sample of interest using principal component analysis of 600 top variable expressed genes in this sample set. Cases with BCR::ABL1-like driver fusions from distinct signaling trajectories (e.g., JAK/STAT and ABL-class) were excluded from this analysis. *Marks cases where RNA-seq driver call could not be validated in independent samples. **(C)** Heatmaps indicate unsupervised clustering (Ward.D2) of gene expression of two representative double-driver cases compared to other samples from both involved subtypes. **(D)** Table illustrating the distribution of genomic drivers identified in double-driver cases of the entire B-ALL data set ($n = 31/1210$; 2.6%). Molecular driver subtypes are characterized as primary or secondary according to the predominating gene expression-based subtype classification.

fractions, highlighting reduced sensitivity at low tumor content (Supporting Information S1; Figure 6). A similar pattern was observed for virtual karyotyping: samples with expression-based aneuploid subtypes and no KaryALL confirmation had lower predicted blast

fractions (median: 60.6%, $P > 0.001$) than automatically classified cases.

For validation, we tested IntegrateALL in 436 independent B-ALL samples from our own two cohorts ($n = 203$, including 98 unselected

routine-diagnostic cases; Munich Leukemia Laboratory) and one public cohort 233,²⁸ representing also pediatric patients. In total, 343 (78.7%) cases were automatically classified and 93 (21.3%) samples were flagged for manual curation (Supporting Information S1: Figure 7), confirming the distribution of our initial test set (Figure 5B). The reasons for manual curation—missed driver (41.9%), discordant driver (14.0%), and additional driver (16.1%)—were comparable, supporting IntegrateALL's applicability to independent data sets.

Dual drivers are called in 2.6% of B-ALL and are associated with one dominant gene expression program

To characterize the full spectrum of double-driver cases, we analyzed combined data across all cohorts ($n = 1210$ samples). IntegrateALL identified 31 (2.6%) samples with two driver calls corresponding to two distinct subtype definitions, confirming the observations from our first cohort (Figure 6D). Most frequently, hyperdiploidy co-occurred with subtypes, in which it is not commonly described (ETV6::RUNX1, $n = 4$; KMT2A, $n = 3$; PAX5 P80R, $n = 3$; and BCR::ABL1-like, $n = 2$). Several BCR::ABL1-like cases harbored driver fusions from different signaling trajectories (e.g., JAK/STAT and ABL1-class drivers or NTRK3; $n = 4$). The BCR::ABL1-like driver P2RY8::CRLF2 occurred in one hyperdiploid case and two ZNF384 cases. Subtype-defining single-nucleotide variants were also identified as additional drivers. Consistent with the primary cohort (Figure 6A,B), a single-expression program predominated in most dual-driver samples, underscoring the need for independent validation to distinguish true double drivers from technical artifacts. The distribution of secondary drivers suggests that several aberrations might function as primary drivers but also as cooperating events—which is well established for hyperdiploidy but might also apply to PAX5 P80R and P2RY8::CRLF2.

Available classifiers capture only part of IntegrateALL cross-validated subtype calls

Comparison of molecular subtype classifier accuracy typically relies on an independent ground truth, which is used for benchmarking individual tool outputs. Subtype definitions in B-ALL mostly rely on the combined presence of a subtype-defining driver call, absence of additional subtype-defining aberrations, and a corresponding gene expression signature. In some subtypes (e.g., BCR::ABL1-like, PAX5alt) a high-confidence gene expression classification is considered sufficient for subtype assignment. We defined our IntegrateALL classification rule set so, that it recapitulates this definition, meaning that the automated subtype classification indeed represents the corresponding ground truth, while in the remaining cases, the decision remains with the user to which extent he would accept less stringent definitions or would perform independent validation experiments. To compare IntegrateALL to other molecular subtype classification tools, we used 974/1210 samples (80.5%) with automatic IntegrateALL classification as ground truth and compared these labels to published gene expression-based classification tools—ALLSorts, ALLSpice, and ALLIUMv2—as well as the multimodal MD-ALL pipeline after harmonizing subtype nomenclature (Supporting Information S2: Table 6).

Compared to our ground truth, the highest overall accuracy for gene expression-based classification was achieved by ALLSorts (0.88), followed by ALLSpice (0.72) and ALLIUMv2 (0.71; Figure 7A–C, Supporting Information S2: Table 8). Outputs from these tools included unassigned/ambiguous classifications as own classes. When

excluding these samples from the benchmarking analysis, performance increased to 0.97, 0.96, and 0.83, respectively (“accuracy called”), highlighting missed and ambiguous classifications as the major limitation of gene expression classifiers.

We next benchmarked IntegrateALL against the MD-ALL pipeline, which combines gene expression, SNVs, and gene fusion calls. On 631/974 samples from our internal cohort with automatic classifications, MD-ALL achieved an accuracy of 0.92 (Figure 7D, Supporting Information S2: Table 7), indicating that 8% of high-confidence IntegrateALL classifications were discordant with MD-ALL despite the multilevel data integration. Most inconsistencies were observed with aneuploid karyotypes. Manual review of RNA-SeqCNV karyotypes and SNParray data, where available, supported the IntegrateALL subtype assignment, suggesting that KaryALL improved classifications in these subtypes, which are difficult to distinguish by gene expression alone.

“Low-hyperdiploid B-ALL” was the only molecular subtype represented in other classifiers (ALLSorts, ALLIUM, and MD-ALL) but not in IntegrateALL. Interestingly, all classifiers identified different samples as low hyperdiploid without any overlap (Supporting Information S2: Table 8), suggesting a need for a more precise definition of this subtype, which has not been included in hematological disease classifications^{1,2} so far.

Finally, we evaluated the agreement of gene expression classifiers, including the SVM gene expression module of MD-ALL, in the 236/1210 cases (19.5%) flagged by IntegrateALL for manual curation. Only 33/236 of these cases (14.0%) received consistent subtype definitions by all four gene expression classifiers, underscoring the need for manual curation and orthogonal validation among those cases (Supporting Information S1: Figure 8A). Conversely, 531/974 samples (54.5%) with automated IntegrateALL classification received identical subtype assignments by all classifiers, indicating both the higher level of agreement between all classifiers in this group and the added value of IntegrateALL's driver calls to confirm an unambiguous subtype classification in the remaining 443/974 cases (45.5%) with disagreements of the gene expression classifiers (Supporting Information S1: Figure 8B).

IntegrateALL's interactive report provides user-friendly access to heterogeneous tool outputs

To ensure consistent and intuitive access to the heterogeneous tool outputs, we developed an interactive HTML report (Figure 8). The interface highlights findings relevant to ALL subtype classification while preserving the ability to explore raw results. A navigation bar enables rapid movement between sections. The report includes a classification summary indicating WHO-HAEM5/ICC alignment (automated classification) or the need for manual curation; the ALL-CatchR gene expression-based subtype allocation; KaryALL classifications for aneuploid subtypes; RNA-Seq quality control metrics; hotspot single-nucleotide variant calls; RNASeqCNV-visualizations (including a karyogram summary); and gene fusion calls from ARRIBA and FusionCatcher presented in sortable tables with accompanying visualizations.

DISCUSSION

Recent updates in B-ALL classification have expanded the number of recognized molecular subtypes to a total of 12 according to WHO-HAEM5 and 27 according to ICC, including five provisional entities.^{1,2} Most of these subtypes are defined by distinct gene

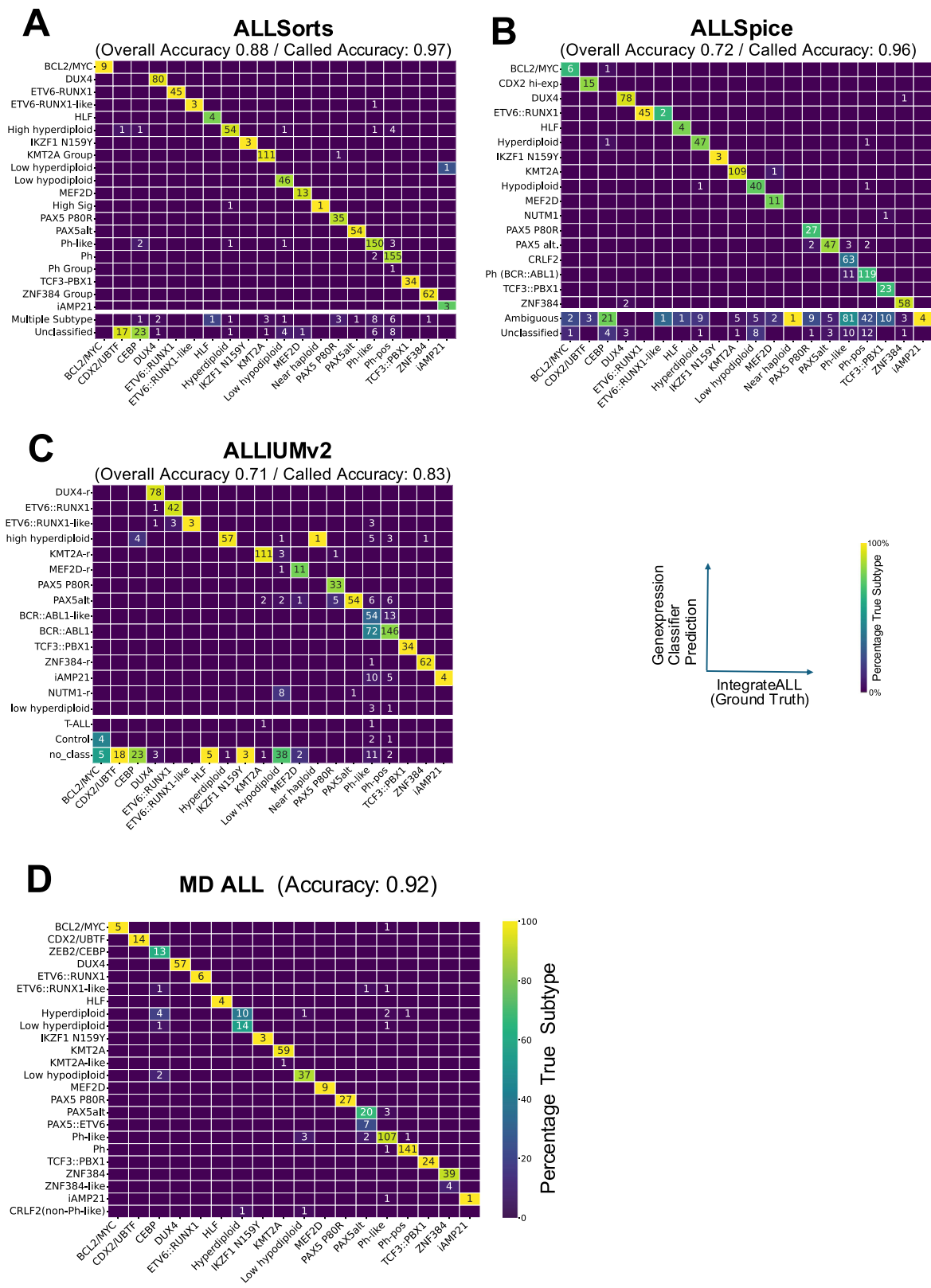


FIGURE 7 IntegrateALL benchmark against other tools and pipelines. We benchmarked four GEP-based classification tools (ALLSpice, ALLSorts, and ALLIUM) against IntegrateALL using 974 automatically classified samples from our internal and external cohort as ground truth. (A–C) Confusion matrices comparing predicted subtypes (Y-axis) to IntegrateALL ground truth (X-axis) for ALLIUM version 2 (A), ALLSorts (B), and ALLSpice (C). After harmonizing subtype nomenclature (Supporting Information S2: Table 6), ALLSorts demonstrated the highest accuracy (0.88), followed by ALLSpice (0.72) and ALLIUMv2 (0.71). (D) Confusion matrix for the MD-ALL multimodal classification pipeline, which integrates gene expression counts, SNVs, and fusions from both FusionCatcher and Cicero fusion callers. We analyzed 631/774 samples from the automatically classified internal IntegrateALL cohort. MD-ALL achieved an accuracy of 0.92, demonstrating robust performance by leveraging multiple data modalities beyond gene expression alone.

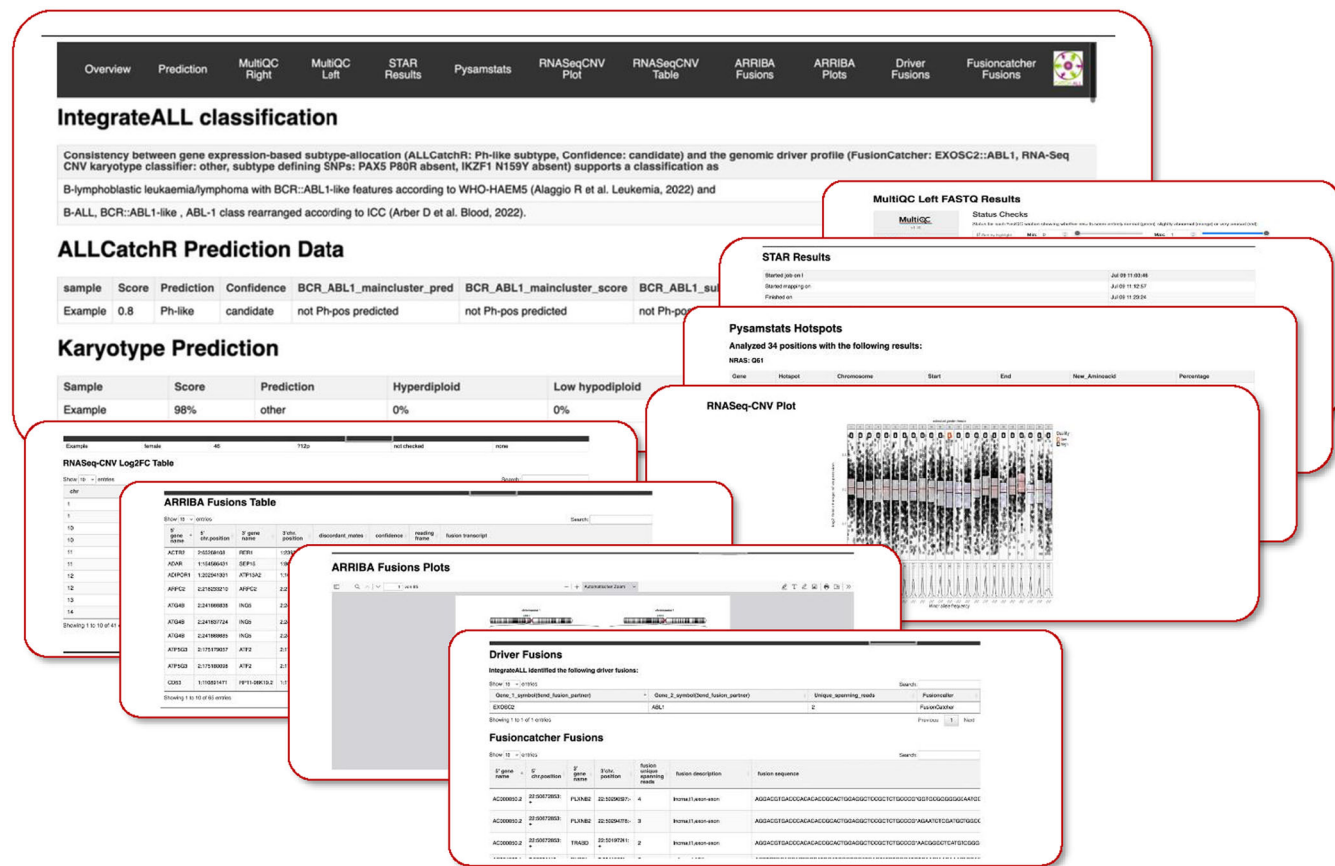


FIGURE 8 Overview of the IntegrateALL interactive report. To facilitate direct and user-friendly access to all data layers, IntegrateALL provides an interactive HTML document shown here for one example case. The top panel shows the navigation bar, linking to key sections: overview, prediction summary, and subtype classification according to ICC and WHO-HAEM5, quality control (MULTIQC), alignment (STAR), SNV analysis (GATK workflow, Pysamstats), copy number profiling (RNASeqCNV; KaryALL), and gene fusion calling (ARRIBA and FusionCatcher). Below, a concordant case is shown, alongside ALLCatchR and KaryALL predictions. The remaining thumbnails summarize individual analysis modules, providing a visual impression of the integrated output layout.

expression profiles coupled with specific genomic driver aberrations. Conventional diagnostic approaches require integration of multiple modalities— cytogenetics, FISH, breakpoint PCRs, and targeted sequencing—to capture all relevant alterations. This process is time-consuming, costly, and may miss cryptic lesions (e.g., *DUX4* fusions) that only manifest as expression signatures. In this context, RNA-seq has emerged as a single-platform alternative capable of detecting gene fusions, sequence mutations, virtual karyotypes, and the overall gene expression signature in one assay. Various recent studies demonstrate the feasibility and clinical value of RNA-seq-based classification, indicating an improved subtype classification beyond conventional diagnostics.^{27,30,33,34}

We and others have developed machine learning classifiers for B-ALL subtype classification (e.g., ALLSorts,⁹ Allium,¹¹ or ALLCatchR⁸). These tools achieve overall classification accuracies of 90%–95% and even higher in well-represented subtypes with clear-cut gene expression signatures. However, some rare subtypes remain a challenge, including iAMP21 or near-haploid ALL and samples with low blast proportions. Ambiguous cases—especially when harboring multiple drivers—might escape expression-based classification. Frameworks have been established that integrate outputs from gene expression classifiers with second data layers to improve classification with either sentinel driver aberrations from RNA variant calling (MD-ALL,³⁵ RaScALL,³⁶ or conventional diagnostics (clinALL¹³) as

well as DNA-methylation profiling (ALLIUM¹¹). These data underscore the improvements achieved for subtype classification by combining gene expression analysis with underlying drivers or independent functional signatures. Our benchmarking demonstrates that IntegrateALL's layered integration strategy achieves 100% accuracy for automatically classified samples, outperforming even multimodal pipelines like MD-ALL (92% accuracy), which relies on additional sequencing modalities. The fully automated end-to-end workflow, from raw FASTQ files to final classification, eliminates the need for manual preprocessing and complex tool integration, addressing a major practical barrier in clinical implementation. This combination of superior precision and operational simplicity positions IntegrateALL as a robust solution for routine diagnostic workflows.

Our newly developed IntegrateALL pipeline fits in this evolving landscape as an open source, end-to-end pipeline rather than a single-prediction algorithm. We aim to provide a workflow environment that leverages state-of-the-art methods for unsupervised, systematic identification of molecular drivers, including gene fusions, expression-based subtypes, SNVs, and karyotype abnormalities. For a systematic subtype allocation, we used published descriptions of individual molecular B-ALL subtypes,^{18,19} together with comprehensive cohort analyses^{16,20,27} and diagnostic classifications,^{1,2} to provide a comprehensive catalogue of genomic B-ALL drivers and corresponding gene expression signatures. An open design allows for

adaptation and, importantly, expansion, as novel B-ALL subtypes are still being identified.³⁷ We do not aim to outperform other existing classifiers, but to deliver a reproducible and transparent workflow that captures the correct subtype in the majority of samples automatically, while flagging atypical cases for manual curation. Validation of IntegrateALL in external cohorts achieved an unambiguous automatic classification in 83.7%, 84.8%, or 73.8% of samples, limiting the number of samples for detailed manual review. This approach acknowledges that no single classifier will meet diagnostic standards at handling the molecular complexity of B-ALL. Instead, by requiring concordance between gene expression subtype and the corresponding driver lesion for an automatic call, IntegrateALL prioritizes precision and interpretability.

Classification of karyotypes based on gene expression has relied so far on expert curation. Although clear enrichment patterns of non-random chromosomal gains and losses have been described,^{38,39} individual cases regularly present deviations from the overall average, and large-scale chromosomal gains/losses impart relatively subtle transcriptional effects. To carry out a systematic and quantitative analysis of aneuploid karyotypes and chr21 amplification/iAMP21, we established KaryALL, an ensemble machine learning classifier for systematic inference for virtual karyotypes from RNA-Seq data based on RNASeqCNV¹⁴ outputs. By training on 395 cases with SNParray-based karyotypes, KaryALL achieved high cross-validation accuracy (~98%) and F1 score (~0.96%) in identifying near-haploid, low-hypodiploid, hyperdiploid, and iAMP21 cases. Importantly, KaryALL was able to resolve the ambiguity between near-Haploid and hypodiploid ALL, which are difficult to separate by gene expression alone. To our knowledge, KaryALL is the first classifier for virtual B-ALL karyotypes providing quantitative confidence in the karyotype call, again aiding in the distinction between clear-cut cases and rare candidates that would need orthogonal validations for a correct subtype assignment.

Typically, B-ALL samples harbor a single genomic driver aberration or a subtype-defining aneuploidy pattern. Targeted approaches for B-ALL classification therefore might stop after identifying the first—probably more frequent or more prominent—driver. IntegrateALL flags these cases for manual curation. Application to large cohorts ($n = 1210$ cases) provides first estimates that 2.6% of B-ALL cases harbor multiple subtype-defining aberrations simultaneously. The co-occurrence of hyperdiploid karyotypes and *ETV6::RUNX1*, *PAX5 P80R*, *KMT2A-*, or *CRLF2*-rearrangements is not unexpected, since hyperdiploidy is also observed in *BCR::ABL1*-positive ALL.³² However, it is a novel observation that hyperdiploidy can be acquired independently of the underlying driver fusion. Less expected was the co-occurrence of *BCR::ABL1*-like ALL driver fusions (e.g., *CRLF2*, *ROS1*, or *IL2RB*) or *EP300::ZNF384* with *PAX5 P80R* mutations. *PAX5 P80R* has so far been described as an exclusive driver with a subtype-defining gene expression signature.^{17,19,40} Our data suggest that this mutation might also cooperate with other drivers during leukemogenesis or that multiclonal driver events co-occur in the same patient. Similarly, the subtype-defining^{41,42} *IKZF1 N159Y* mutation was also observed to co-occur with *BCR::ABL1*, *IGH::DUX4*, or *KMT2A::AFF1* gene fusions on the RNA level, while independent genomic validation failed in one tested case of our cohort. IntegrateALL also identified two driver fusions from different signaling types (JAK/STAT-driven plus ABL class or other) in *BCR::ABL1*-positive/-like cases, confirming case reports with similar observations.^{43–46} Notably, the *BCR::ABL1*-like driver *P2RY8::CRLF2* seems to be especially prone to co-occurrence with other drivers.⁴⁶

While these reports are mostly anecdotal, IntegrateALL identifies such double-driver cases systematically. Remarkably, our gene expression analyses identified one dominant subtype-specific gene expression signature in most double-driver cases, suggesting that one driver remains in control of oncogenic signaling. Sample swapping or

other technical artifacts remain a possibility. However, the non-random enrichment patterns of specific dual drivers—partly confirmed by validation in independent samples of our cohort—seem to be more in line with co-occurrence in B-ALL. Importantly, IntegrateALL identifies such rare candidates—likely missed by more focused approaches—for further orthogonal validation. In all cases reported in the literature, patients with double drivers tended to be high risk and often required combination therapy (e.g., TKI plus chemotherapy),^{43,44} underscoring the need for systematic unsupervised identification.

Research requires reproducible workflows. IntegrateALL provides a systematic pipeline of standard tools used for B-ALL research. The Snakemake environment facilitates implementation of novel applications for specific research questions, which can then build upon a systematic and comparable baseline characterization. Free accessibility of all raw outputs enables unsupervised analysis beyond established subtype definitions and drivers. An open design allows for seamless implementation of novel classification rules.

In conclusion, the integration of machine learning into the analysis of B-ALL RNA-Seq data represents an important step toward a more systematic and data-driven classification of leukemia subtypes. IntegrateALL facilitates diagnostic workflows by enabling large-scale, standardized profiling while simultaneously identifying complex cases that require manual curation. By incorporating gene expression-based subtype classification, karyotype prediction, and secondary driver detection, the pipeline provides a comprehensive framework for refining the molecular characterization of B-ALL.

The results demonstrate that computational models can enhance the accuracy and consistency of subtype classification while also capturing additional genomic complexity that may influence disease progression. The ability to distinguish primary from secondary drivers further supports a more nuanced interpretation of cases with multiple alterations. These findings highlight the potential of machine learning-assisted approaches to improve leukemia diagnostics and provide a scalable method for integrating diverse genomic data in clinical and research settings.

ACKNOWLEDGMENTS

The large language model ChatGPT (OpenAI, USA) was used for language editing without interfering with any content of the manuscript. The large language model Claude (Anthropic, USA) was used for code refactoring and optimization without interfering with any scientific content or methodology.

AUTHOR CONTRIBUTIONS

Nadine Wolgast: Writing—original draft; validation; Methodology; visualization; writing—review and editing; software; formal analysis; data curation; conceptualization; investigation. **Thomas Beder:** Methodology; writing—review and editing; software; conceptualization; investigation. **Mayukh Mondal:** Writing—review and editing; software; validation. **Wencke Walter:** Resources; writing—review and editing; validation. **Stephan Hutter:** Validation; writing—review and editing. **Sonja Bendig:** Writing—review and editing; validation. **Jan Kässens:** Validation; writing—review and editing. **Björn-Thore Hansen:** Methodology; writing—review and editing. **Katharina Iben:** Writing—review and editing; validation. **Sebastian Wolf:** Resources; writing—review and editing; validation. **Anjali Cremer:** Resources; writing—review and editing; validation. **Malwine Barz:** Writing—review and editing; validation; methodology. **Martin Neumann:** Resources; writing—review and editing. **Nicola Gökbüget:** Resources; writing—review and editing. **Claudia Haferlach:** Writing—review and

editing; resources; validation. **Monika Brüggemann:** Resources; writing—review and editing; validation. **Claudia D. Baldus:** Conceptualization; investigation; writing—review and editing; supervision; funding acquisition. **Alina M. Hartmann:** Writing—original draft; conceptualization; supervision; project administration; validation; writing—review and editing; funding acquisition; investigation. **Lorenz Bastian:** Conceptualization; investigation; writing—original draft; methodology; project administration; supervision; writing—review and editing; validation; data curation; funding acquisition; formal analysis; visualization.

CONFLICT OF INTEREST STATEMENT

Monika Brüggemann reports honoraria from Amgen, BD, Janssen, and Pfizer for participation in speakers' bureaus, as well as travel support from these companies. She has also served on advisory boards for Amgen, AstraZeneca, Hello Healthcare, and Incyte. The remaining authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are partly published^{16,17} and are openly available in Zenodo at <https://doi.org/10.5281/zenodo.18310354>. Additional raw sequencing data are made available upon request.

The code used in this work is available in a publicly accessible repository at <https://github.com/NadineWolgast/IntegrateALL> under the MIT License, ensuring open access for modification and distribution. For detailed guidance on implementing and running the pipeline, users can refer to the comprehensive instructions provided in the repository's README file, available at <https://github.com/NadineWolgast/IntegrateALL/tree/main#readme>. To facilitate reproducibility and ease of deployment, a Dockerfile is provided in the repository, enabling users to build a containerized environment with all necessary dependencies pre-configured.

ETHICS STATEMENT

All patients provided written informed consent in accordance with the Declaration of Helsinki. Data and biosamples were collected for the storage and use of residual material for medical research purposes under the Broad Informed Consent framework implemented at the University Medical Center Schleswig-Holstein (UKSH), Campus Kiel, and approved by the local ethics committee.

FUNDING

This study was funded in part by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) project number 444949889 (KFO 5010/1 Clinical Research Unit "CATCH ALL" to MN, MBr, CDB, AHM, and LB), and project number 413490537 (Clinician Scientist Program in Evolutionary Medicine) to B-TH and by the Deutsche Krebshilfe (DKH, German Cancer Aid) project number 70115443 to LB. LB was funded by the Faculty of Medicine of Kiel University within the 'Advanced Clinician Scientist Program'. Open Access funding enabled and organized by Projekt DEAL.

SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article.

ORCID

Nadine Wolgast  <https://orcid.org/0000-0002-3068-3899>

Thomas Beder  <https://orcid.org/0009-0000-2278-5054>

Mayukh Mondal  <https://orcid.org/0000-0003-0122-0323>

Wencke Walter  <https://orcid.org/0000-0002-5083-9838>

Stephan Hutter  <https://orcid.org/0000-0003-2432-8898>

Sonja Bendig  <https://orcid.org/0009-0003-0644-9140>

Jan Kässens  <https://orcid.org/0000-0001-7840-1160>

Björn-Thore Hansen  <https://orcid.org/0000-0003-0482-4472>

Sebastian Wolf  <https://orcid.org/0009-0008-9993-9750>

Anjali Cremer  <https://orcid.org/0000-0003-4701-842X>

Malwine Barz  <https://orcid.org/0000-0003-1581-2020>

Nicola Gökbüget  <https://orcid.org/0000-0003-2291-8245>

Claudia Haferlach  <https://orcid.org/0000-0002-6333-5049>

Monika Brüggemann  <https://orcid.org/0000-0001-5514-5010>

Claudia D. Baldus  <https://orcid.org/0000-0002-0748-834X>

Alina M. Hartmann  <https://orcid.org/0009-0002-0638-0000>

Lorenz Bastian  <https://orcid.org/0000-0002-1487-9437>

REFERENCES

- 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(7):1720-1748. doi:10.1038/s41375-022-01620-2.
- 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. doi:10.1182/blood.2022015850
- Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
- Nicorici D, Satalan M, Edgren H, et al. FusionCatcher—a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *bioRxiv*. Preprint posted online November 19, 2014. doi:10.1101/011650
- Uhrig S, Ellermann J, Walther T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res*. 2021;31(3):448-460. doi:10.1101/GR.257246.119
- Miles A, Mattioni M. *pysamstats: A Python Utility for Calculating Statistics Against Genome Positions Based on Sequence Alignments from a SAM or BAM file*. Zenodo; 2015. doi:10.5281/ZENODO.34480
- Van der Auwera G, O'Connor B. *Genomics in the Cloud*. O'Reilly Media, Inc.; 2020.
- Beder T, Hansen BT, Hartmann AM, et al. The gene expression classifier ALLcatchR identifies B-cell precursor ALL subtypes and underlying developmental trajectories across age. *HemaSphere*. 2023;7(9):e939. doi:10.1097/HS9.0000000000000939
- Schmidt B, Brown LM, Ryland GL, et al. ALLSorts: an RNA-Seq subtype classifier for B-cell acute lymphoblastic leukemia. *Blood Adv*. 2022;6(14):4093-4097. doi:10.1182/bloodadvances.2021005894
- Mäkinen VP, Rehn J, Breen J, Yeung D, White DL. Multi-cohort transcriptomic subtyping of B-cell acute lymphoblastic leukemia. *Int J Mol Sci*. 2022;23(9):4574. doi:10.3390/ijms23094574
- Krali O, Marincevic-Zuniga Y, Arvidsson G, et al. Multimodal classification of molecular subtypes in pediatric acute lymphoblastic leukemia. *NPJ precision oncology*. 2023;7(1):131. doi:10.1038/s41698-023-00479-5
- Gu Z, Hu Z, Jia Z, Liu J, Mao A, Han H. MD-ALL: an integrative platform for molecular diagnosis of B-cell acute lymphoblastic leukemia. *Research Square*. Preprint posted online April 14, 2023. doi:10.21203/rs.3.rs-2798895/v1
- Tang M, Antić Ž, Fardzadeh P, et al. An artificial intelligence-assisted clinical framework to facilitate diagnostics and translational discovery in hematologic neoplasia. *EBioMedicine*. 2024;104:105171. doi:10.1016/j.ebiom.2024.105171

14. Bařinka J, Hu Z, Wang L, et al. RNAseqCNV: analysis of large-scale copy number variations from RNA-seq data. *Leukemia*. 2022;36(6):1492-1498.
15. M3lder F, Jablonski KP, Letcher B, et al. Sustainable data analysis with Snakemake. *F1000Research*. 2021;10:33. doi:10.12688/f1000research.29032.1
16. Bastian L, Hartmann AM, Beder T, et al. UBTF::ATXN7L3 gene fusion defines novel B cell precursor ALL subtype with CDX2 expression and need for intensified treatment. *Leukemia*. 2022;36(6):1676-1680. doi:10.1038/s41375-022-01557-6
17. Bastian L, Schroeder MP, Eckert C, et al. PAX5 biallelic genomic alterations define a novel subgroup of B-cell precursor acute lymphoblastic leukemia. *Leukemia*. 2019;33(8):1895-1909. doi:10.1038/s41375-019-0430-z
18. Harvey RC, Tasian SK. Clinical diagnostics and treatment strategies for Philadelphia chromosome-like acute lymphoblastic leukemia. *Blood Adv*. 2020;4(1):218-228. doi:10.1182/bloodadvances.2019000163
19. Gu Z, Churchman ML, Roberts KG, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet*. 2019;51(2):296-307. doi:10.1038/s41588-018-0315-5
20. Brady SW, Roberts KG, Gu Z, et al. The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat Genet*. 2022;54(9):1376-1389. doi:10.1038/s41588-022-01159-z
21. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham Bioinformatics; 2010.
22. Ewels P, Magnusson M, Lundin S, K3ller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi:10.1093/bioinformatics/btw354
23. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-Learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830. doi:10.5555/1953048.2078195
24. Lemaitre G, Nogueira F, Aridas CK. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *arXiv*. Published online September 21, 2016. <http://arxiv.org/abs/1609.06570>
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357. doi:10.1613/jair.953
26. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016;13-14:785-794. doi:10.1145/2939672.2939785
27. Walter W, Shahswar R, Stengel A, et al. Clinical application of whole transcriptome sequencing for the classification of patients with acute lymphoblastic leukemia. *BMC Cancer*. 2021;21(1):886. doi:10.1186/s12885-021-08635-5
28. Autry RJ, Paugh SW, Carter R, et al. Integrative genomic analyses reveal mechanisms of glucocorticoid resistance in acute lymphoblastic leukemia. *Nat Cancer*. 2020;1(3):329-344. doi:10.1038/s43018-020-0037-3
29. Yu CH, Wu G, Chang CC, et al. Sequential approach to improve the molecular classification of childhood acute lymphoblastic leukemia. *J Mol Diagn*. 2022;24(11):1195-1206. doi:10.1016/j.jmoldx.2022.08.001
30. Tran TH, Langlois S, Meloche C, et al. Whole-transcriptome analysis in acute lymphoblastic leukemia: a report from the DFCI ALL consortium protocol 16-001. *Blood Adv*. 2022;6(4):1329-1341. doi:10.1182/bloodadvances.2021005634
31. Harrison CJ, Moorman AV, Broadfield ZJ, et al. Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *Br J Haematol*. 2004;125(5):552-559. doi:10.1111/j.1365-2141.2004.04948.x
32. Bastian L, Beder T, Barz MJ, et al. Developmental trajectories and cooperating genomic events define molecular subtypes of BCR::ABL1-positive ALL. *Blood*. 2024;143(14):1391-1398. http://ashpublications.org/blood/article-pdf/143/14/1391/2219824/blood_bld-2023-021752-main.pdf
33. Farnoud N, Liosis K, Leongamornlert D, et al. Diagnostic utility of comprehensive RNA-Seq analysis in adult B-ALL. *Blood*. 2023;142(Suppl 1):1601. doi:10.1182/blood-2023-190519
34. Boer JM, Koudijs MJ, Kester LA, et al. Challenging conventional diagnostic methods by comprehensive molecular diagnostics: a nationwide prospective comparison in children with ALL. *JCO Precis Oncol*. 2025;9:e2400788. doi:10.1200/PO-24-00788
35. Hu Z, Jia Z, Liu J, Mao A, Han H, Gu Z. MD-ALL: an integrative platform for molecular diagnosis of B-acute lymphoblastic leukemia. *Haematologica*. 2024;109(6):1741-1754. doi:10.3324/haematol.2023.283706
36. Rehn J, Mayoh C, Heatley SL, et al. RaScALL: rapid (Ra) screening (Sc) of RNA-seq data for prognostically significant genomic alterations in acute lymphoblastic leukaemia (ALL). *PLoS Genet*. 2022;18(10):e1010300. doi:10.1371/journal.pgen.1010300
37. Bendig S, Hartmann AM, Wessels W, et al. IGH::FENDRR and specific KRAS mutations define a novel B-ALL molecular subtype with poor chemotherapy response. *Blood*. 2026;2025031102. doi:10.1182/blood.2025031102
38. Haas OA, Borkhardt A. Hyperdiploidy: the longest known, most prevalent, and most enigmatic form of acute lymphoblastic leukemia in children. *Leukemia*. 2022;36(12):2769-2783. doi:10.1038/s41375-022-01720-z
39. Holmfeldt L, Wei L, Diaz-Flores E, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat Genet*. 2013;45(3):242-252. doi:10.1038/ng.2532
40. Passet M, Boissel N, Sigaux F, et al. PAX5 P80R mutation identifies a novel subtype of B-cell acute lymphoblastic leukemia with favorable outcome. *Blood*. 2019;133:280-284. doi:10.1182/blood-2018-10-882142
41. Li JF, Dai YT, Lilljebj3rn 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. doi:10.1073/pnas.1814397115
42. Alvey EN, Yap KL, Rathbun P, Fitzpatrick C, Gong S, Richardson AI. B-cell lymphoblastic leukemia/lymphoma with mutated *IKZF1* N159Y: clinical and genetic features of an emerging entity. *Haematologica*. 2025;110:3078-3082. doi:10.3324/haematol.2024.286626
43. Jain N, Lu X, Daver N, et al. Co-occurrence of CRLF2-rearranged and Ph+ acute lymphoblastic leukemia: A report of four patients. *Haematologica*. 2017;102(12):e514-e517. doi:10.3324/haematol.2016.161000
44. He G, Lei Y, Huang D, Gao J, Yang R. Philadelphia chromosome-like acute lymphoblastic leukemia with concomitant rearrangements of CRLF2 and ABL1: a pediatric case report. *BMC Pediatr*. 2024;24(1):517. doi:10.1186/s12887-024-04991-w
45. Balatzenko G, Guenova M, Kalinova I, Belcheva M, Hristozova H, Kaleva V. Simultaneous occurrence of ETV6-RUNX1 and BCR-ABL1 (e1a2) transcripts in a child with B-cell acute lymphoblastic leukemia. *Cancer Genet*. 2013;206(3):97-101. doi:10.1016/j.cancergen.2013.01.004
46. Russell LJ, Jones L, Enshaei A, et al. Characterisation of the genomic landscape of CRLF2-rearranged acute lymphoblastic leukemia. *Genes Chromosom Cancer*. 2017;56(5):363-372. doi:10.1002/gcc.22439