SUPPLEMENTAL INFORMATION FOR

Transcriptomic analysis reveals pro-inflammatory signatures associated with acute myeloid leukemia progression

Svea Stratmann[1*], Sara A. Yones[2*], Mateusz Garbulowski[2], Jitong Sun[1], Aron Skaftason[3], Markus Mayrhofer[4], Nina Norgren[5], Morten Krogh Herlin[6], Christer Sundström[1], Anna Eriksson[7], Martin Höglund[7], Josefine Palle[8], Jonas Abrahamsson[9], Kirsi Jahnukainen[10], Monica Cheng Munthe-Kaas[11,12], Bernward Zeller[12], Katja Pokrovskaja Tamm[13], Lucia Cavelier[1], Jan Komorowski[2,14,15,16], Linda Holmfeldt[1,17]

[1]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden

[2]Science for Life Laboratory, Department of Cell and Molecular Biology; Uppsala University, Sweden

[3]Department of Molecular Medicine and Surgery; Karolinska Institutet, Stockholm, Sweden

[4]National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Sweden

[5]Department of Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Umeå University, Sweden

[6]Department of Clinical Medicine - Department of Pediatrics and Adolescent Medicine, Aarhus University, Denmark

[7]Department of Medical Sciences, Uppsala University, Sweden

[8]Department of Women's and Children's Health, Uppsala University, Sweden

[9]Department of Pediatrics, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

[10]Children's Hospital, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland

[11]Norwegian Institute of Public Health, Oslo, Norway

[12]Division of Pediatric and Adolescent Medicine, Oslo University Hospital, Oslo, Norway

[13]Department of Oncology and Pathology, Karolinska Institutet and Karolinska University hospital, Sweden

[14]Swedish Collegium for Advanced Study, Uppsala, Sweden

[15]Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

[16]Washington National Primate Research Center, Seattle, WA, USA

[17]The Beijer Laboratory, Uppsala, Sweden

[*]These authors contributed equally to this work.

Table of contents

Supplemental Methods

Cohort and sample characteristics

Inclusion criteria for this study were; all acute myeloid leukemia (AML; excluding acute promyelocytic leukemia [APL]) cases with available relapse- or primary resistant (PR) RNA material of sufficient quality and yield from the Nordic countries. Samples were collected from 1995 through 2016 from the following biobanks: U-CAN[1]; Clinical Pathology, Uppsala University Hospital, Sweden; Nordic Society of Paediatric Haematology and Oncology (www.nopho.org), with all of these sample collections being part of Uppsala Biobank; as well as a sample collection at Astrid Lindgren's Children's Hospital, Stockholm, Sweden, part of the Karolinska Institute Biobank. Further details on clinical and biological characteristics are summarized in **Supplemental Tables 2-3**. Genomic characterization of the entire study cohort was reported previously[2] (data available via controlled access: doi.org/10.17044/scilifelab.12292778), including bone marrow (BM) derived normal stromal cells as well as complete remission BM samples as a source of germline DNA.

All patients were diagnosed according to the WHO criteria for AML[3,4], and classified according to the ELN-risk classification[5] for adult AML and the NOPHO-DBH AML 2012 Protocol (EudraCT Number 2012-002934-35) for pediatric AML. Event-free survival (EFS) was defined as the time from initial diagnosis to first relapse or initial treatment failure. Short EFS was set at <6 months for adults and <12 months for pediatric patients. PR was defined as treatment failure without reaching first complete remission, while persistent relapse (R-P) samples were acquired post-relapse treatment from patients not achieving complete remission after the respective relapse. For the pediatric cohort, a sample was defined as treatment resistant if the patient did not achieve complete remission after intensive treatment. Corresponding resistance data were largely missing for the adult cohort.

Next-generation transcriptomic sequencing

Transcriptomic analysis was performed by RNA-sequencing (RNA-seq) on 122 tumor samples (**Supplemental Table 1**) and five CD34+ BM control samples (from here-on referred to as BM-controls; **Supplemental Table 4**). Extracted RNA was qualified by automated electrophoresis using a TapeStation 4200 (Agilent Technologies, Santa Clara, CA, USA) and quantified by a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Except for AML008-PR (RNA integrity number [RIN]=5.8), AML017-R2 (RIN=6.9) and AML043-D

(RIN=7.5), only RNA samples with a RIN≥8 were included in the study. Library preparation and sequencing were carried out at the SNP&SEQ Technology Platform, SciLifeLab, National Genomics Infrastructure (NGI), Uppsala, Sweden.

RNA-seq libraries were prepared from 500ng total RNA for 75 samples from 47 adult patients, and from 450ng total RNA for 47 samples from 23 pediatric patients (**Supplemental Table 2**), using the TruSeq stranded total RNA library preparation kit with ribosomal depletion by RiboZero Gold (Illumina, San Diego, CA, USA) according to the manufacturer's protocol (#15031048). Sequencing of adult samples was carried out on the Illumina HiSeq2500 platform, generating paired-end 125 base pair (bp) reads using v4 sequencing chemistry. An average of 40.1 million reads per sample was generated (range: 15.9 – 64.7; median: 38.4). Sequencing of pediatric samples was carried out on the Illumina NovaSeq6000 platform (S2 flowcell), generating paired-end 100bp reads using the v1 sequencing chemistry. An average of 43.1 million reads per sample was generated (range: 10.5–160.2; median: 29.3). BM-control samples from five individual healthy donors were sequenced in technical duplicates using both platforms (Illumina HiSeq2500 and NovaSeq6000).

Raw paired-end sequencing reads were aligned using the nf-core/rnaseq (v.1.0; ref. [6]) pipeline written in Nextflow[7]. Briefly, raw reads were adapter trimmed with the help of trim-galore (v.0.5.0; ref.[8]) using standard parameters and mapped to the reference genome (hg19) using STAR (v.2.6.1; ref.[9]). Duplicate reads were estimated with Picard´s MarkDuplicates (v.2.18.14; ref.[10]) and Dupradar (v.1.8.0; ref.[11]) and marked for downstream processing. Gene counts were retrieved with the help of FeatureCounts (v.1.6.2; ref.[12]). Spanning splicing events were hard-clipped utilizing the GATK (v.4.0.12; ref.[13]) tool SplitNCigarReads and mapping qualities were reassigned by the GATK tool ReassignOneMappingQuality. Quality of reads was determined using fastqc (v.0.11.7; ref.[14,15]) and RSeQC (v.2.6.4; ref.[16]) and quality metrics were summarized with the help of MultiQC (v.1.6). Single nucleotide variants (SNVs) and small insertion and deletion mutations (InDels; <50bp) were called by HaplotypeCaller (GATK) using default settings for RNA-seq data, and further filtered utilizing VariantFiltration (GATK). SNVs and small InDels were filtered against: (i) filtering clusters of at least three SNVs within a window of 35 nucleotides, (ii) Fisher Strand values greater than 30.0, and (iii) low quality reads with quality score less than 30. RNA-seq fusion transcripts were called via STAR-Fusion (v.1.5.0; ref.[17]) following pre-defined settings.

Filtering and manual curation of transcriptomic variants

First, known common single nucleotide polymorphisms (SNPs) were excluded by filtering the RNA-seq HaplotypeCaller (GATK) output against dbSNP (build 138; ref.[18]). In a second step, variants were removed unless they fulfilled the following criteria: (i) variant allele frequency (VAF) greater than or equal to 0.1; (ii) present in less than three Swegen[19] samples; (iii) present in less than 10 reads in all BM-controls; (iv) present in less than two normal samples from the pool of all whole genome sequencing (WGS) normal controls within our cohort (n=60; ref.[2]). Rescuing was performed for variants that were flagged according to Cosmic70 (ref.[20]) or ClinVar[21], or were validated as somatic on the genomic level by WGS or whole exome sequencing (WES), or were present in another sample that passed the above filtering criteria. Finally, for the current study, we focused solely on protein-coding variants.

Subsequently, all potential somatic variants were manually validated utilizing the UCSC genome browser[22] and by inspecting the sequencing reads at the respective region, for which a genomic and/or transcriptomic variant was reported, using Integrative Genomics Viewer (IGV, v.2.5.3; ref.[23]). Remaining normal variants were identified and removed by comparing the data against their respective patient-matched normal WGS or WES sample and the BM-control samples (**Supplemental Tables 7** and **8**).

Manual curation and technical validation of RNA fusions

The output of STAR-fusions was further filtered, and transcript fusions characterized by one or more of the following criteria were excluded from the study: (i) Fusion transcripts involving uncharacterized genes, immunoglobulin genes or long noncoding RNA genes; (ii) Fusion transcripts that were also found in BM-control samples; (iii) Fusion transcripts with a minimum FFPM (fusion fragments per million total reads) <0.1; and (iv) Fusion transcripts identified in healthy tissue based on FusionHub (https://fusionhub.persistent.co.in/home.html as of October 17 2020). Fusion transcripts were rescued if present in a patient-matched sample that passed the above stated FFPM filtering criterion. See **Supplemental Table 9** for the final fusion list.

Technical validation of a subset of putative fusion transcripts was performed by reverse transcriptase (RT)-PCR using primers targeting the respective area of interest on cDNA from leukemia cells (**Supplemental Table 20**).

Fusion genes were visualized at the cohort level using Circos (v.0.63-9; ref.[24]), while fusions, copy number alterations and sequence mutations were visualized in a sample specific manner

using Shinycircos[25] (as of September 7, 2020) for AML028 and AML071, incorporating data also from WGS analysis.

Pre-processing of RNA read counts for gene expression- and machine learning analysis

Gene count matrices for each cohort were first filtered for protein-coding genes. Genes that had low expression across most samples were investigated and removed using R v.4.0.1 (ref.[26]). To compute a threshold for filtering out the non- and lowly expressed genes, library sizes of each gene were estimated for each sample using edgeR R package v.3.28.1 (ref.[27]). The median of the library sizes for each gene was then computed and normalized by dividing the median over one million. Following the edgeR best practices, a gene was deemed as expressed if it had five or more samples with greater than or equal to 10/(Median of the library size of the gene). After filtering away the non- and lowly expressed genes, the gene expression matrix was normalized using the trimmed mean of M-values normalization method (TMM[28]).

Validation cohorts

The Cancer Genome Atlas (TCGA) AML cohort[29] was used to validate the association between disease outcome among adult AML patients and the expression levels of DEGs as well as genes identified through machine learning-based analysis in our study. Gene expression profiles for TCGA LAML (phs000178) were downloaded from the National Cancer Institute GDC data portal using the RTCGAToolbox[30]. In total, 162 samples with available RNA-seq data were included, while APL-samples and samples from patients below the age of 19 at initial diagnosis, were excluded (**Supplemental Table 10F**).

The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) AML cohort (phs000465) was utilized for validation purposes among pediatric AML patients. Expression data for 316 AML tumor samples were downloaded through https://portal.gdc.cancer.gov/projects as of November 16 2018. Twenty-two samples were excluded from the analysis according to one of the following reasons: (i) patient age at AML onset ≥19; and/or (ii) insufficient metadata for the respective analysis. In total, 254 diagnosis samples were used to generate Kaplan-Meier plots for investigation of association between with gene expression levels and EFS as well as OS (**Supplemental Table 10E**). Further, 29 diagnosis samples and 38 relapse samples, including 29 patient-matched diagnosis-relapse pairs, were utilized for machine learning-based analysis (**Supplemental Table 10D**).

Raw counts from both validation cohorts were further processed and analyzed as described above for the local adult and pediatric cohorts.

Differential gene expression analysis using Qlucore

Pre-processed and normalized genes were annotated, and the gene length for each gene was calculated from the union of all isoforms of each gene via the corresponding GTF-file utilizing GenomicRanges (v.1.40.0; ref.[31]), rtracklayer (v.1.48.0; ref.[32]), and Rsamtools (v.2.4.0; ref.[33]). The calculated gene lengths were subsequently used to adjust the TMM-normalized data for the gene length, utilizing Qlucore omics explorer v.3.6 (Qlucore AB, Lund, Sweden). Thereafter, read counts for a total of 15546 protein-coding genes identified for the local adult and pediatric R/PR AML cohorts were log2 transformed using Qlucore (normalization Z-score [mean=0, var=1]). Normalization was performed separately for each conducted analysis as detailed in **Supplemental Table 10**. The resulting data were batch corrected (ref.[34] applied through Qlucore) for the applied sequencing method (Illumina HiSeq2500 vs Illumina NovaSeq6000) and patient sex. No further confounding effects were detected.

Gene ontology enrichment analysis

Gene Ontology (GO) enrichment analysis was carried out using Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla[35,36]; http://cbl-gorilla.cs.technion.ac.il/ as of October 17, 2020). A target list of up- or downregulated genes (|log2 fold change [log2FC]|>1; corresponding to a minimum fold change of +/-2) was compared to the background of all expressed, protein-coding genes (n=15546) using the standard Hyper Geometric statistics with a P-value threshold of 0.01. To correct for multiple comparisons, the Benjamini-Hochberg[37] method was used.

Interpretable supervised learning to obtain rule-based classifiers for disease states

Overview of the analysis pipeline

A graphical overview of the analysis pipeline is given in **Supplemental Figure 10**. In brief, RNA-seq data for the local adult cohort (diagnosis: n=22; relapse: n=42), the local pediatric cohort (diagnosis: n=17; relapse: n=22), and a pediatric validation cohort (TARGET; diagnosis [n=29] and relapse [n=38] samples part of phs000465 at https://portal.gdc.cancer.gov/projects) were individually used as training data to create three machine learning models to discern between diagnosis and relapse states (**Supplemental Table 10C** and **D**). The results from the models using the local cohorts were analyzed individually and the co-predictive interactions

between features and their values were visualized in the form of rule networks (**Figure 4** and **Supplemental Figure 13**) using VisuNet (v.1.3.5; ref.[38]). A rule network is constructed from nodes that represent genes, and edges that represent connections between genes. Within a network, co-predictive genes are visible as highly connected nodes, which also are called hubs. This graphic representation is a way to visualize a rule-based model and its statistics. Importantly, rule-based networks differ from co-expression networks in that each decision class, here diagnosis and relapse, have separate networks and use genes with different expression levels. Genes in rule-based networks are co-predictors of a certain decision class, meanwhile genes in co-expression networks are co-expressed but not necessarily co-related to the decision class. Due to the small number of samples for the local pediatric cohort, the predictive features from the models on the local pediatric- and TARGET data were merged in order to increase the power of the machine learning models. Following merging, new models were created for the local pediatric- and TARGET cohorts, utilizing the newly merged predictive features for the respective dataset. Rule-based networks were then built for the new models and network analysis approaches were used to compare the similarity of networks for the models to discover co-predictive patterns that are comparable in both cohorts (**Figure 5** and **Supplemental Figure 14**). More details for each of these steps are given in the following seven sub-sections below.

Pre-processing of RNA-seq data

Genes were pre-processed and near zero variance genes were removed. Removal of near zero variance genes was based on the following characteristics: (i) if they had very few unique expression values relative to the number of samples based on calculating a unique expression value percentage, and (ii) if the frequency ratio of the most common expression value to the frequency of the second most one was large. The function nearZeroVar in the R caret package version 6.0-86 with default values for cutoff was used[39]. The pre-processed gene expression data were checked and corrected for any batch effects and other sources of variation using SVA v.3.34.0 (Ref.[40]) and variancePartition v.1.19.17 (Ref.[41]) R packages. Correction was performed for sequencing lane batches for the adult cohort. No further confounding effects were detected.

Data discretization and feature selection

The expression values thereafter underwent data discretization. Equal frequency binning was used to discretize the data into three levels. To convert continuous features into discrete ones, equal frequency binning first sorts the values, and then divides them into equally sized bins.

Importantly, during the machine learning process, discretization is performed on the training set and then cuts are applied on the test set. Data discretization is an essential step in rough sets theory that is a basis in the R.ROSETTA algorithm. Subsequently, the Monte Carlo Feature Selection algorithm (MCFS; rmcfs R package v.1.2.5 [Ref.[42]]) was used for feature selection, as it is well-suited for data that have small numbers of samples but thousands of features (here; genes). This step is essential to reduce the noise in the data and rank the most important features for classification. The relative importance of a specific feature was assessed over multiple classification trees, which were built from randomly sampled training sets. Then the features were ranked based on their highest relative importance, which represents the feature's classification ability.

Optimizing number of selected features for rule-based learning

The number of significant features based on the cutoff methods used by MCFS highly differed, ranging from 0 to 600 features. In order to select the optimal number of significant features for building a predictive model for each dataset using the MCFS ranked list, iterative computational rounds were performed (referred to in **Supplemental Figure 10** as Feature Boosting), resulting in a minimum set of significant features required to distinguish between disease states (i.e. diagnosis and relapse). The features were incrementally added to build several rule-based models for each dataset (i.e. cohort), and the selected features that were used to build the model with the best overall accuracy were chosen for downstream analysis. The highest accuracy was gained for the models built with 50 features for the local adult cohort and 60 features for the local pediatric cohort (**Supplemental Figure 16**).

Constructing rule-based models

Rule-based models were built using the R.ROSETTA R package (v.2.2.9; ref.[43]). A rule-based model is a set of transparent IF-THEN rules calculated from reducts that are minimal subsets of features maintaining the indiscernibility[44]. For estimating reducts, R.ROSETTA may use several different algorithms, which from here-on are referred to as reducers. In this work, the Genetic and Johnson reducers[43] were applied. These reducers allow for estimating co-predictive rules and were successfully applied in previous studies[43,45-47]. In brief, the Genetic reducer is an evolutionary-based optimization algorithm, while the Johnson reducer is a greedy algorithm. The main difference between these algorithms is that the Genetic approach is stochastic, while Johnson is deterministic. Herein, the Genetic reducer was used in order to interpret the models for the local adult and pediatric cohorts (**Supplemental Tables 16** and **17**). Models used for

merging features and network comparisons were based on the Johnson reducer (**Supplemental Tables 18** and **19**). The performance of each rule generated depends on the number of samples in the dataset that support both the antecedent and the consequence of the rule (left-hand side and right-hand side, respectively), in addition to the accuracy of the rule. The rule accuracy is computed by dividing the number of objects that satisfy the consequence of the rule by the number of objects that satisfy the antecedent part of the rule, as further detailed by Garbulowski et al.[43]. The model performance was based on the total mean accuracy and the area under the receiver operating characteristic curve of the model generated from five-, and three-fold cross-validations for the adult and pediatric cohort, respectively (**Supplemental Figure 16**). Such a rule-based model from each dataset was used for further analysis and visualization.

Validating rule-based models

Model validation was performed by applying a permutation test. This was done by randomly shuffling the decision label (Diagnosis or Relapse) 1000 times for the decision table used to build the models utilized for network comparisons. Each time the shuffled decision table was used to build a rule-based model. The accuracies from each run were used to build a distribution. A threshold of 0.05 and confidence interval of 0.95 were used to determine the significance of the P-value. The mean and standard deviation and the standard error for the normal distribution were computed. The accuracy of the original model was compared to the mean (M) and standard error (SE) of the distribution as M±SE. If the accuracy of the original model was <M-SE or >M+SE, the P-value in this case was $P<0.05$ and proven to be significant.

Rule-based heat maps for evaluating classifiers

A binary matrix was constructed, where rules were oriented as the columns and samples as the rows. The matrix was used to cluster samples and construct a heat map using the pheatmap R package (v1.0.12; ref.[48]) based on the rules using asymmetric binary distance as a distance measure with hierarchical clustering. Hierarchical clustering on the support of each object per rule showed visible clusters of diagnosis and relapse samples for both the adult and pediatric models (**Supplemental Figures 17** and **18**).

Network-based comparisons and hubs visualization

To perform a comparison between cohorts, rule-based network structures were evaluated. Connection values of nodes were used for clustering of the decision classes for each dataset as proposed by Garbulowski et al.[45]. The node connection value reflects a co-predictive strength

of each node in the network. The clustering was performed on the most informative nodes, using Kendall rank correlation coefficient as a distance metric. Additionally, based on the clustering, topmost hubs were selected from the network and visualized as arc diagrams using the R package arcdiagram (v.0.1.12; ref.[49]). Here, an arc diagram displays a hub gene and all its associations estimated by the rule model. Importantly, a hub was defined as a gene with a high number of connections in the network that indicates appearance in a great number of rules. For each arc diagram values of nodes and arcs reflected values of nodes and edges from the network. In this study, by visualizing hubs, arc diagrams allowed for investigation of specific co-predictive mechanisms. To visualize hubs in the form of arc diagrams, an additional R function was implemented based on a VisuNet output. The function is attached to the VisuNet R package and publicly available at https://github.com/komorowskilab/VisuNet. To the best of our knowledge, we show the application of arc diagrams as a novel way of visualizing selected fragments of rule-based networks.

Tumor purity assessment

Next-generation sequencing-based tumor purity assessment was manually performed based on patient-matched genomic material[2] that was extracted together with the RNA from the respective samples (**Supplemental Tables 1** and **2**). Purity for whole genome sequenced samples were assessed manually based on their available somatic genomic aberrations, and how the relative effect on sequence coverage (deletion from two to one copy), the allele ratio of heterozygous SNPs (in regions with copy-neutral loss-of-heterozygosity) and somatic SNVs (in diploid regions) would theoretically scale with tumor purity. The estimated purity for whole exome sequenced samples was based on solely somatic SNVs present in diploid regions. The next-generation sequencing-based purity results were further compared to morphology-based purity assessment information on May Grünwald and Giemsa stained cells post cryopreservation and, if applicable, post immune-based depletion of non-tumor cells (**Supplemental Table 2**).

Calculation and visualization of statistical significance

The following statistical calculations were performed using Qlucore omics explorer v.3.6 with default settings, if not otherwise specified. Principal Component Analysis (PCA; according to ref.[50-52]) and t-Distributed Stochastic Neighbor Embedding (t-SNE; according to ref.[53,54]) were used to visualize the high-dimensional and unsupervised data, after centering and scaling the variables to zero mean and unit variance (mean=0, variance=1). Hierarchical clustering was

performed using log2-transformed normalized values following the Euclidean metric on normalized variables (mean=0, variance=1). Hierarchical clustering and associated heat maps were constructed following ref.[55]. Further, genes were ranked according to their R/R^2-statistic values (R-statistics; **Supplemental Tables 11** and **14**), which Qlucore computes according to the coefficient of partial determination. Volcano plots were used to identify the highest ranked differentially expressed genes among sample groups. Venn diagrams were utilized to inspect the intersection of differentially expressed genes between the cohorts. Qlucore applies the Benjamini-Hochberg[37] method to correct for multiple testing and calculates the fold change (FC) from the difference between the arithmetic averages over each group.

In order to calculate significant differences observed in gene expression levels between two or more groups, GraphPad Prism version 7.02 and 9.0.2 were used and results were visualized in the form of scatter plots with mean and standard deviation, or spaghetti plots. First, normality was tested for sample groups containing more than 30 values following the D´Agostino & Pearson calculation. Next, one of the following statistical tests was performed: (i) Unpaired t-test for a two-group comparison on dichotomous variables; (ii) Mann-Whitney test for a two-group comparison on non-parametric data; (iii) Kruskal-Wallis test for multi-group comparisons on non-parametric data followed by Dunn´s correction for multi-group comparisons; or (iv) Wilcoxon matched-pairs significant rank test for a patient-matched two-group comparison on non-parametric data. All P-values are given for two-sided tests. Kaplan-Meier plots were used to visualize EFS and 5-year overall survival rates, and putative differences between low- and high-expression of the respective genes were calculated using Log-rank (Mantel-Cox) test (**Supplemental Table 13**).

Sample usage for various analyses

Different sample sets were exploited in the respective analyses, in order to maximize the number of samples included for each investigation. Detailed information regarding samples included in each of the sub-groups is present in **Supplemental Table 10**.

Supplemental Results

Genomic and transcriptomic landscape of R/PR AML

We previously analyzed the composition of genomic alterations via WGS or WES for all 122 leukemia samples included in this study[2]. As a means to investigate the potential of using RNA-seq as an alternative to WGS/WES with regards to detection of SNVs and small InDels, we examined the overlap between alterations identified at the DNA- and RNA levels (**Supplemental Tables 7** and **8**). Thirty-nine percent (715/1841) of somatic protein-coding SNVs and small InDels detected at the genomic level were located in regions not being expressed at a sufficient level (<3 reads), and could thus not confidently be identified by RNA-seq (**Supplemental Figure 1**). Out of the remaining 1126 variants, 117 genomic variants (10.4%) could not be validated at the transcriptomic level, including sub-clonal variants in *GATA2, KMT2A* and *NF1,* as well as clonal frameshift variants in *TP53* and *WT1*. The variants not validated at the RNA-level showed a lower median variant allele frequency at the DNA level (VAF_{DNA}) compared to variants identified also by RNA-seq (median VAF_{DNA}: 0.32 and 0.46, respectively). Approximately half (n=535) of the remaining 1009 genomic variants were correctly called by HaplotypeCaller (GATK[13]) at the RNA level, while the rest were identified by manual inspection of the corresponding region in the RNA-seq reads utilizing Integrative Genomics Viewer (IGV[23]). The main difference between these two groups was a lower frequency of small InDels among the variants reported by HaplotypeCaller compared to those solely detected via IGV (11.4% and 37.8%, respectively). Nonsense- and frameshift variants with a premature stop codon, potentially resulting in nonsense-mediated RNA decay, were as expected more often not detected at the transcriptomic level. We found 17 additional somatic protein-coding, non-recurrent variants at the RNA level that were not reported by WGS/WES, all of which belonged to one of the following three groups: 1) variant region not covered by WES (n=6), 2) missed by WES variant calling (n=8), or 3) mitochondrial variant (n=3; filtered out due to criteria regarding normal DNA controls, not customized to the in general very high coverage seen for mitochondrial genes in WGS/WES data).

This analysis revealed the feasibility of RNA-seq to determine the mutational status of R/PR AML. Nevertheless, accurate variant detection by RNA-seq requires deeper coverage as well as improvement in variant calling from RNA-seq data to confidently perform classification and prognostication based on that data.

Supplemental Table legends

Supplemental Tables 1-20 are present in a separate Supplemental document, including a content list, abbreviations and a legend regarding color coding on the first sheet of that document.

Supplemental Table 1: Study cohort sample overview. Summary of patient-matched longitudinal AML samples and normal control samples included in the study cohort, as well as overlapping genomic sequencing methods performed for the respective samples. BMS, Bone marrow derived stromal cells; CR, Complete remission; D, Diagnosis; G, Genomic data availability; PR, Primary resistant; R1/2/3, Sequential relapses; R1/2-P, Persistent relapse sample; T, Transcriptomic data availability; WES, Whole exome sequencing; WGS 30X, Whole genome sequencing, aiming at >30X coverage; WGS 90X, Whole genome sequencing, aiming at >90X coverage.

Supplemental Table 2: Study cohort sample characteristics. Detailed information regarding the characteristics of the 122 tumor samples and five healthy control samples included in the current study. Sample purity and cell viability are given in intervals of 10 and 25%, respectively. BM, Bone marrow; D, Diagnosis; HSCT, Hematopoietic stem cell transplantation; NOPHO, Nordic Society of Paediatric Haematology and Oncology; PB, Peripheral blood; PR, Primary resistant; R1/2/3, Sequential relapses; R1/2-P, Persistent relapse sample; RIN, RNA integrity number (Agilent Tape Station); U-CAN, Uppsala Umeå Comprehensive Cancer Consortium, Sweden; WES, Whole exome sequencing; WGS 30X, Whole genome sequencing, aiming at >30X coverage; WGS 90X, Whole genome sequencing, aiming at >90X coverage.

Supplemental Table 3: Clinical information. Summary of the clinical characteristics of patients included in the local AML study cohort. Allo, Allogeneic; Auto, Autologous; CR, Complete remission; D, Diagnosis; F, Female; HSCT, Hematopoietic stem cell transplantation; M, Male; M0-M7, The French-American-British (FAB) classification of AML; MDS, Myelodysplastic syndromes; NOS, Not otherwise specified; R1/2/3, Sequential relapses; t-AML, Treatment related AML; WBC, White blood cell count; VP, Current treatment protocol.

Supplemental Table 4: Characteristics of CD34+ BM-control samples. Summary of the characteristics of CD34+ BM samples from five individual healthy donors used as normal counterparts throughout the study. Cell viability is given in intervals of 25%. CD34+ BM cells

were purchased through AllCells Inc via Nordic BioSite. BM, Bone marrow; F, Female; M, Male.

Supplemental Table 5: Antibody information. Antibodies (BD Biosciences, San Jose, CA, USA; Thermo Fisher Scientific, Waltham, MA, USA) used for purification of patient-derived AML samples by immune-based depletion of non-tumor cells.

Supplemental Table 6: RNA-seq statistics. Sequencing statistics and QualiMap results for samples analyzed by RNA-seq (Adult cohort: Illumina HiSeq2500; Pediatric cohort: NovaSeq6000). D, Diagnosis; PR, Primary resistant; R1/2/3, Sequential relapses; R1/2-P, Persistent relapse sample.

Supplemental Table 7: SNVs and small InDels detected by RNA-seq. Manually curated somatic single nucleotide variants (SNVs) and small insertions and deletions (InDels; <50bp) derived from the RNA-seq analysis (HaplotypeCaller) and the overlap with analysis on whole genome- and whole exome sequenced samples. D, Diagnosis; PR, Primary resistant; R1/2/3, Sequential relapses; R1/2-P, Persistent relapse sample; VAF, Variant allele frequency; WES, Whole exome sequencing; WGS, Whole genome sequencing.

Supplemental Table 8: Comprised metadata and RNA-seq- and WGS/WES results. Various sample- and clinical information, combined with data regarding variants recurrently identified in the R/PR cohort, on a per sample basis. Both transcriptomic and genomic results are overlaid. CNA, Copy number alteration; CN-LOH, Copy-neutral Loss-of-heterozygosity; D, Diagnosis; F, Female; FAB, The French-American-British classification of AML; InDel, Insertion or Deletion mutation; M, Male; N/A, Data not available; PR, Primary resistant; R1/2/3, Sequential relapses; R-P, Persistent relapse sample; WES, Whole exome sequencing; WGS-30X, Whole genome sequencing, aiming at >30X coverage; WGS-90X, Whole genome sequencing, aiming at >90X coverage.

Supplemental Table 9: Fusion transcripts in R/PR AML. Summary of manually curated fusion transcripts detected by StarFusion. Chr, Chromosome; D, Diagnosis; FFPM, Fusion fragments per million total reads; gPCR; Polymerase chain reaction on genomic DNA; INV, Inversion; PR, Primary resistant; R1/2/3, Sequential relapses; R1/2-P, Persistent relapse sample; WES, Whole exome sequencing; WGS, Whole genome sequencing.

Supplemental Table 10: Sample usage for generation of various analyses. Tumor samples included for various analyses. (**A**) Differential gene expression analysis between samples associated with short versus long EFS for the local cohort. (**B-D**) Diagnosis versus relapse samples for the local cohort (**B** and **C**), and the TARGET cohort (**D**). (**E-F**) Diagnosis samples used to generate Kaplan-Meier plots for EFS and OS assessments for the TARGET (**E**) and TCGA (**F**) cohorts. D, Diagnosis; EFS, Event-free survival; OS, Overall survival; PR, Primary resistant; R1/2/3, Sequential relapses; R-P, Persistent relapse sample; TARGET, Therapeutically Applicable Research to Generate Effective Treatments; TCGA, The Cancer Genome Atlas.

Supplemental Table 11: DEGs associated with short vs. long EFS. Summarized are all differentially expressed protein-coding genes between samples associated with short versus long EFS with a P-value <0.05. A fold change >1.0 indicates up-regulation in samples associated with short EFS, while a fold change <1.0 indicates down-regulation. Only diagnosis samples were included in the comparison, and adult and pediatric cases were combined. See **Supplemental Table 10A** for details regarding samples included for generating the data in this table. DEGs, Differentially expressed genes; EFS, Event-free survival. FDR, False discovery rate; R-statistic, Square roots of the R2-statistics (coefficient of determination).

Supplemental Table 12: GO-analysis of DEGs between short vs. long EFS-associated samples. Detailed results of the GO enrichment analysis, utilizing GOrilla, for samples associated with short versus long EFS. DEGs with a P-value <0.05 and a |log2 fold change| >1 were analyzed against the background of all expressed protein-coding genes included in this study. B, Total number of genes associated with a specific GO term; b, Number of genes in the intersection; DEG, Differentially expressed gene; EFS, Event-free survival; FDR, False discovery rate; GO, Gene ontology; N, Total number of protein-coding genes (background list); n, Number of genes in the target set.

Supplemental Table 13: Statistics associated with survival analyses. Survival analysis-associated statistics for Kaplan-Meier plots depicting EFS and overall survival with regards to low versus high expression of *CD6, GLI2*, *IL1R1, INSR, ST18,* and *ZNF773*. Local, TARGET and TCGA refer to the respective analyzed cohort. EFS, Event-free survival; OS, Overall survival.

Supplemental Table 14: DEGs between patient-matched diagnosis and relapse samples. Summarized are all differentially expressed protein-coding genes between diagnosis and relapse samples with a P-value <0.05. A fold change >1.0 indicates up-regulation at diagnosis compared to relapse, while a fold change <1.0 indicates down-regulation. Only patient-matched diagnosis and relapse samples were included in the comparison (adult: n=22 diagnosis-relapse pairs; pediatric: n=17 diagnosis-relapse pairs). See **Supplemental Table 10B** for details regarding samples included for generating the data in this Table. DEGs, Differentially expressed genes; FDR, False discovery rate; R-statistic, Square roots of the R2-statistics (coefficient of determination).

Supplemental Table 15: GO-analysis of DEGs between patient-matched diagnosis and relapse samples. Detailed results of the GO enrichment analysis, utilizing GOrilla, for diagnosis versus relapse samples. DEGs with a P-value <0.05 and a |log2 fold change| >1 were analyzed against the background of all protein coding expressed genes included in this study. B, Total number of genes associated with a specific GO term; b, Number of genes in the intersection; DEG, Differentially expressed gene; FDR, False discovery rate; GO, Gene ontology; N, Total number of protein-coding genes (background list); n, Number of genes in the target set.

Supplemental Tables 16-19: Machine learning model rules for diagnosis and relapse in various cohorts and comparisons. The respective table describes the generalized rules output from training the machine learning model on the respective cohort samples to differentiate between diagnosis and relapse. The features column describes the left-hand side of the rule (LHS; The genes), the levels describe the discretized gene expression value for each feature in the rule (1, low; 2, medium; 3, high), and the decision is the right-hand side (RHS) or the consequence of the rule. The rules quality is represented mainly by SupportLHS and SupportRHS, which represent the number of samples supporting the rule, the rule accuracy and the P-value of the rule.

Supplemental Table 16: Machine learning model rules for diagnosis and relapse in adult AML. The table describes the generalized rules output from training the machine learning model on the adult AML cohort samples to differentiate between diagnosis and relapse. The Genetic reducer was applied. See **Supplemental Table 10C-upper** for details regarding samples included for generating the data in this table.

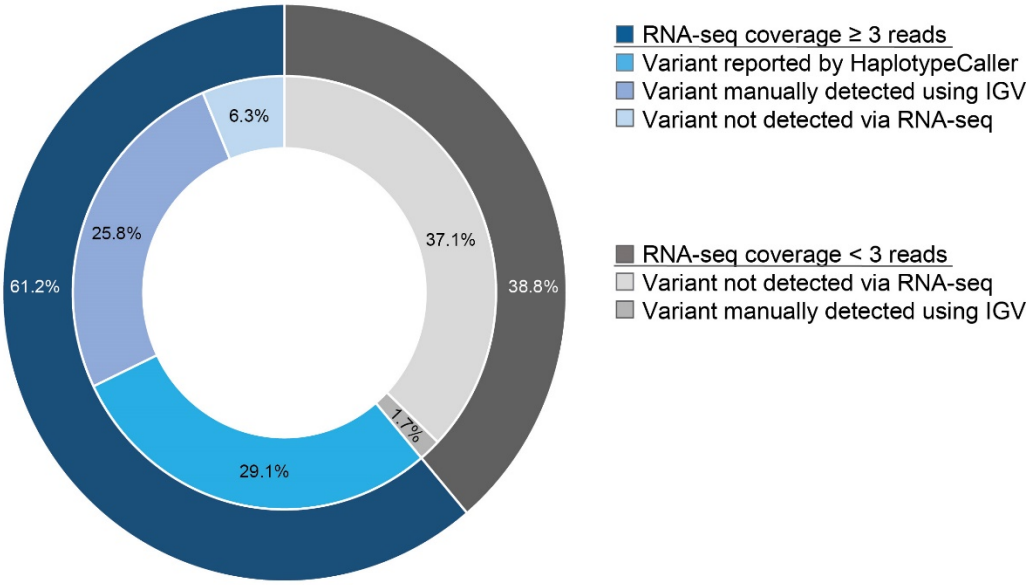
Supplemental Table 17: Machine learning model rules for diagnosis and relapse in pediatric AML. The table describes the generalized rules output from training the machine learning model on the local pediatric AML cohort samples to differentiate between diagnosis and relapse. The Genetic reducer was applied. See **Supplemental Table 10C-lower** for details regarding samples included for generating the data in this table.

Supplemental Table 18: Machine learning model rules for diagnosis and relapse in pediatric AML (features merged with TARGET). The table describes the generalized rules output from training the machine learning model on samples from the local pediatric AML cohort using the merged MCFS features from both the local pediatric cohort and the TARGET cohort to differentiate between diagnosis and relapse in the local pediatric cohort. The Johnson reducer was applied. See **Supplemental Table 10C-lower** and **D** for details regarding samples included for generating the data in this table.

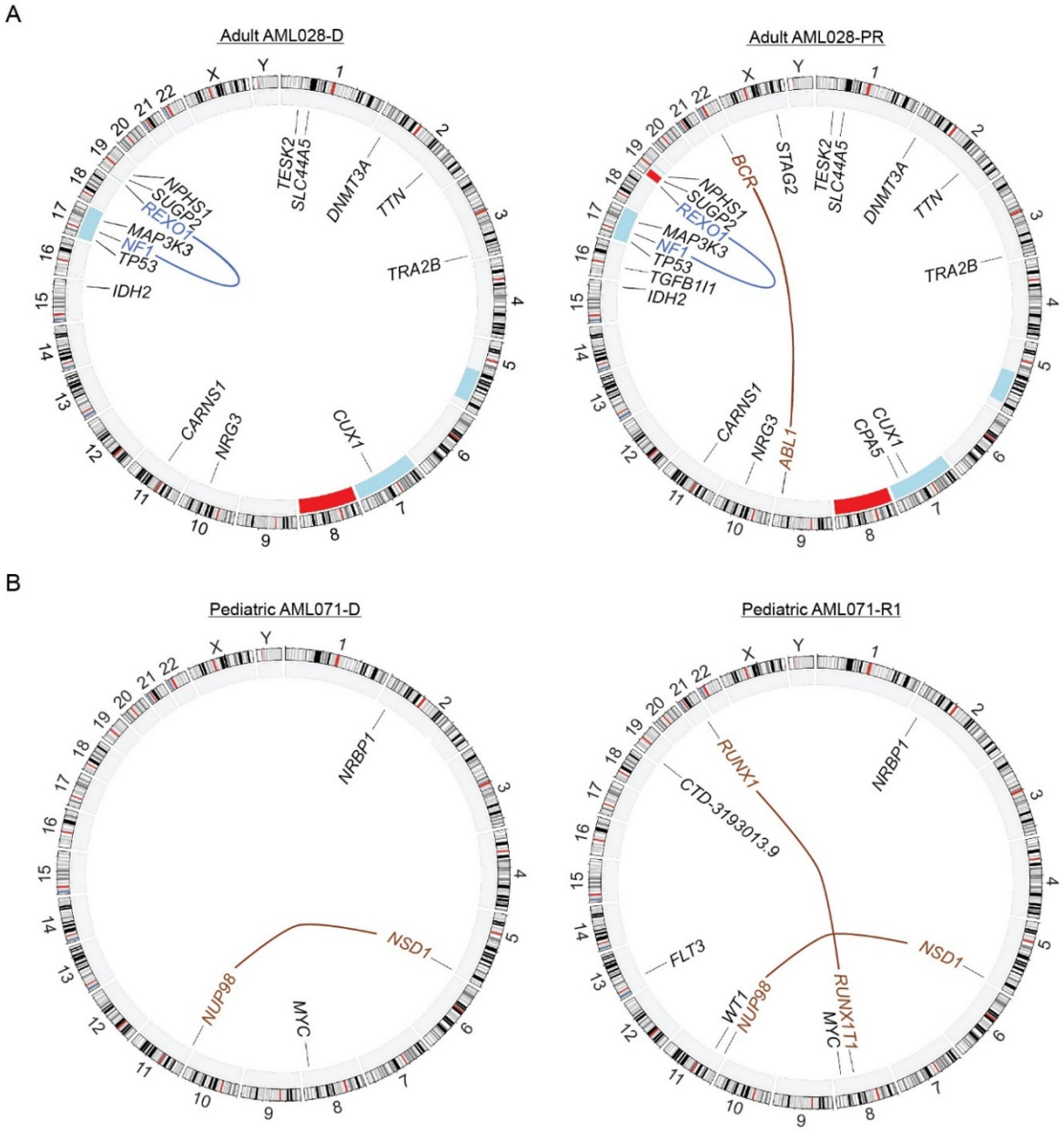
Supplemental Table 19: Machine learning model rules for diagnosis and relapse in the TARGET AML cohort (features merged with Local pediatric). The table describes the generalized rules output from training the machine learning model on samples from the TARGET cohort using the merged MCFS features from both the local pediatric AML cohort and the TARGET cohort to differentiate between diagnosis and relapse in the TARGET cohort. The Johnson reducer was applied. See **Supplemental Table 10C-lower** and **D** for details regarding samples included for generating the data in this table.

Supplemental Table 20: Verification of transcriptomic fusion events and associated primer information. Detailed information of primers used for technical verification of putative fusion transcripts. D, Diagnosis; ITD, Internal tandem duplication; PR, Primary resistant; R1, Sequential relapses; R1-P, Persistent relapse sample; RT-PCR, Reverse transcriptase polymerase chain reaction; WT, wild type.

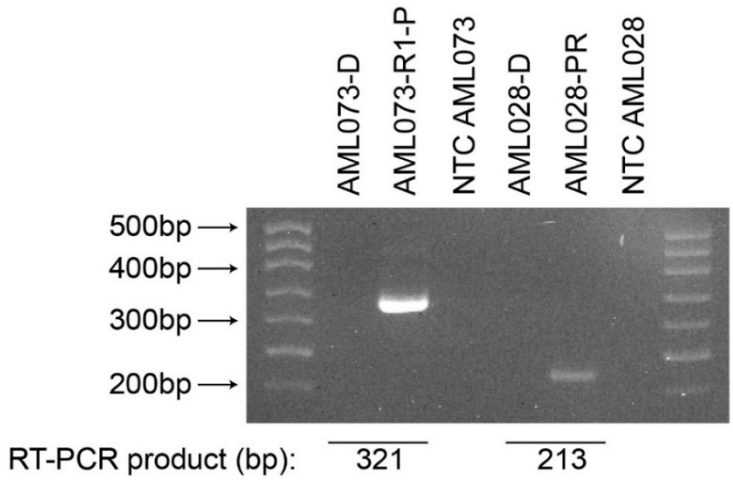
Supplemental Figures



Supplemental Figure 1. Detection of SNVs and small InDels by RNA-seq. The overlap of protein-coding variants detected via WGS/WES and RNA-seq is shown. Variants with a minimum coverage of three reads surrounding the variant location based on RNA-seq are highlighted in blue color, and the remaining variants are shown in grey. IGV[23], Integrative genomics viewer; InDel, Insertion or deletion mutation; SNV, Single nucleotide variant; WES, Whole exome sequencing; WGS, Whole genome sequencing; .

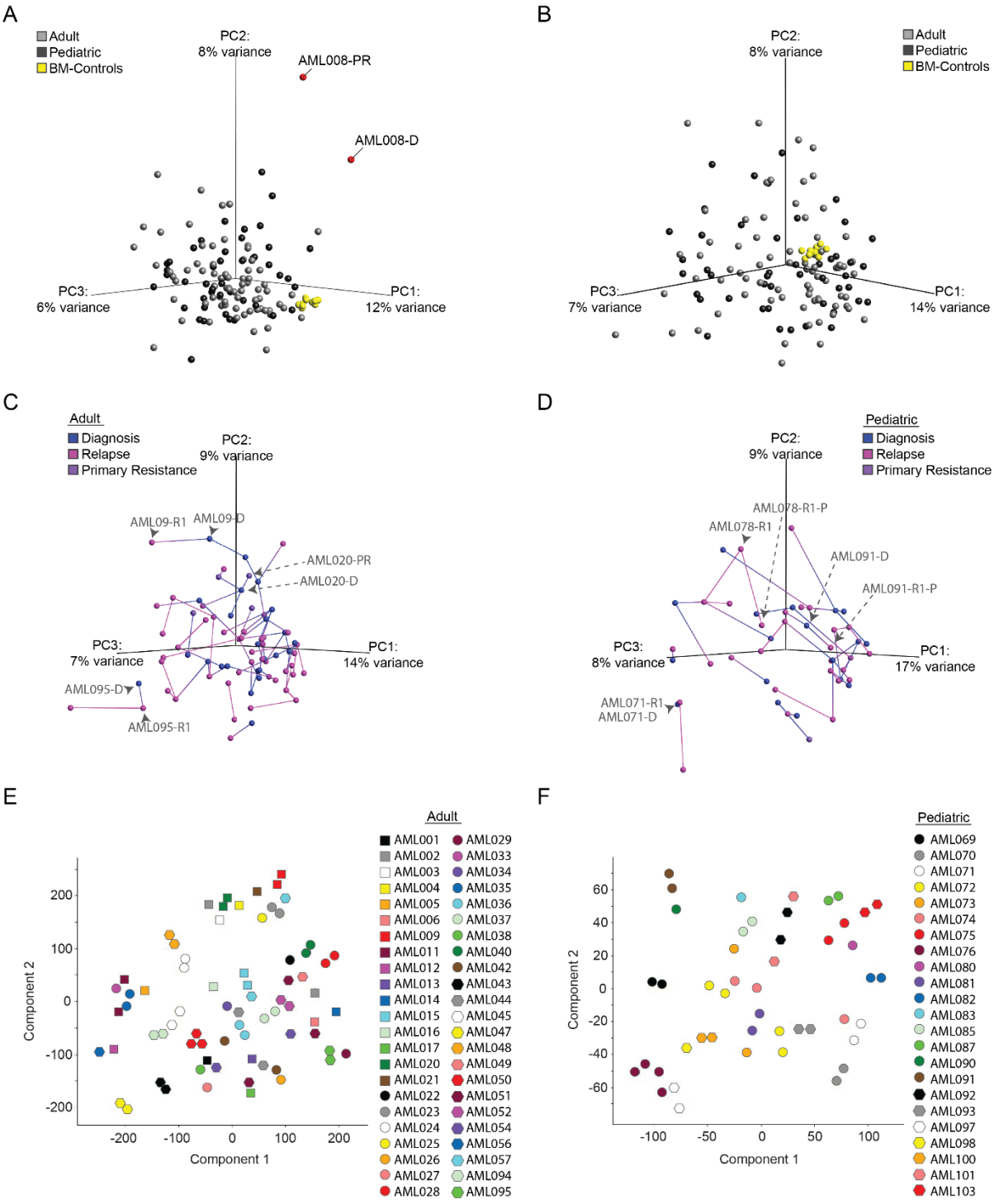


611 **Supplemental Figure 2. Changes in the mutational landscape during AML progression.**
612 **A-B)** Circos plots of sequential AML samples represented by the adult case AML028 (**A**), with
613 samples from diagnosis and primary resistance, as well as the pediatric case AML071 (**B**), with
614 samples at diagnosis and first relapse. Copy number alterations, derived by WGS analysis, are
615 plotted in the inner ring, including heterozygous deletions (light blue) and amplifications (red).
616 Genes affected by sequence mutations (black; data derived from WGS- and RNA-seq analysis)
617 and gene fusion events (brown for in-frame fusions and blue for frameshift fusions; data derived
618 from RNA-seq analysis) are plotted inside the circle. The outer ring depicts chromosome
619 idiograms (based on Circos package data UCSC.hg19.chr), with the chromosomal numbers
620 indicated. D, Diagnosis; PR, Primary resistant; R1, Relapse 1; WGS, Whole genome
621 sequencing.



622

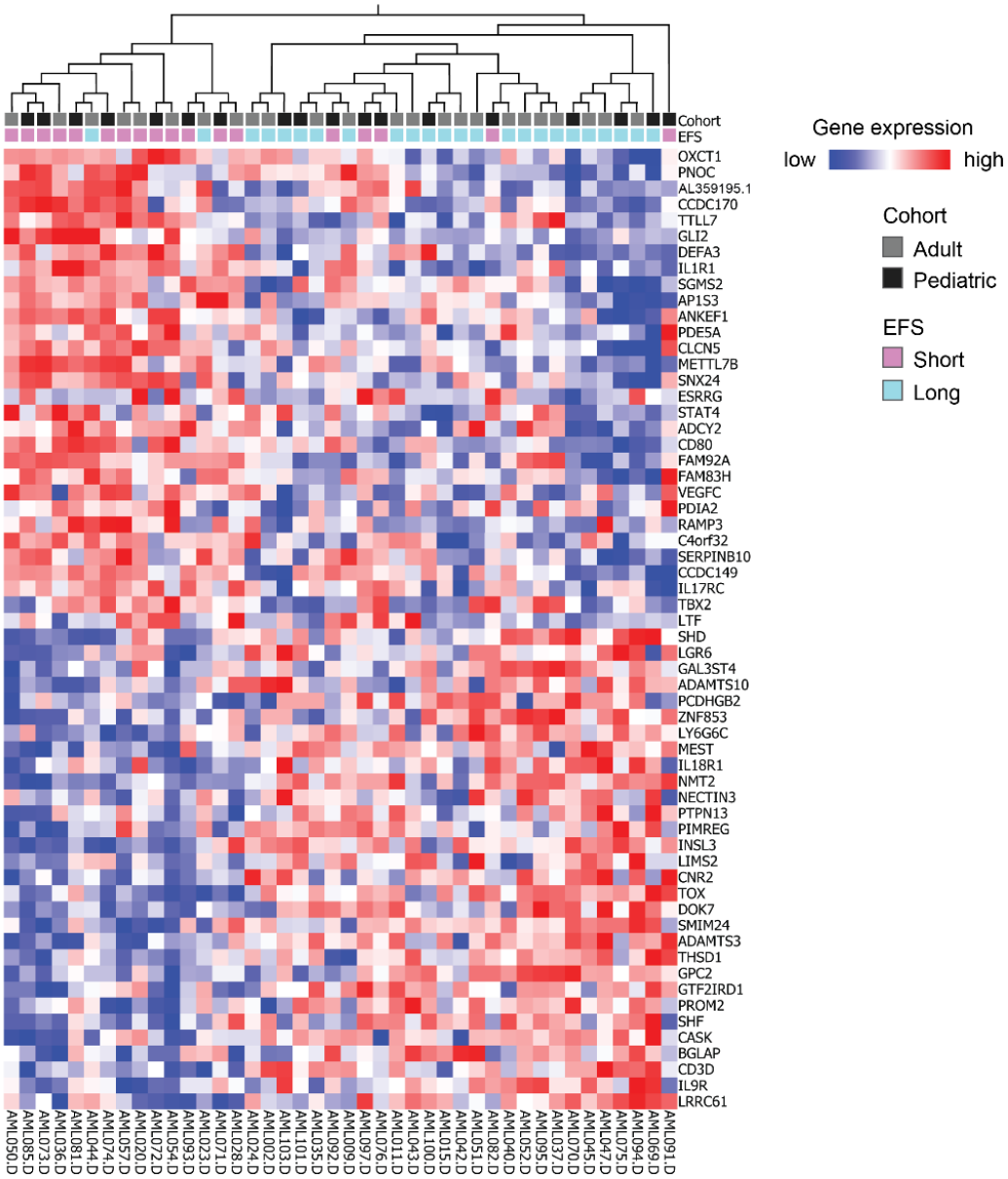
623 **Supplemental Figure 3. *BCR-ABL1* fusions in treatment persistent AML.** RT-PCR for
624 investigation of the presence of *BCR-ABL1* fusion transcripts in R/PR AML, with primers used
625 listed in **Supplemental Table 20**. bp, Base pairs; D, Diagnosis; NTC, Non-template PCR
626 control; PR, Primary resistant; R1-P, Persistent relapse sample.



627

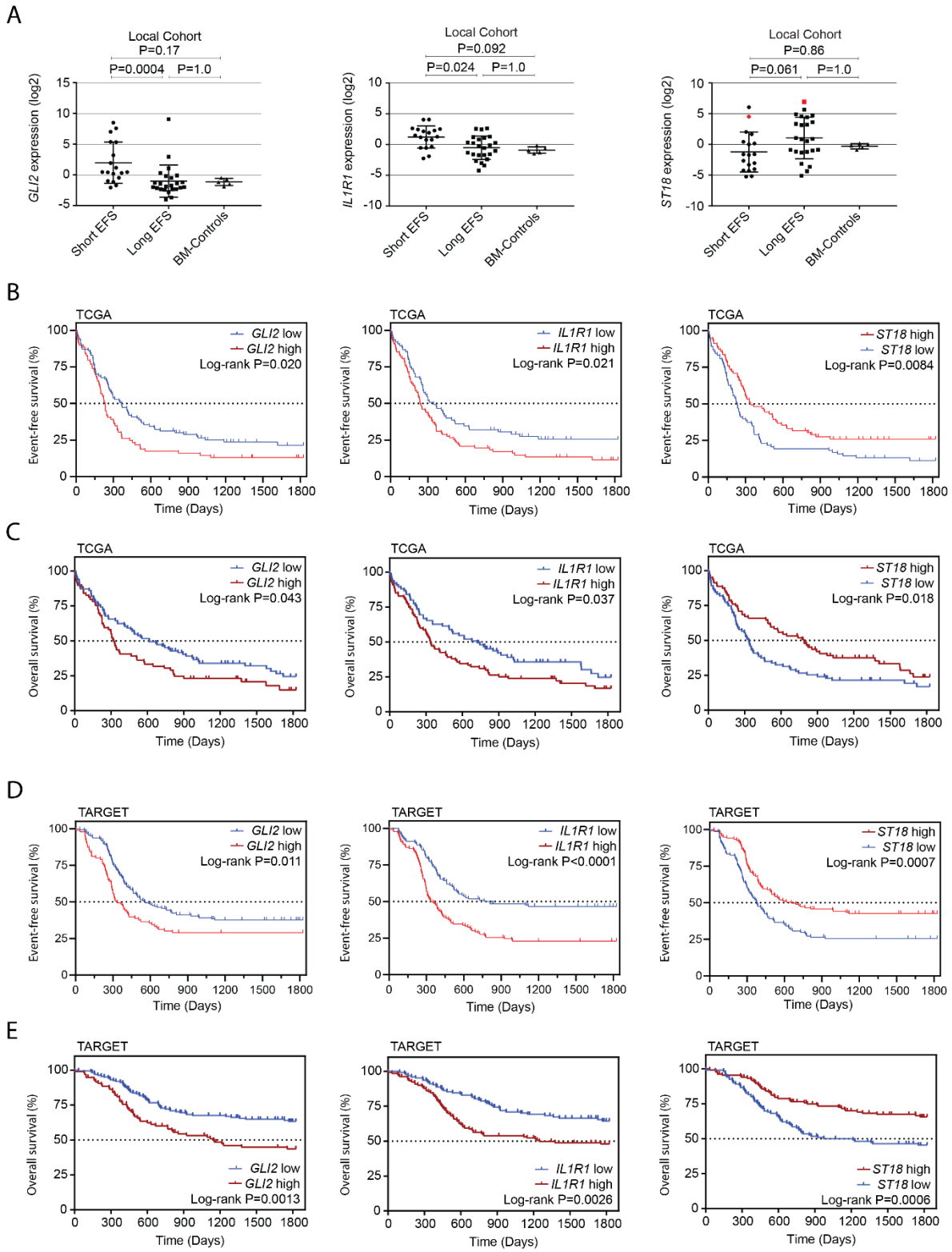
628 **Supplemental Figure 4. Unsupervised clustering of the R/PR AML cohort. A)**
629 Identification of sequential tumor samples from case AML008 as outliers via principal
630 component analysis (PCA). **B)** PCA analysis of all combined adult (light gray) and pediatric
631 (dark gray) samples, post exclusion of AML008-D/PR, with a distinct cluster of BM-control
632 samples (yellow; including two technical replicates each). **C-F)** Neighboring information in the
633 form of PCA plots and T-distributed stochastic neighbor embedding plots (t-SNE;
634 perplexity=7) showing similarity between sequential patient-matched tumor samples in adults

(**C** and **E**) and children (**D** and **F**). Visualization and underlying calculations were performed using Qlucore omics explorer v.3.6. BM-controls, normal CD34 expressing bone marrow cell control samples; D, Diagnosis; PR, Primary resistant; R1, First relapse; R1-P, Treatment persistent first relapse.



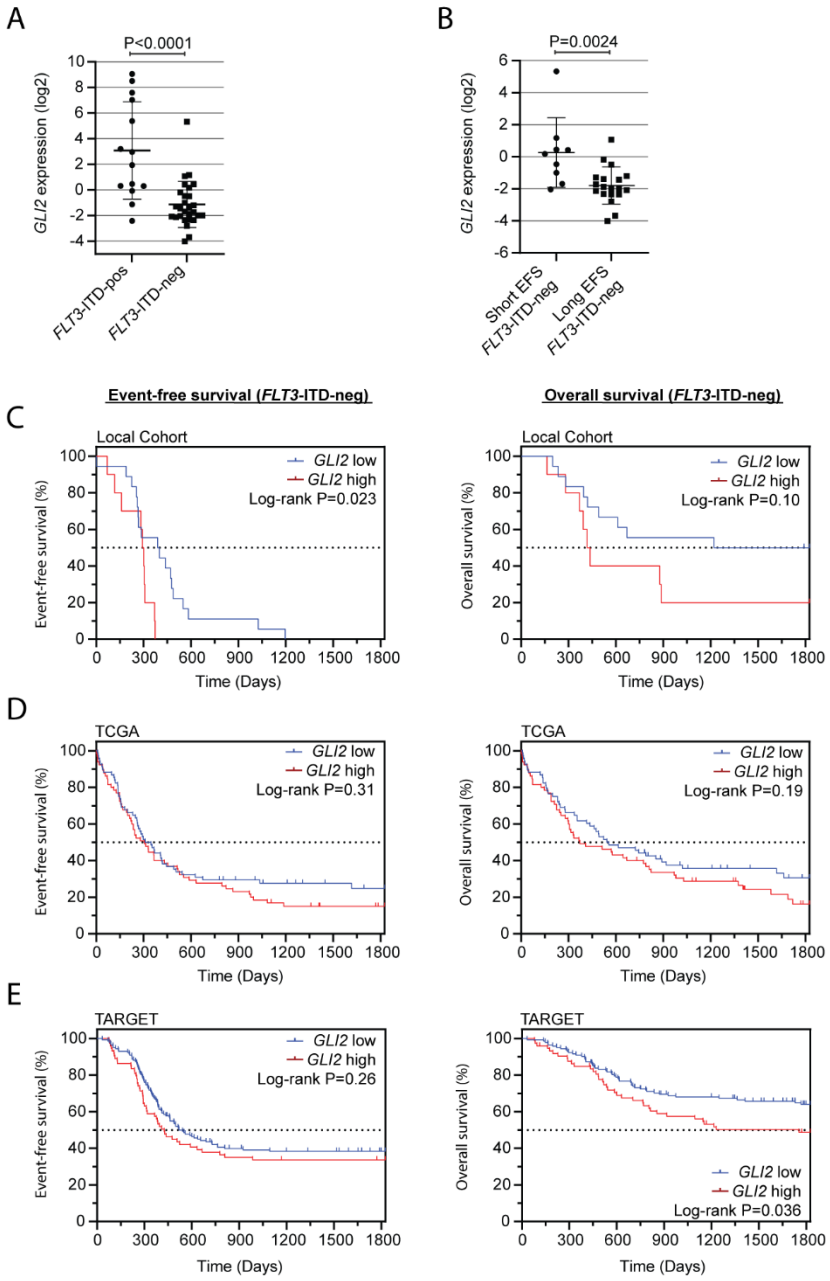
639

640 **Supplemental Figure 5. Differential gene expression between short versus long EFS-**
641 **associated samples.** Heat map and hierarchical cluster analysis of the top 60 differentially
642 expressed genes (DEGs; corresponding to: P ≤0.0071 and |FC|≥2.5) between samples
643 associated with short versus long EFS. Adult and pediatric diagnosis samples are combined (X-
644 axis). Genes are ranked according to their R-statistic values (Y-axis), with upregulated genes
645 depicted in red and downregulated genes in blue. Visualization and underlying statistical
646 calculations were performed using Qlucore omics explorer v.3.6. EFS, Event-free survival; FC,
647 Fold change. See **Supplemental Table 10A** for details regarding samples included in this
648 figure, and **Supplemental Table 11** for all DEGs.



Supplemental Figure 6. Expression levels of *GLI2*, *IL1R1* and *ST18* correlate with outcome in adult and pediatric AML. A) Scatter plots with mean and SD presenting gene expression data of *GLI2*, *IL1R1* and *ST18* comparing samples associated with short versus long EFS within the local cohort. Samples highlighted in red for *ST18* harbor an inversion on

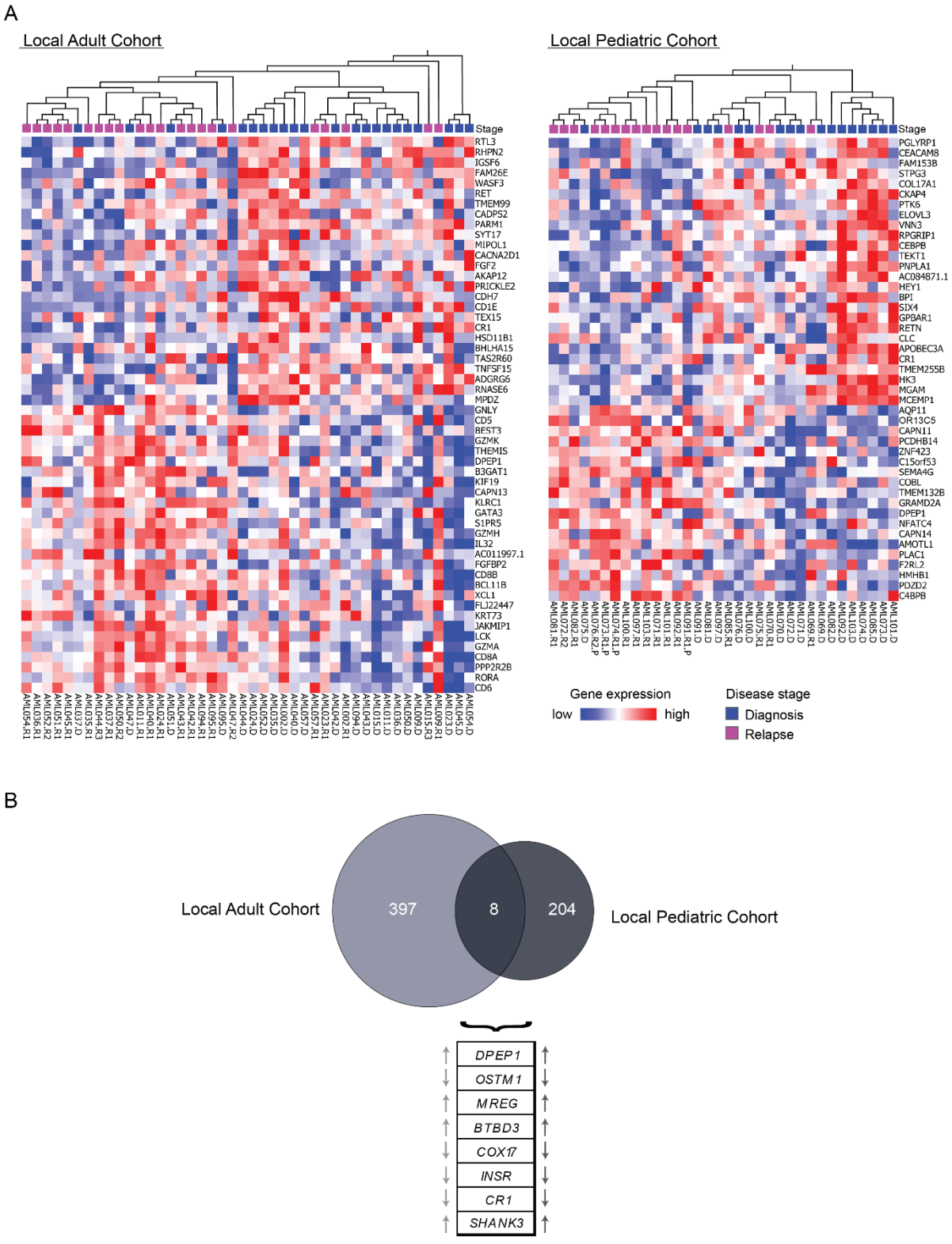
chromosome 16, leading to a *CBFB-MYH11* gene fusion. The expression values for the BM-control samples are given as the average of two technical replicates. Applied statistical test: Kruskal-Wallis test followed by Dunn´s correction for multi-group comparisons. The Y-axis represents log2 transformed, TMM normalized expression of mRNA. **B**-**E**) Kaplan-Meier plots depicting EFS (**B** and **D**) and overall survival (**C** and **E**), comparing high and low *GLI2*, *IL1R1* and *ST18* expression within the TCGA (**B** and **C**) and TARGET (**D** and **E**) cohorts. High and low gene expression was discretized based on the mean expression for the respective gene over all samples included in the analysis. Applied statistical test: Log-rank (Mantel-Cox) test. EFS, Event-free survival; SD, Standard deviation; TMM, Trimmed mean of M-values. See **Supplemental Table 10A** and **E-F** for details regarding samples included in this figure, and **Supplemental Table 13** for accompanied statistical results.



665

666 **Supplemental Figure 7. Elevated *GLI2* expression is associated with poor outcome**
667 **independent of *FLT3*-ITD status. A-B**) Scatter plots with mean and SD depicting *GLI2*
668 normalized expression values within the local cohort of (**A**) samples according to their *FLT3*-
669 ITD mutational status, and (**B**) comparing *FLT3*-ITD-negative samples associated with short
670 versus long EFS. Applied statistical test: Mann-Whitney test. The Y-axis represents log2
671 transformed, TMM normalized mRNA expression. **C-E**) Kaplan-Meier plots presenting the
672 EFS and overall survival, comparing high and low *GLI2* expression in *FLT3*-ITD-negative
673 cases within the local (**C**), TCGA (**D**) and TARGET (**E**) cohorts. High and low gene expression
674 was discretized based on the mean expression over all samples included in the analysis. Applied
675 statistical test: Log-rank (Mantel-Cox) test. EFS; event-free survival; *FLT3*-ITD, *FLT3* internal

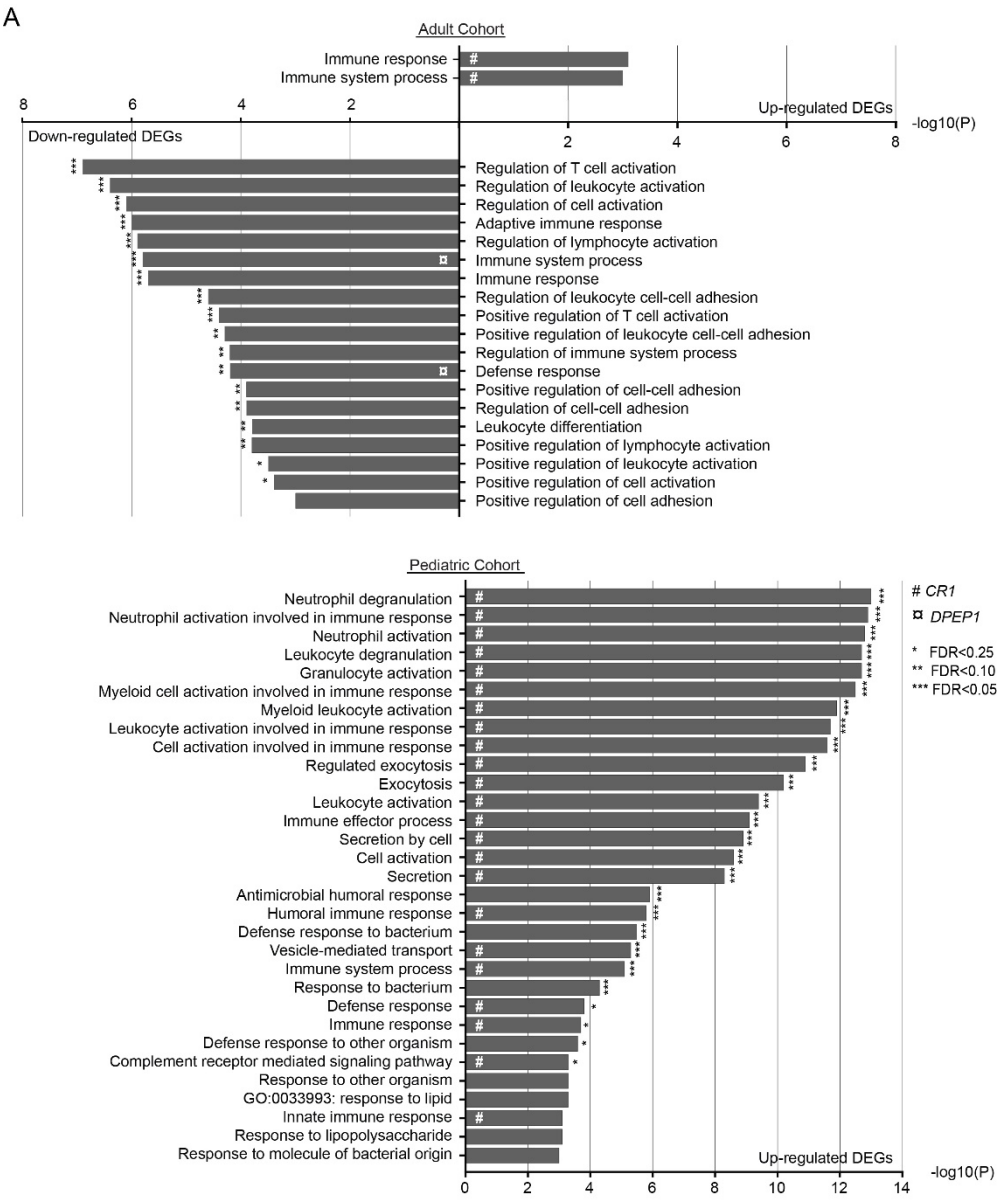
676 tandem duplication; neg, negative; pos, positive; SD, Standard deviation; TMM, Trimmed
677 mean of M-values.

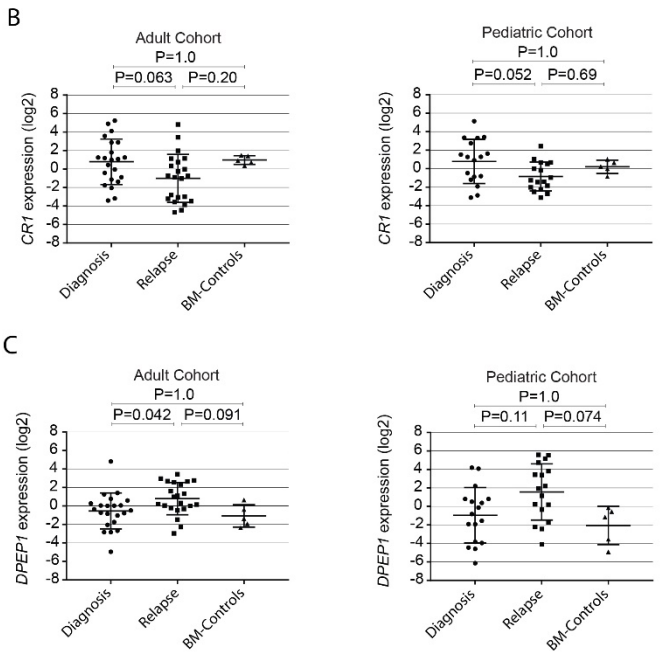


Supplemental Figure 8. Differential gene expression between paired diagnosis and relapse samples. A) Heat maps and hierarchical cluster analysis of the top ranked DEGs between paired diagnosis and relapse samples for adult (P≤0.03, |FC|≥2; n=54 genes) and pediatric (P≤0.03, |FC|≥2; n=45 genes) samples (X-axis). Genes are ranked according to their R-statistic values

(Y-axis), with upregulated genes depicted in red and downregulated genes in blue. **B**) Venn diagram of diagnosis versus relapse specific DEGs in the local adult and pediatric cohorts. Included are all DEGs with a P-value <0.05, independent of their respective fold change. Genes that form the intersection between the adult and pediatric cohorts are indicated, with arrows depicting directionality of expression at relapse. Visualization and underlying statistical calculations were performed using Qlucore omics explorer v.3.6. D, Diagnosis; DEGs, Differentially expressed genes; FC, Fold change; R1/2/3, Relapse 1/2/3; R1/2-P, Persistent relapse sample. See **Supplemental Table 10B** for details regarding samples included in this figure, and **Supplemental Table 14** for all DEGs.

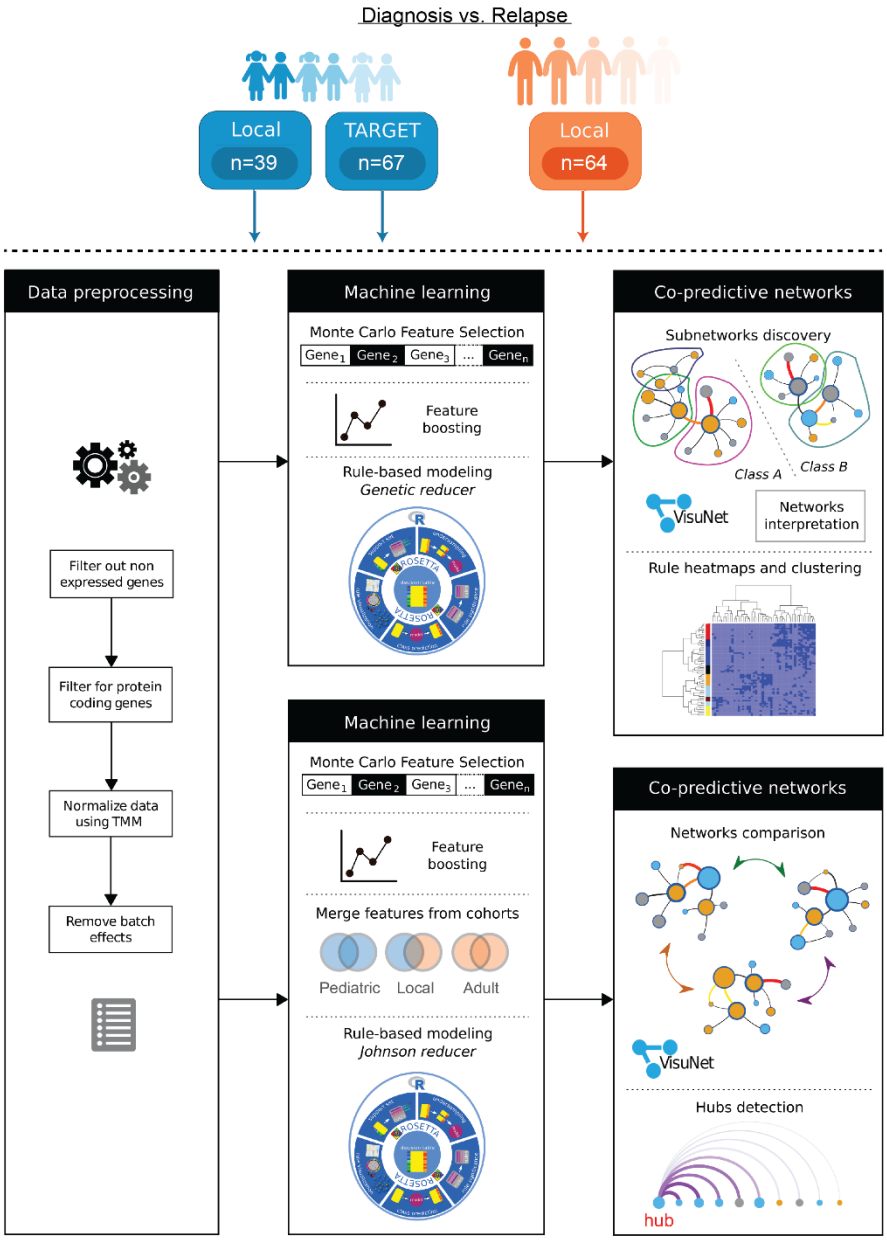
692


A
Adult Cohort
Immune response
Immune system process
Down-regulated DEGs
Up-regulated DEGs
-log10(P)
Regulation of T cell activation
Regulation of leukocyte activation
Regulation of cell activation
Adaptive immune response
Regulation of lymphocyte activation
Immune system process
Immune response
Regulation of leukocyte cell-cell adhesion
Positive regulation of T cell activation
Positive regulation of leukocyte cell-cell adhesion
Regulation of immune system process
Defense response
Positive regulation of cell-cell adhesion
Regulation of cell-cell adhesion
Leukocyte differentiation
Positive regulation of lymphocyte activation
Positive regulation of leukocyte activation
Positive regulation of cell activation
Positive regulation of cell adhesion
Pediatric Cohort
Neutrophil degranulation
Neutrophil activation involved in immune response
Neutrophil activation
Leukocyte degranulation
Granulocyte activation
Myeloid cell activation involved in immune response
Myeloid leukocyte activation
Leukocyte activation involved in immune response
Cell activation involved in immune response
Regulated exocytosis
Exocytosis
Leukocyte activation
Immune effector process
Secretion by cell
Cell activation
Secretion
Antimicrobial humoral response
Humoral immune response
Defense response to bacterium
Vesicle-mediated transport
Immune system process
Response to bacterium
Defense response
Immune response
Defense response to other organism
Complement receptor mediated signaling pathway
Response to other organism
GO:0033993: response to lipid
Innate immune response
Response to lipopolysaccharide
Response to molecule of bacterial origin
Up-regulated DEGs
-log10(P)
CR1
¤ DPEP1
* FDR<0.25
** FDR<0.10
*** FDR<0.05


B
Adult Cohort
P=1.0
P=0.063
P=0.20
CR1 expression (log2)
Diagnosis
Relapse
BM-Controls
Pediatric Cohort
P=1.0
P=0.052
P=0.69
CR1 expression (log2)
Diagnosis
Relapse
BM-Controls

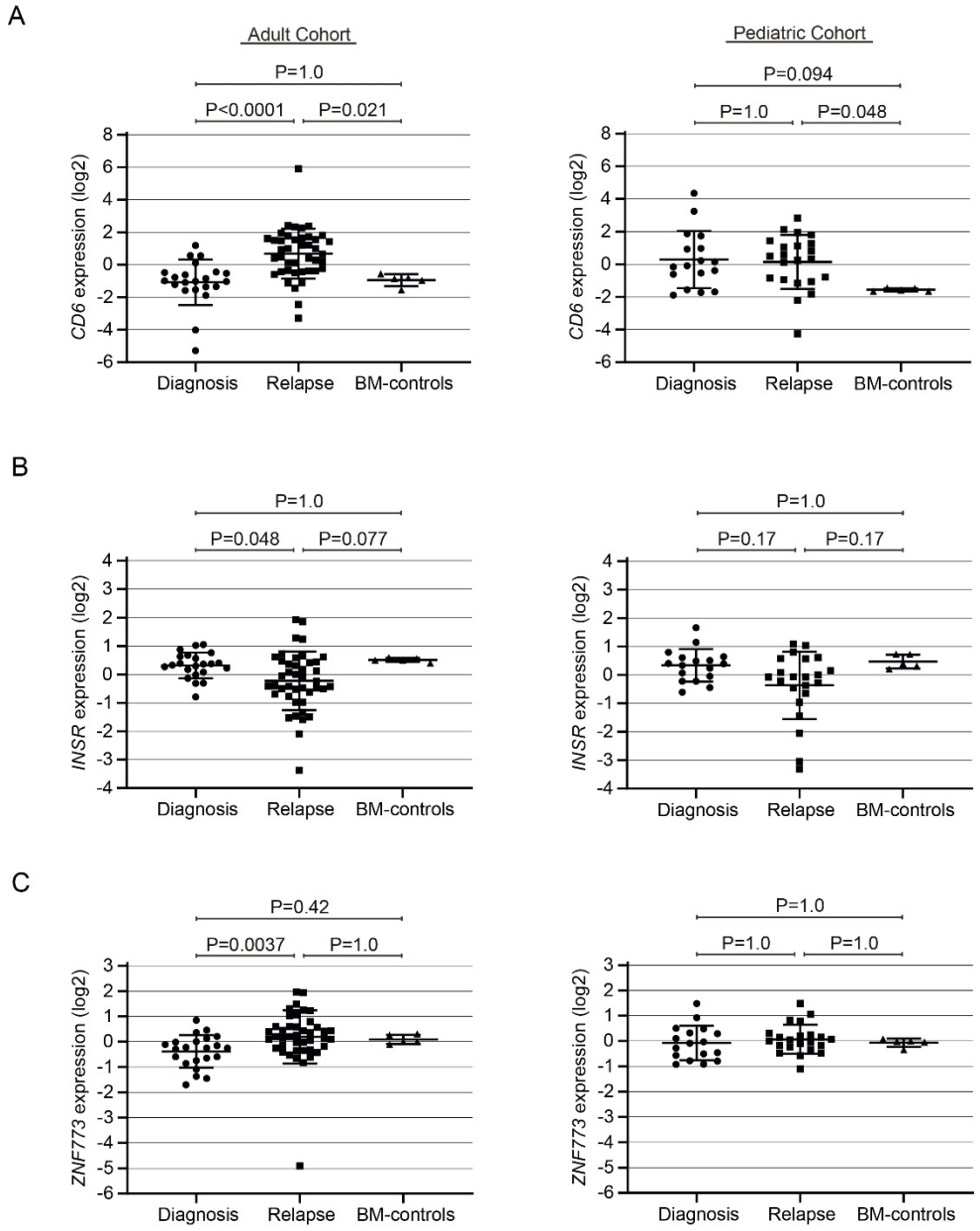
C
Adult Cohort
P=1.0
P=0.042
P=0.091
DPEP1 expression (log2)
Diagnosis
Relapse
BM-Controls
Pediatric Cohort
P=1.0
P=0.11
P=0.074
DPEP1 expression (log2)
Diagnosis
Relapse
BM-Controls

Supplemental Figure 9. Gene expression of highly ranked genes associated with diagnosis or relapse. A-B) GO-analysis of DEGs between diagnosis and paired relapse samples for (**A-upper**) adult and (**A-lower**) pediatric samples. GO-terms presented above the X-axis are enriched among genes upregulated at diagnosis compared to relapse, while pathways below the X-axis are enriched among downregulated genes. **B-C)** Scatter plots with mean and SD presenting gene expression data comparing diagnosis and relapse samples for *CR1* (**B**) and *DPEP1* (**C**) among the local adult and pediatric cohorts, including the BM-control samples. Applied statistical test: Kruskal-Wallis test followed by Dunn´s correction for multi-group comparisons. The Y-axis represents log2 transformed, TMM normalized expression of RNA. DEGs, Differentially expressed genes; FDR, False discovery rate (* FDR<0.25, ** FDR<0.1, *** FDR<0.05; Benjamini–Hochberg correction); SD, Standard deviation; TMM, Trimmed mean of M-values. # *CR1*, ¤ *DPEP1*. See **Supplemental Table 10B** for details regarding samples included in this figure, and **Supplemental Table 14** for all DEGs.

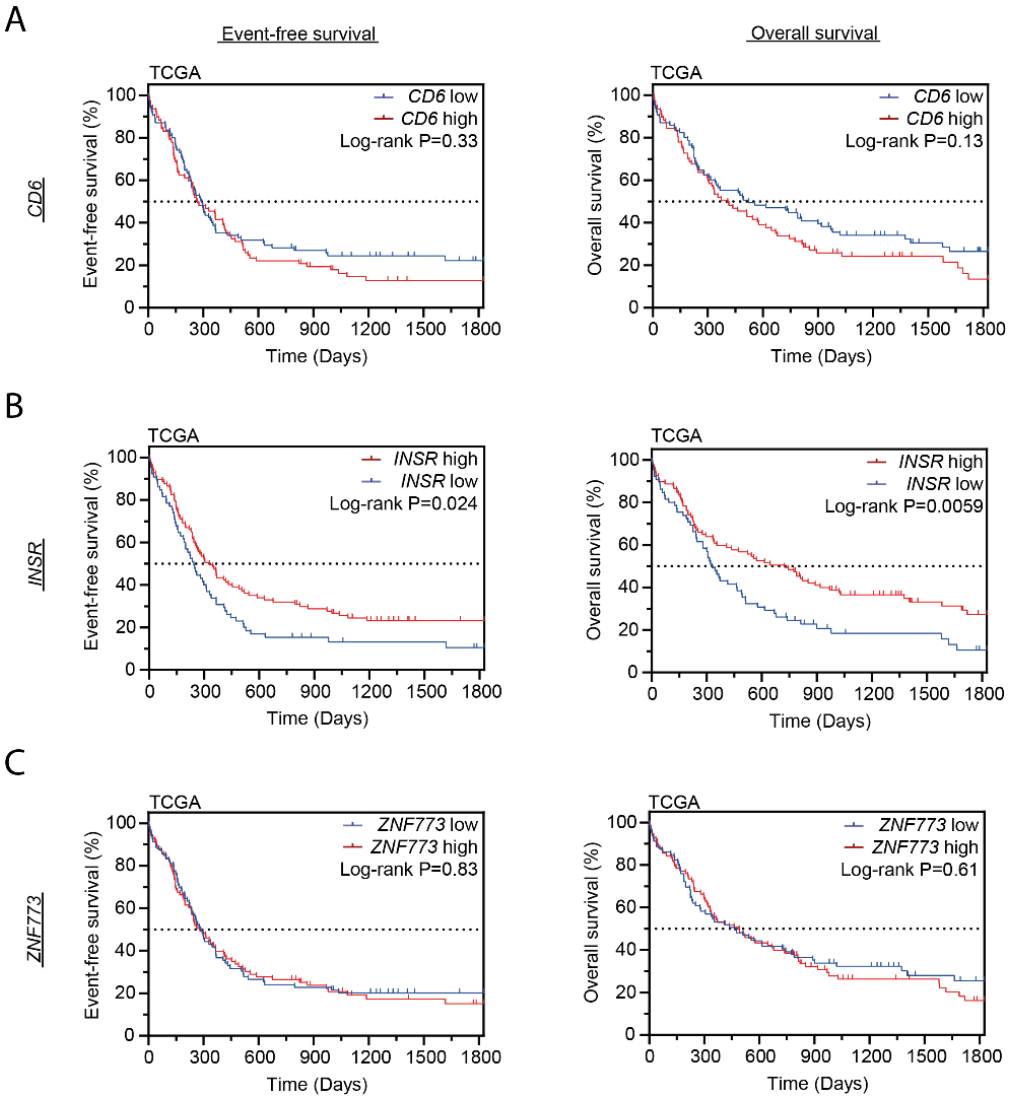


Supplemental Figure 10. Workflow for feature prediction using machine learning-based analysis. The top level of the workflow shows the number of samples used for each cohort. Initial data filtration steps are depicted on the left-hand side as detailed in Supplemental Methods – "Interpretable supervised learning to obtain rule-based classifiers for disease states" – "Pre-processing of RNA-seq data". The upper middle level of the workflow shows four sequential steps for analyzing the data for each cohort and biological interpretation. The lower middle level shows the pipeline for constructing new models based on a merged list of features. This step was performed in order to validate cohorts and reveal common co-predictors. Detailed description of the steps included in the middle level of this figure are given in the following sub-sections in the Supplemental Methods: "Data discretization and feature selection",

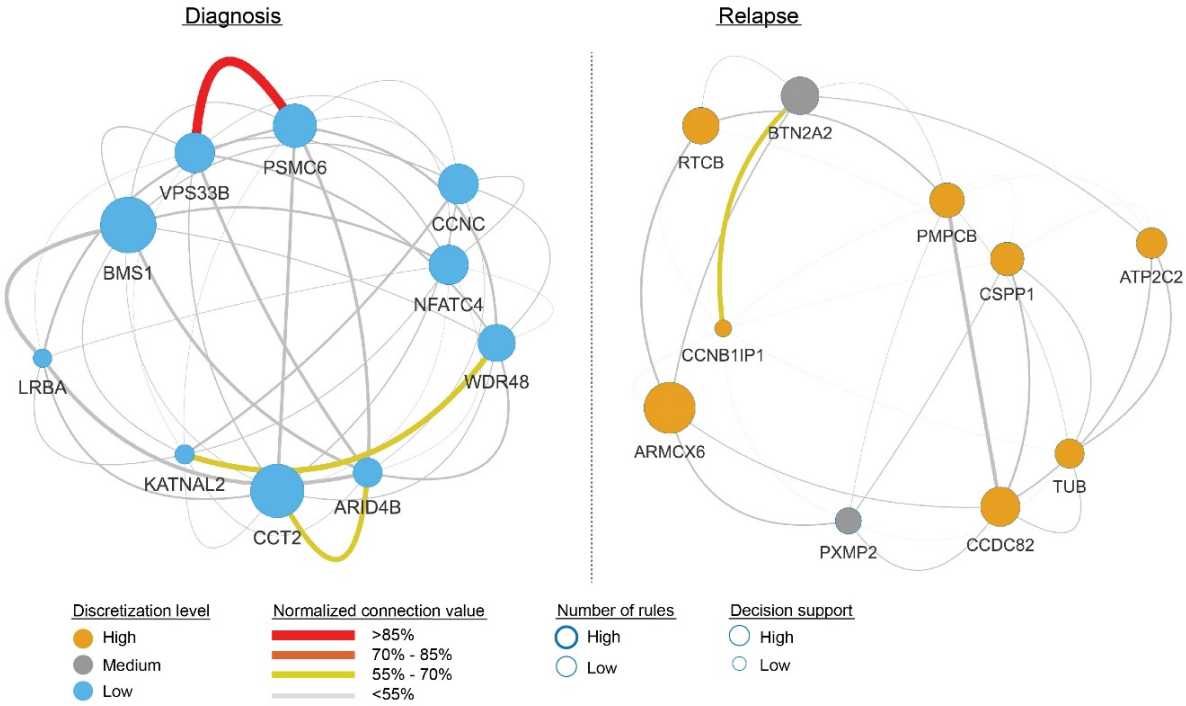
"Optimizing number of selected features for rule-based learning", "Constructing rule-based models", and "Validating rule-based models". The right-hand side depicts generation of rule-based networks, heat maps, arc plots, as well as network comparisons, with detailed description of these steps given in the Supplemental Methods sub-sections "Overview of the analysis pipeline", "Rule-based heat maps for evaluating classifiers", and "Network-based comparisons and hubs visualization". This figure was designed using resources from www.flaticon.com. See **Supplemental Table 10C** and **D** for details regarding samples included in the machine learning-based analyses, and **Supplemental Tables 16-19** for all rules generated through these analyses.



Supplemental Figure 11. Relapse-specific differential expression of *CD6*, *INSR* and *ZNF773* in adult AML. Scatter plots with mean and SD depicting *CD6* (**A**), *INSR* (**B**) and *ZNF773* (**C**) expression between diagnosis, relapse and BM-control samples, for adult (left) and pediatric (right) cases. Applied statistical test: Kruskal-Wallis test followed by Dunn´s correction for multi-group comparisons. The Y-axis represents log2 transformed, TMM normalized expression of mRNA. SD, Standard deviation; TMM, Trimmed mean of M-values. See **Supplemental Table 10C** for details regarding samples included in this figure.

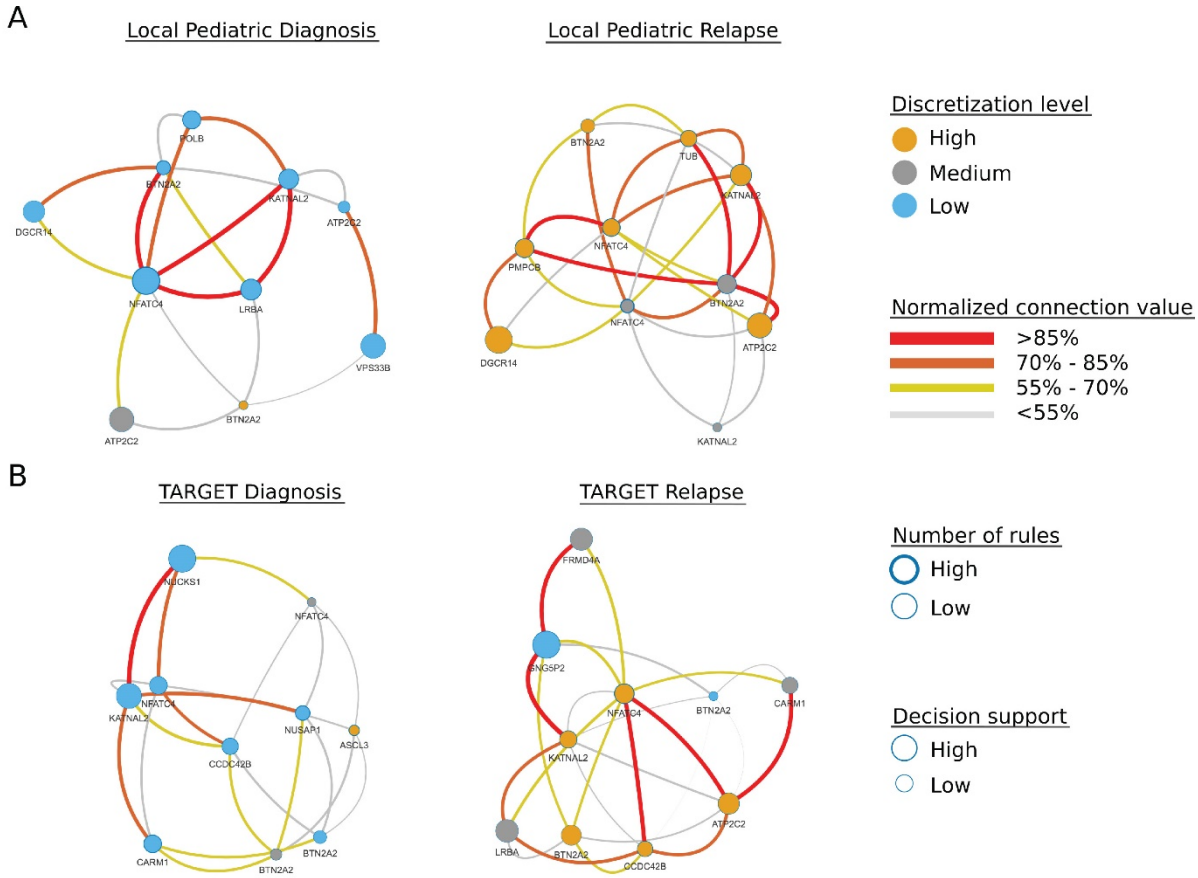


734 **Supplemental Figure 12. Low *INSR* expression is associated with worse disease outcome**
735 **for the TCGA cohort. A-C)** Kaplan-Meier plots depicting event-free survival (left) and overall
736 survival (right), comparing high and low *CD6* (**A**), *INSR* (**B**) and *ZNF773* (**C**) expression within
737 the TCGA cohort. High and low gene expression was discretized based on the mean expression
738 for the respective gene over all samples included in the analysis. Applied statistical test: Log-
739 rank (Mantel-Cox) test. See **Supplemental Table 10F** for details regarding samples included
740 in this figure, and **Supplemental Table 13** for accompanied statistical results.



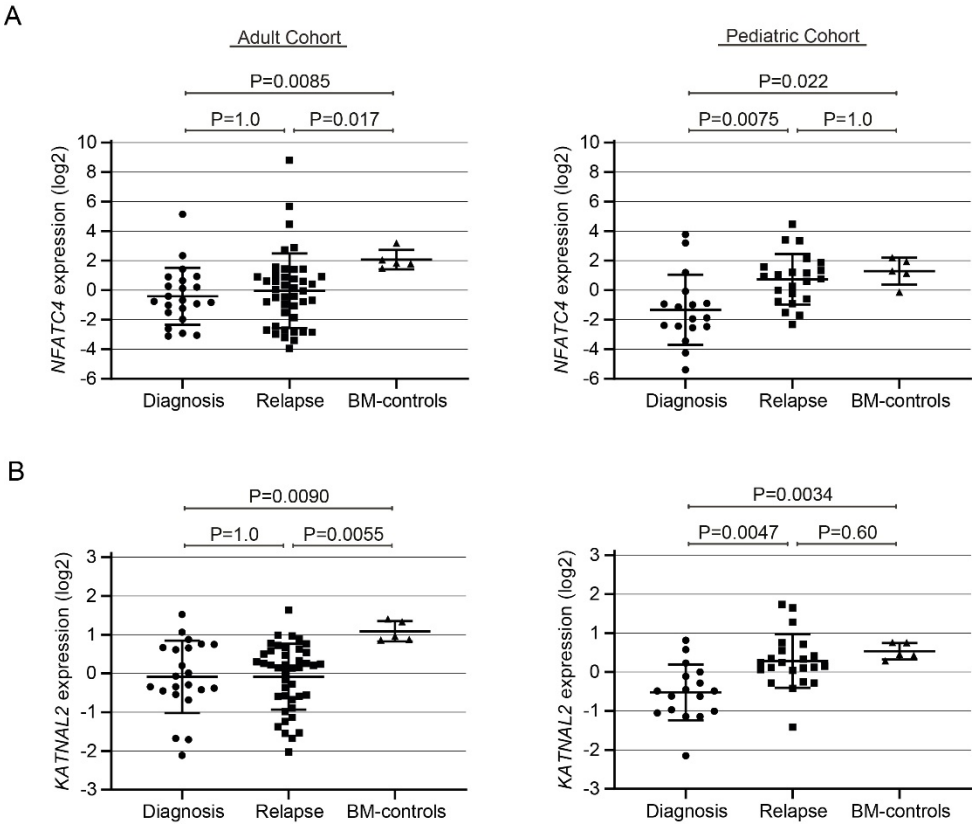
741

742 **Supplemental Figure 13. Co-predictive features detected by machine learning-based**
743 **analysis in pediatric AML.** Relationships between co-predictive features associated with
744 diagnosis (left) and relapse (right) among pediatric AML cases are visualized utilizing VisuNet.
745 The Genetic reducer was applied. The color of the nodes shows the expression level, with three
746 bins for high (orange), medium (grey) and low (blue) expression. The rule support is shown by
747 the size of the respective node, while the support for each connection is visualized by the
748 thickness and color of the connective line. Rules were filtered according to FDR <0.05. See
749 **Supplemental Table 10C-lower** for details regarding samples included for generating the data
750 in this figure. FDR, False discovery rate.



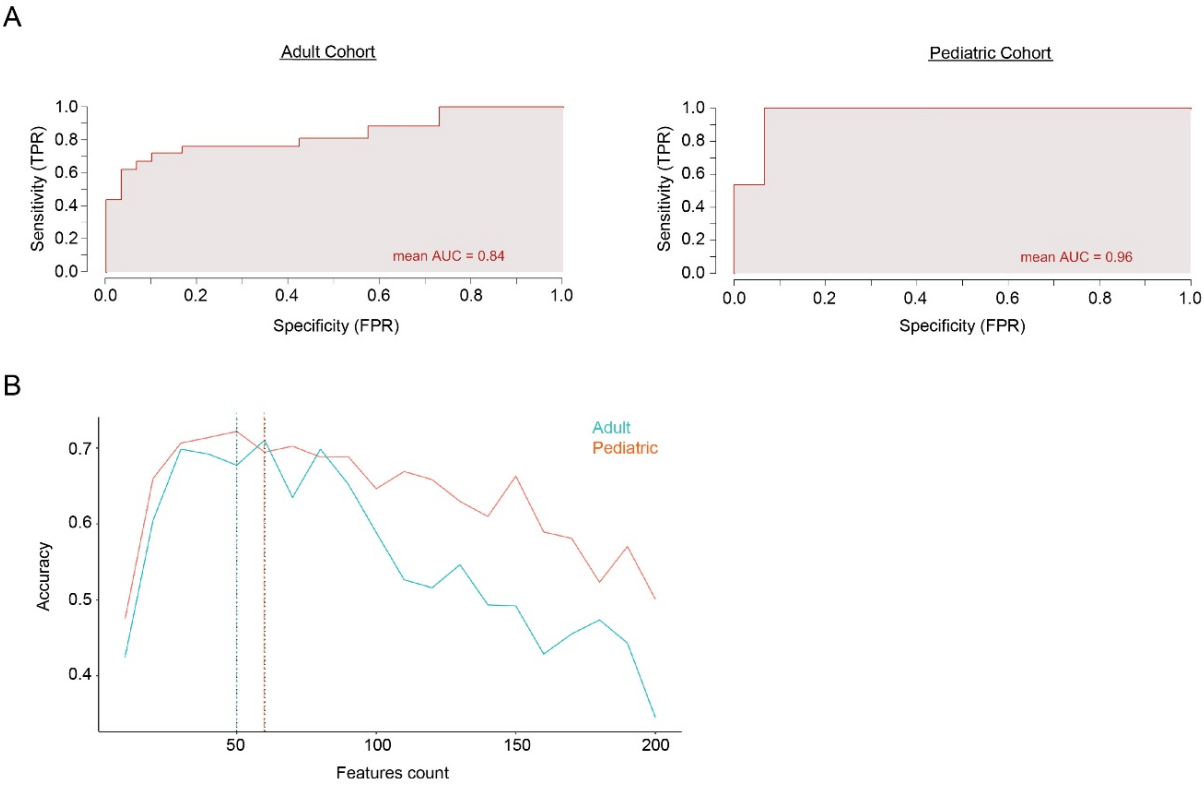
751

752 **Supplemental Figure 14. Co-predictive features detected by machine learning-based**
753 **analysis in local pediatric and TARGET AML cohorts. A-B)** Relationships between top 10
754 co-predictive genes associated with diagnosis (left) and relapse (right) among pediatric local
755 (**A**) and TARGET (**B**) AML cases are visualized utilizing VisuNet. The Johnson reducer was
756 applied on the respective dataset after merging features identified for the separate datasets. The
757 color of the nodes shows the expression level, with three bins for high (orange), medium (grey)
758 and low (blue) expression. The rule support is shown by the size of the respective node, while
759 the support for each connection is visualized by the thickness and color of the connective line.
760 See **Supplemental Table 10C-lower** and **D** for details regarding samples included for
761 generating the data in this figure.

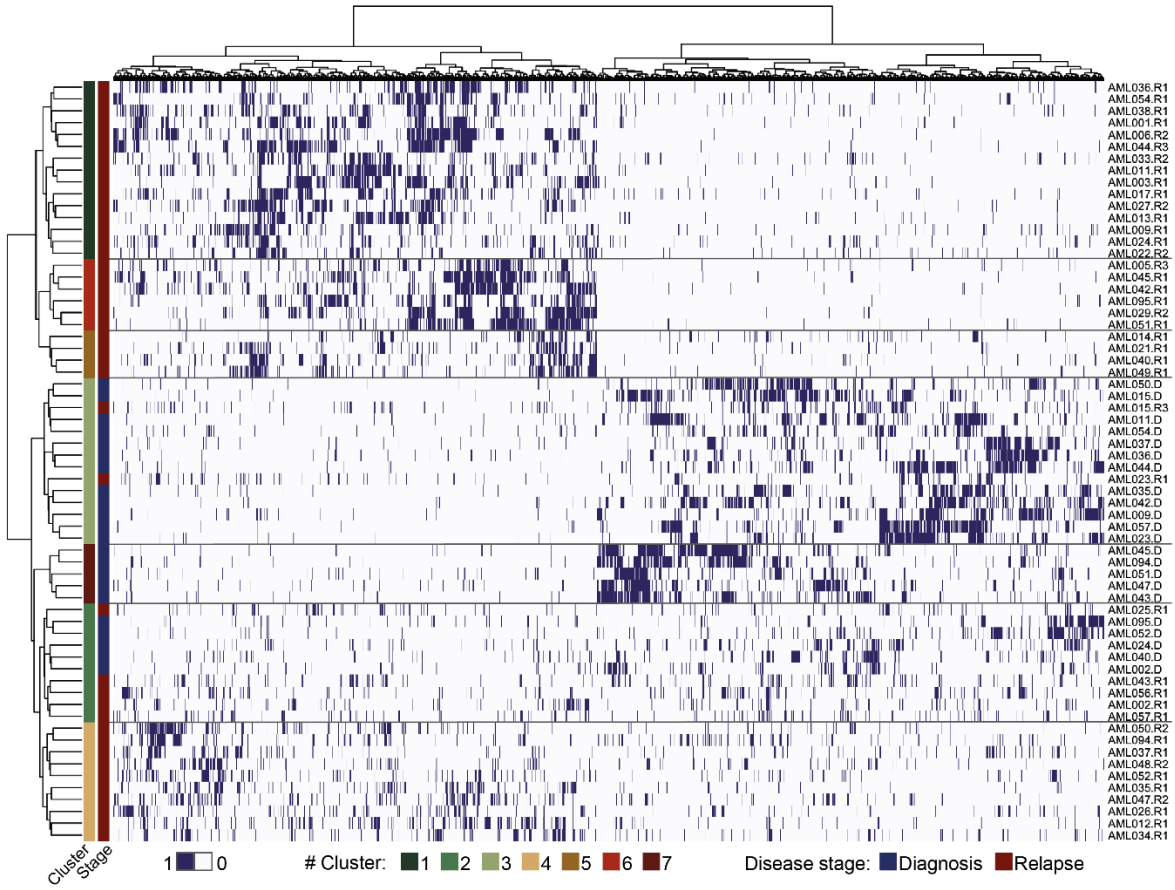


762

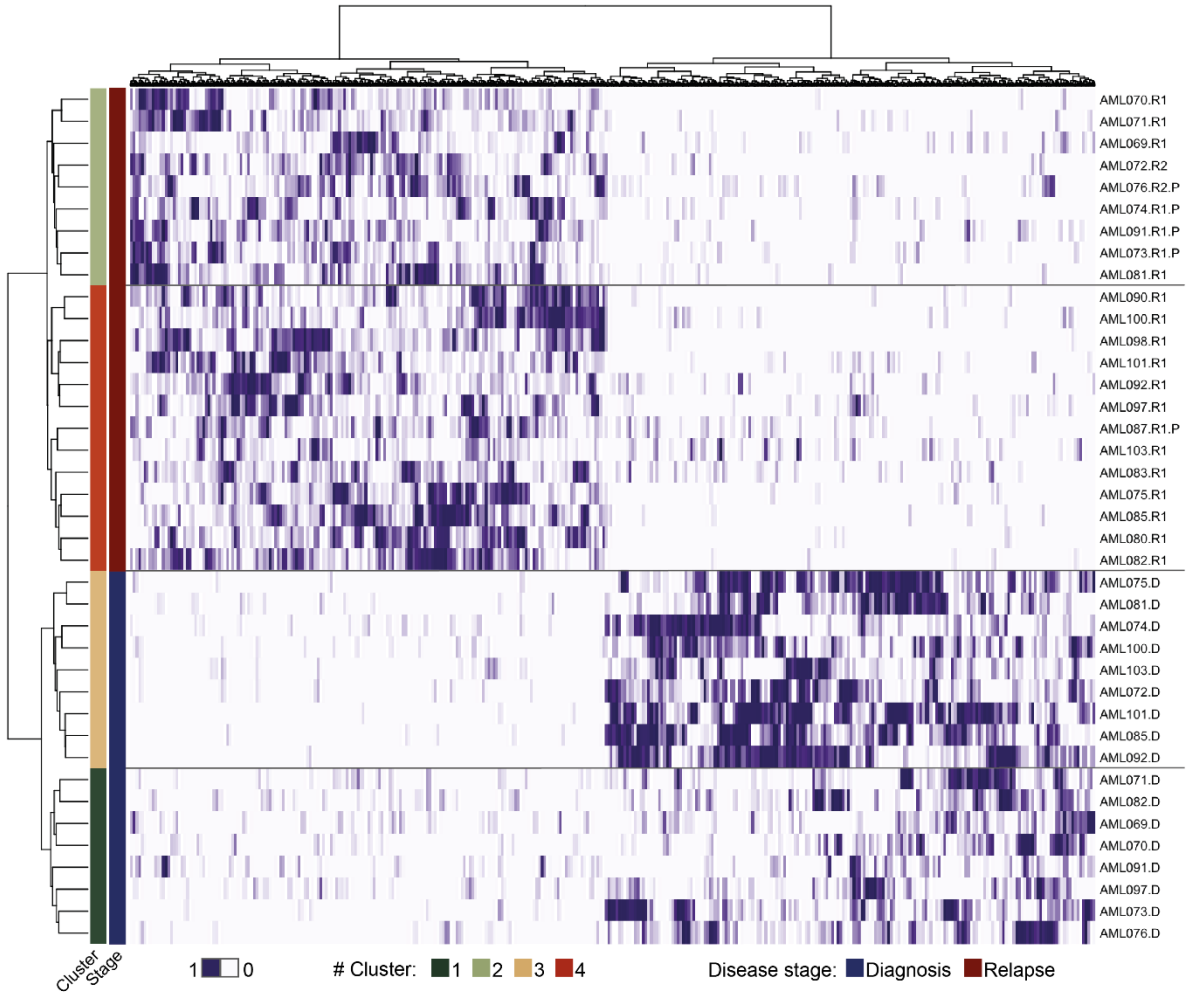
763 **Supplemental Figure 15. Downregulation of *NFATC4* and *KATNAL2* at diagnosis in**
764 **pediatric AML.** Scatter plots with mean and SD showing the log2 transformed, TMM
765 normalized expression values in diagnosis-, unpaired relapse- and BM-control samples for
766 *NFATC4* (**A**) and *KATNAL2* (**B**) for the adult (left) and local pediatric (right) AML cohorts.
767 Applied statistical test: Kruskal-Wallis test followed by Dunn´s correction for multi-group
768 comparisons. SD, Standard deviation; TMM, Trimmed mean of M-values. See **Supplemental**
769 **Table 10C** for details regarding samples included for generating the data in this figure.



Supplemental Figure 16. Statistical evaluation of machine learning-based results. A) Model performance based on the area under the receiver operating characteristic curves (AUCs) for rule based models for the adult cohort (mean AUC = 0.84) and for the pediatric cohort (mean AUC = 0.96). **B**) Accuracies for all the models built using the first 200 top features from MCFS. The highest accuracy is shown for the models built with 50 features for the adult cohort and 60 features for the local pediatric cohort. FPR, False positive rate; MCFS, Monte Carlo feature selection; TPR, True positive rate.



Supplemental Figure 17. Heat map showing clustering of adult AML diagnosis and relapse samples based on the Genetic reducer rule model. The heat map is based on a binary matrix for supported samples per each rule from the rule-based model. The samples that belong to the support set of the rule is given 1 and the rest are assigned 0. An asymmetric binary distance is computed using the matrix for the hierarchical clustering. The heat map shows a good separation of diagnosis and relapse samples based on the rules. The heat map was constructed using the pheatmap R package (v1.0.12). D, Diagnosis; R1/2/3, Relapse 1/2/3. See **Supplemental Table 16** for all the rules that this heat map is based on, and **Supplemental Table 10C-upper** for details regarding samples included in this figure.



Supplemental Figure 18. Heat map showing clustering of pediatric AML diagnosis and relapse samples based on the Genetic reducer rule model. The heat map is based on a binary matrix for supported samples per each rule from the rule-based model. The samples that belong to the support set of the rule is given 1 and the rest are assigned 0. An asymmetric binary distance is computed using the matrix for the hierarchical clustering. The heat map shows a perfect separation of diagnosis and relapse samples based on the rules. The heat map was constructed using the pheatmap R package (v1.0.12). D, Diagnosis; R1/2, Relapse 1/2; R1/2.P, Persistent relapse sample. See **Supplemental Table 17** for all the rules that this heat map is based on, and **Supplemental Table 10C-lower** for details regarding samples included in this figure.

Supplemental References

1. Glimelius B, Melin B, Enblad G, et al. U-CAN: a prospective longitudinal collection of biomaterials and clinical information from adult cancer patients in Sweden. *Acta Oncol*. 2018;57(2):187-194.
2. Stratmann S, Yones SA, Mayrhofer M, et al. Genomic characterization of relapsed acute myeloid leukemia reveals novel putative therapeutic targets. *Blood Adv*. 2021;5(3):900-912.
3. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
4. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood*. 2002;100(7):2292-2302.
5. Dohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447.
6. Patel H, Ewels P, Peltzer A, et al. nf-core/rnaseq: nf-core/rnaseq versi. Vol. 1.1. http://doi.org/10.5281/zenodo.4323183: Zenado; 2020.
7. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. 2017;35(4):316-319.
8. Krueger F. Trim Galore. https://github.com/FelixKrueger/TrimGalore.
9. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2012;29(1):15-21.
10. Broad_Institute. Picard Toolkit. http://broadinstitute.github.io/picard/: Broad Institute; 2019.
11. Sayols S, Scherzinger D, Klein H. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics*. 2016;17(1):428.
12. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930.
13. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*. 2013;43(1):11.10.11-11.10.33.
14. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; 2010.
15. FastQC; 2015.
16. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012;28(16):2184-2185.
17. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biology*. 2019;20(1):213.
18. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic acids research*. 2001;29(1):308-311.
19. Ameur A, Dahlberg J, Olason P, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *European Journal of Human Genetics*. 2017;25(11):1253-1260.
20. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*. 2018;47(D1):D941-D947.

21. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980-985.
22. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
23. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnology*. 2011;29(1):24-26.
24. Krzywinski MI, Schein JE, Birol I, et al. Circos: An information aesthetic for comparative genomics. *Genome Research*. 2009.
25. Yu Y, Ouyang Y, Yao W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics*. 2017;34(7):1229-1231.
26. Team RDC. R: A language and environment for statistical computing.: R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org; 3.5.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009;26(1):139-140.
28. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 2010;11(3):R25.
29. Cancer Genome Atlas Research N, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059-2074.
30. Samur MK. RTCGAToolbox: A New Tool for Exporting TCGA Firehose Data. *PLOS ONE*. 2014;9(9):e106397.
31. Lawrence M, Huber W, Pagès H, et al. Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology*. 2013;9(8):e1003118.
32. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*. 2009;25(14):1841-1842.
33. Morgan M. Rsamtools. Bioconductor version: Release (3.11); 2020:Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.4.0,.
34. Wichura MJ. The Coordinate-Free Approach to Linear Models. Cambridge: Cambridge University Press; 2006.
35. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering Motifs in Ranked Lists of DNA Sequences. *PLOS Computational Biology*. 2007;3(3):e39.
36. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009;10(1):48.
37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289-300.
38. Smolinska K, Garbulowski M, Diamanti K, et al. VisuNet: an interactive tool for rule network visualization of rule-based learning models. GitHub repository: https://github.com/komorowskilab/VisuNet; 2021.
39. Kuhn M. Building Predictive Models in R Using the caret Package. *2008*. 2008;28(5):26.
40. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883.
41. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*. 2016;17(1):483.

42. Draminski M, Rada-Iglesias A, Enroth S, Wadelius C, Koronacki J, Komorowski J. Monte Carlo feature selection for supervised classification. *Bioinformatics*. 2008;24(1):110-117.
43. Garbulowski M, Diamanti K, Smolińska K, et al. R.ROSETTA: an interpretable machine learning framework. *BMC Bioinformatics*. 2021;22(1):110.
44. Komorowski J, Pawlak Z, Polkowski L, Skowron A. Rough Sets: A Tutorial. In: Rough Fuzzy Hybridization: A New Trend in Decision Making. *Springer Verlag, Singapore City*. 1999:3-98.
45. Garbulowski M, Smolinska K, Diamanti K, et al. Interpretable Machine Learning Reveals Dissimilarities Between Subtypes of Autism Spectrum Disorder. *Frontiers in Genetics*. 2021;12(73).
46. Yones SA, Annett A, Stoll P, et al. Interpretable machine learning identifies paediatric Systemic Lupus Erythematosus subtypes based on gene expression data. *bioRxiv*. 2021:2021.2006.2003.446884.
47. Torabi Moghadam B, Dabrowski M, Kaminska B, Grabherr MG, Komorowski J. Combinatorial identification of DNA methylation patterns over age in the human brain. *BMC Bioinformatics*. 2016;17(1):393.
48. Kolde R. pheatmap: Pretty Heatmaps. CRAN: CRAN; 2019.
49. Sanchez G. Introduction to the R package arcdiagram. CiteSeerX; 2016.
50. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1901;2(11):559-572.
51. Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. 1933;24(6):417-441.
52. Jolliffe IT. Principal component analysis. New-York: Springer-Verlag; 2002.
53. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of machine learning research*. 2008;9(11).
54. Van Der Maaten L. Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research*. 2014;15(1):3221-3245.
55. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*. 1998;95(25):14863-14868.