**Comprehensive Transcriptome Profiling of Cryptic *CBFA2T3-GLIS2* Fusion-positive AML Defines Novel Therapeutic Options – A COG and TARGET Pediatric AML Study**

Smith, J., L., Ries, R., E., …, Meshinchi, S.

**Supplemental Methods**

**Statistical Methods.** Data were current as of March 31, 2018 for outcome analyses. Overall survival (OS) and event-free survival (EFS) were calculated from Kaplan-Meier estimates. OS was defined as the time from study entry to death from any cause. EFS was defined as time from study entry to induction failure, relapse, or death. Cox proportional hazard regression models were used to estimate hazard ratios (HR) for univariate and multivariate analyses of OS and EFS. The log-rank statistic was used to compare differences between groups for OS and EFS. The significance of observed differences in proportions was tested by Pearson’s chi-square test. Fisher’s exact test was used when data were sparse. The Kruskal-Wallis test was used to determine the significance between differences in medians for continuous variables.

**Flow Cytometry and Cytotoxicity Assays**. Flow cytometry data from 437 pediatric AML patients from AAML0531 for 13 cell surface antigens (CD45, D34, CD38, HLADR, CD11b, CD36, CD15, CD13, CD14, CD33, CD7, CD56, CD117) were provided by Hematologics, Inc. (Seattle, WA) as mean fluorescent intensity (MFI) values. Methods for flow cytometry and the myeloid immune marked panel above was previously reported (18). Cytotoxicity assays were carried out on primary AML samples from bone marrow or peripheral blood by Notable Labs translational drug discovery platform (Foster City, CA). Cytotoxicity assays utilized the CD56 antibody clone m906 as previously described (19). The antibody controls included chKTi-SPP-DM1 (20) and human IgG-SAP and goat IgG-ZAP (Advanced Targeting Systems, San Diego, CA, #IT-36, #IT-35). Briefly, fresh never frozen samples were red blood cell lysed to remove red blood cells and then plated at 15,000 cells per well in 15ul of serum free medium with cytokines in 384-well plates. Frozen primary samples were thawed in Iscoves’ Modified Dulbecco’s medium (IMDM) with 20% fetal bovine serum (FBS) with cytokines and plated at 30,000 cells per well in 15ul of IMDM/20%FBS/cytokine medium. Plated primary cells were treated with ADC in triplicate and three days post-exposure the samples were stained with appropriate antibodies (Panel: CD45, CD14, CD66B, HLA-DR, CD11B, CD3/CD19/DAPI, CD33, CD56, CD38) and read out on an Intellicyt iQue Plus flow cytometer. Live cells were gated using FSC/SSC and DAPI exclusion, and then further defined by cell surface marker expression. Average absolute counts and standard deviation for each population were calculated, as well as counts normalized to a vehicle-only (DMSO) control for every condition tested.

***CBFA2T3-GLIS2* Transduced Cord Blood.** Human cord blood samples for research were obtained from normal deliveries under Swedish Medical Center Institutional Review Board (Seattle,WA) approval and upon informed donor consent. Samples were red blood cell lysed and enriched for CD34+ fraction using CliniMACS CD34 MicroBeads (Miltenyi Biotec, Germany, #130-046-702). The CD34+ cord blood cells were transduced with a pRRL lentivirus encoding the *CBFA2T3-GLIS2* fusion transcript and GFP (50). Transduced and mock-transduced control cells were cultured in bulk in a myeloid promoting condition (IMDM with 15% FBS and 10 ng/mL of each growth factor SCF, TPO, FLT3L, IL3, and IL6). The transduction efficiency was approximately 9% at MOI of 50 and expression of the fusion transcript was confirmed by RT-PCR, using the primers: forward 5-[6-FAM]CGAAGGGCCTCAGCTAGACGTG-3 and reverse 5-CTCGGGCTTGACATGGTAAT-3. Growth kinetics were evaluated by flow cytometry, and growth rates calculated using GraphPad Prism exponential growth analysis. Differences in proliferation were assessed using a paired two-tailed student’s t-test. Surface protein expression was assessed by multidimensional flow cytometry, as described above (Hematologics Inc, Seattle, WA)(18).

**RNA-sequencing Library Construction.** TotalRNA from diagnostic primary peripheral blood or bone marrow samples was extracted and purified using the QIAcube automated system with AllPrep DNA/RNA/miRNA Universal Kits (QIAGEN, Valencia, CA, #80224). The mRNA libraries were prepared for 75bp strand-specific, paired-end sequencing using the ribodepletion 2.0 protocol by the British Columbia Genome Sciences Center (BCGSC, Vancouver, BC). Libraries were sequenced on the Illumina HiSeq 2000/2500 and aligned to Hg19 (GRCh37-lite) reference genome using BWA v0.5.7 with default parameters, except the addition of "-s" option, and duplicated reads were marked with Picard Tools. Gene level coverage analysis was performed using the BCGSC pipeline v1.1, and composite gene models were generated using Ensembl v69 annotations. Gene counts were quantified as the sum of the fraction of each sequence read inside all exons in the merged gene or transcript model.

The microRNA libraries were produced using the miRNA 3.0 protocol by the British Columbia Genome Sciences Center (BCGSC). Libraries were size-selected to enrich the miRNA-containing fraction and adapter contaminants were removed. Size-selected libraries were sequenced on the Illumina HiSeq 2000 with 31-bp reads. Sequence reads were trimmed to remove adapter sequences; any trimmed reads shorter than 15bp were discarded. The trimmed reads were aligned to Hg19 (GRCh37-lite) using BWA v0.5.7 and duplicated reads were marked with Picard Tools. Only perfect alignments with no mismatches were retained for mature miRNA quantification. Reads were annotated using miRBase v20 and normalized to million miRNA-aligned reads (RPM).

**Screening of *CBFA2T3-GLIS2* Fusion.** The *CBFA2T3-GLIS2* fusion transcript was detected by either A) Fragment length analysis, or B) fusion detection algorithms, STAR-fusion v1.1.0 and TransAbyss v1.4.10 (22,23).

For fragment length analysis, RNA was extracted as above and reverse transcription was performed on total RNA followed by PCR using a fluorescently labeled forward primer on cDNA of the *CBFA2T3-GLIS2* fusion transcript using the following primers: forward 5-[6-FAM]CGAAGGGCCTCAGCTAGACGTG-3 and reverse 5-CTCGGGCTTGACATGGTAAT-3. The denaturing, annealing, and extension steps were performed at 94°C for 30 s, 62.5°C for 45 s, and 72°C for 45 s, respectively, for 40 cycles with an initial 3-min denaturation step at 94°C and a final 5-min extension step at 72°C. The house-keeping gene BCR was used to evaluate the quality of cDNA. Tandem amplification of BCR was performed for all specimens. Fragment length analysis of the PCR product was accomplished using the Applied Biosystems 3730xl DNA Analyzer.

For samples with RNA-sequencing data, STAR-fusion and TransAbyss fusion detection software was employed. STAR-fusion was run using default parameters with the pre-made GRCh37 resource library with Gencode version 19 annotations (<https://data.broadinstitute.org/Trinity/CTAT_RESOURCE_LIB/>) (22). TransAbyss required fusions to have breakpoint reads ≥ 1, flanking pairs ≥ 2 counts, and spanning reads ≥ 2 counts (23). Fusions detected computationally were verified using RT-PCR, with primers listed above. For one case, the forward primer 5’-GCAGAAGAGTGGAAGCACCT-3’ was used for verification due to a non-typical breakpoint. Fusion gene expression was quantified and visualized on IGV using Fusion Inspector v.1.2.0 (Broad Institute, Cambridge, MA) (22). The median spanning fusion fragment count was 46.5 reads (range: 3.0 - 195.0), and median junction supporting count was 51 reads (range: 6.0 - 190.0). Abundance of the fusion fragments per million mapped reads (FFPM) was 1.0 FFPM (range: 0.2 - 5.1).

**Differential Expression, Hierarchical Clustering, Gene-Set Enrichment Analysis, and miRNA-mRNA Interactions.** Differential expression analysis was performed using mRNA-sequencing or miRNA-sequencing data from 1,049 diagnostic AML patients and 62 healthy normal marrows. Differentially expressed genes and miRNAs between *CBFA2T3-GLIS2* positive and negative patients were identified using the Limma v.3.36.1 R package with trimmed mean of M values (TMM) normalized gene counts. Genes with absolute log2 fold change > 1 and Benjamini–Hochberg adjusted p-values < 0.05 (false discovery rate, FDR) were retained for further analysis.

For cluster analysis, the read counts were TMM normalized and converted to log2 scale with a count of 1 added to avoid taking the log of zero. Unsupervised hierarchical clustering was performed using Euclidean distance matrices with the ward.D2 linkage algorithm. Classical multidimensional scaling was performed with mean fluorescent intensity (MFI) values from flow cytometry data using the vegan v2.5-4 R package; MFI values were log2 transformed MFI with an added count of 1 prior to analysis.

Gene-set enrichment analysis was completed with log2 TMM normalized counts with an added count of 1 as input. Gene-set enrichment was performed using the ‘unpaired’ comparison in the GAGE v2.30.0 R package (24), which tests for differential expression of gene-sets by contrasting all possible combinations of fusion-positive to fusion-negative samples. Gene-sets from the Molecular Signatures Database (MSigDB) and the KEGG pathway database were used and non-redundant gene-sets were extracted for further analysis and identification of core genes that contribute to the pathway enrichment using the GAGE package.

Regulatory interactions between miRNA-mRNA were investigated by selecting pairs of differentially expressed miRNAs and genes with significant anti-correlation using Spearman’s *rho* using log2+1 transformed TMM normalized counts. Significance of anti-correlations were adjusted for multiple hypothesis testing using the Benjamini–Hochberg method. The anti-correlated miRNA-miRNA pairs were then investigated for predicted and validated interactions by querying the databases: DIANA microT CDS, EIMMo, Microcosm, miRDB, miRanda, PITA, rna22, TargetScan, miRecords, miRTarBase, PicTar, and TarBase using the anamiR v1.10.0 and multiMiR v1.4.0 R packages, followed by gene ontology (GO) term enrichment for target genes using the goseq v1.32.0 package.

**miRNA Knockdown, Proliferation Assays, and QPCR.**  The patient derived MO7E cell line, which natively possesses the *CBFA2T3-GLIS2* fusion, was purchased commercially in June 2017 (DSMZ, Germany, RRID: CVCL\_2106, Cat# ACC-104). Cell line authentication was performed by DSMZ using short tandem repeat (STR) DNA typing and aliquots derived from the commercial source were cultured for 30 days prior to inhibitor exposure. Screening for *Mycoplasma* was not performed as cell growth and morphology appeared normal and time in culture was less than 6 months from resuscitation. Cells were cultured in RPMI-1640 media (Gibco, Gaithersburg, MD) with 10% FB Essence (Seradigm VWR, Radnor, PA) plus 10 ng/mL rhIL-3 (Peprotech, Rocky Hill, NJ). They were exposed to miRCURY LNA miRNA Power inhibitors (QIAGEN, Germantown, MD, #339132 and #339137) directed against 1) *hsa-miR-224-5p,* 2) *hsa-miR-224-3p,* 3) *hsa-miR-452-5p*, or 4) *hsa-miR-452-3p*, as well as negative controls (A and B) at 800 nM final concentration for 72 hours using non-assisted uptake. RNA was extracted using the miRNeasy Mini Kit (QIAGEN, #217004) and the level of knockdown for each miRNA was measured using miRNA QPCR assays purchased from Thermo Fisher (Waltham, MA, #4398987, #A25576, #4427975) and run on a CFX384 Touch Real Time QPCR Instrument (Bio-Rad, Hercules, CA). All data was analyzed using a delta delta CT calculation with RNU49 used as the endogenous control and the mean of the negative control LNAs used as the reference value. Proliferation assays were then carried out using the CellTiter-Glo Luminescent Cell Viability Assay (Promega, Madison, WI, #G7570). Briefly, cells were plated in triplicate at 20,000 cells/well in 96-well plates in growth media with cytokine and LNA inhibitors. The luminescent values were obtained at 0, 24, 48, 72, and 96 hours using a Bio-Tek Synergy 2 plate reader (Winooski, VT). The proliferation rate for each sample was calculated by dividing each time point value by the mean of the time 0 value. Differences in proliferation rates were assessed using a paired Student’s t-Test.