A Genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets

Sophia Adamia¹, Benjamin Haibe-Kains²*, Patrick M Pilarski³*, Michal Bar-Natan⁴, Samuel Pevzner⁵, Herve Avet-Loiseau ⁶, Laurence Lode ⁷, Sigitas Verselis ⁸, Edward A. Fox ⁸, John Burke⁹, Ilene Galinsky¹⁰, Ibiayi Dagogo-Jack¹¹, Martha Wadleigh¹,¹⁰, David P Steensma¹,¹⁰, Gabriela Motyckova¹,¹⁰, Daniel J Deangelio¹,¹⁰, John Quackenbush¹², Richard Stone¹,¹⁰ and James D Griffin¹,¹⁰.

¹ Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA;
² Bioinformatics and Computational Genomics Laboratory, Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada
³ Department of Computing Science, University of Alberta, Edmonton, AB, Canada;
⁴ Institute of Hematology, Davidoff Cancer Center, Beilinson Hospital, Rabin Medical Center, Petah-Tikva, Israel.
⁵ Dana-Farber Cancer institute, Center for Cancer Systems Biology and Department of Genetics, Harvard Medical School, Boston University school of Medicine and Biomedical Engineering Department, Boston University, Boston, MA;
⁶ Unité de Génomique du Myélome, Laboratoire UGM, University Hospital, CHU Rangueil, Toulouse, France;
⁷ Hematology Laboratory, University Hospital; and INSERM U892, Nantes, France;
⁸ Molecular Diagnostics Laboratory, Dana Farber Cancer Institute, Boston, MA;
⁹ Biotique Systems inc;
¹⁰ Adult Leukemia Program, Dana Farber Cancer Institute, Boston, MA;
¹¹ Brigham and Women's Hospital, Boston, MA;
Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA.

**Running title:** Aberrant splicing in AML-novel markers and potential drug targets

**Keywords:** splicing, acute myeloid leukemia, biomarkers, therapeutic targets.

**Financial support:** SA is supported by the ASH scholar award for Clinical/Translational Research, LLS Translational Research program grant, and supported by the AACR-Amgen, Inc. Fellowship in Clinical/Translational Cancer Research; BHK is supported by internal funding at IRCM; PMP is supported by the Alberta Innovates Centre for Machine Learning; JQ was supported by a grant from the National Library of Medicine of the US National Institutes of Health (R01 LM010129-01).

**Discloser conflict of interests:** JDG receives research support and is a consultant for Novartis Pharma AG. The other authors declare that they have no conflict of interests related to the work reported in this manuscript.

**Corresponding Authors:** Sophia Adamia, PhD

Dana-Farber Cancer Institute
450 Brookline Avenue
Boston, MA, USA 02215
sophia_adamia@dfci.harvard.edu
Tel- 617 632 6083; Fax- 617 632 4388

James D Griffin, MD
Dana-Farber Cancer Institute
450 Brookline Avenue
Boston, MA, USA 02215
James_Griffin@dfci.harvard.edu
Tel- 617 632 3360; Fax- 617 632 4388
Translational relevance

Despite a century of medical advances in pharmaceutical drugs and diagnostic methods, acute myeloid leukemia (AML) remains an incurable disease. Standard treatment for AML is very toxic and includes intensive multi-agent chemotherapy and/or hematopoietic stem cell transplantation. Neither of these therapies appears to be effective, as few patients are cured. To improve outcomes, there is a desperate need for new treatments specifically targeted to the genetic lesions that cause the disease. Also, more effective drug design requires an expanded view of the molecular complexity that underlies AML. To address this issue we performed genome-wide splicing analysis in patients with AML and identified widespread alternative splicing changes in these patients compared to normal donors (NDs). Our results suggest that aberrant splicing is a common characteristic of AML. Our results also suggest that splice variant transcripts that are results of aberrant splicing can be used as selective disease markers. These markers provide potential targets for either therapeutic antibodies or for the delivery of cytotoxic drugs or toxins to leukemic cells.
Abstract

Purpose: Despite new treatments, acute myeloid leukemia (AML) remains an incurable disease. More effective drug design requires an expanded view of the molecular complexity that underlies AML. Alternative splicing (AS) of RNA is used by normal cells to generate protein diversity. Growing evidence indicates that aberrant splicing of genes plays a key role in cancer. We investigated genome-wide splicing abnormalities in AML and based on these abnormalities we aimed to identify novel potential biomarkers and therapeutic targets.

Experimental Design: We used genome-wide AS screening to investigate AS abnormalities in two independent AML patient cohorts (DFCI and UHN) and normal donors (NDs). Selected splicing events were confirmed through cloning and sequencing analysis, and than validated in 193 AML patients.

Results: Our results show that ~29% of expressed genes genome-wide were differentially and recurrently spliced in AML patients compared to NDs bone marrow CD34+ cells. Results were reproducible in two independent AML cohorts. In both cohorts, annotation analyses indicated similar proportions of differentially spliced genes encoding several oncogenes, tumor suppressor proteins, splicing factors and heterogeneous-nuclear-ribonucleoproteins, proteins involved in apoptosis, cell proliferation, and spliceosome assembly. Our findings are consistent with reports for other malignances and indicate that AML-specific aberrations in splicing mechanisms are hallmark of AML pathogenesis.

Conclusions: Overall, our results suggest that aberrant splicing is a common characteristic for AML. Our findings also suggest that splice variant transcripts that are the result of splicing aberrations create novel disease markers and provide potential targets for small molecules or antibody therapeutics for this disease.
**Introduction**

Acute myeloid leukemia (AML) is a heterogeneous neoplasm characterized by the accumulation of myeloid blasts both in the bone marrow (BM) and peripheral blood (PB) of patients. The leukemic blasts are arrested at various stages of granulocytic and monocytic differentiation. Even though these blasts are blocked at different stages of differentiation, they are stem cells and have a natural ability to proliferate. During the proliferation process AML stem cells accumulate various genetic and epigenetic abnormalities, including aberrations in pre-mRNA processing.

Pre-mRNA processing, referred to as alternative RNA splicing (AS), is a critical determinant of protein diversity (1, 2). AS produces multiple transcripts and, as a result, multiple proteins from a single gene. Pre-mRNA splicing is executed in the nucleus by spliceosomes (1, 3, 4). The efficiency of this process is controlled by classical trans- and cis-splicing elements (2, 5, 6). Most studies of AS in AML patients have focused on individual genes, including AML1, Gfi1b, CD96, survivin-2B, AML-ETO and GSK3beta (7-13). Studies that report splice variants of these genes also document the role of some of these variants in disease biology, while others can be used as prognostic or diagnostic biomarkers (13-20). Recently, frequent splicing machinery pathway mutations in myelodysplasia (MDS), AML and chronic lymphocytic leukemia (CLL) were identified (21-30). Despite novel discoveries, genome-wide AS has not been extensively examined in patients with AML.

In present study, we evaluated genome-wide AS events in patients with AML. AS was found to be a common event for AML involving many genes in AML patients. Genome-wide, we found
aberrant patterns of splicing in AML patients compared to BM CD34+ cells from NDs in up to 29% of the annotated genes (P<0.05). Some of these aberrant splicing events are highly recurrent in patients and affect genes encoding many oncogenes and tumor suppressor genes, or genes involved in regulation of apoptosis, the cell cycle, and cell differentiation. Overall, our study identifies widespread splicing abnormalities in AML, which points to a disruption in the mechanisms regulating in the splicing process. In the future, determining the underlying causes and consequences of aberrant splicing in AML could enhance our understanding of disease pathogenesis. Certain of the more common aberrant splice variants may generate new targets for the development of novel therapeutics for AML.

Materials and methods

Study design and patient cohort. This study includes a total of 228 samples from patients with AML; 193 samples were obtained from patients who were recruited at the DFCI (Dana-Farber Cancer Institute, USA) and 35 patients recruited at the UHN (the Hospital de Nantes, France)). From these samples, 66 were analyzed on the Affymetrix Human Exon 1.0ST Arrays (31 from DFCI and 35 from UHN). Of the 228 samples 193 were used in validation studies and 29 in longitudinal study. All patient samples were taken at diagnosis after approval from the Institutional Review Boards, with written informed consent. Supplementary Tables S1 describes demographic, clinical and molecular characteristics of the AML patients that were included in the exon array analysis; patient records were reviewed retrospectively to verify diagnosis and status. A control group of 12 NDs of both genders were also included: 8 BM samples of NDs were obtained from AllCells (California) and 4 PB samples were obtained from healthy volunteers at the DFCI.
Tissue and cell preparation. All samples from patients and NDs were purified using Ficoll–Hypaque gradient centrifugation. PB mononuclear cells and/or BM cells obtained from AML patients were enriched using the EasySep Human Progenitor Cell Enrichment Kit, which is designed to isolate hematopoietic progenitor cells by negative selection (Stem Cell Technologies). From ND BM samples, CD34+ cells were obtained using the EasySep Human CD34 Positive Selection Kit (Stem Cell Technologies). After cell selection the presence of CD34+CD38- and CD34+CD38+, cell populations were evaluated by flow cytometry, which is routinely done in our laboratory.

Monocytes and neutrophils were isolated from four PB samples obtained from healthy volunteers using the EasySep Human Monocyte or Neutrophil Enrichment Kits (Stem Cell Technologies). Cells were resuspended in RLT lysis buffer or in Trizol followed by the total RNA isolation procedure using an AllPrep RNA/DNA Isolation Kit (Qiagen) or Trizol (Life Technologies) reagent according to the manufacturer’s recommendations. Patient samples included in the longitudinal studies were not enriched.

Microarray Hybridization. Prior to the Affymetrix GeneChip Whole Transcript Sense Target Labeling assay, total RNA samples were evaluated on the 2100 Bioanalyzer (Agilent) using the RNA 6000 Nano Kit (Agilent). Subsequently, samples were hybridized to the Affymetrix Human Exon 1.0ST Arrays. All procedures were carried out at the DFCI Microarray core facility according to the recommended protocols. Hybridized arrays were scanned on an Affymetrix GeneChip Scanner 3000 and visually inspected for hybridization artifacts.
Exon array analysis. Data from the exon array experiments were analyzed with the XRay software package. On the exon array, there are approximately four probes per exon and roughly 40 probes per gene. Multiple probes per exon enable “exon-level” analysis and identification of putative splice variants. For each gene, exon (probe-set) expression was analyzed using mixed model analysis of variance (ANOVA) according to the linear model:

\[ Y = M + D + E + C + ED + EC + \text{err} \]

where “M” is a global mean of expression for a gene, “D” is the probe-set expression detected in AML sample versus NDs, “E” is the expression attributed to each exon, and “C” is hybridization effect (random factor). Genes with significant “D” (AML vs NDs) effect are said to show significant AML differential gene expression. However, genes with significant “ED” p-values are said display significant AML vs NDs differential splicing. Computation was performed such that, in general, positive ED scores represent cases of higher exon inclusion (or expression) in AML vs NDs and the opposite is true for negative ED terms. All p- and significance values were corrected for multiple test effects using the Benjamini and Hochberg False Discovery Rate (FDR) method.

Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR). RT-PCR and capillary electrophoresis and DNA fragment analysis were performed as previously described by Adamia et al. (31). The CD13 gene-specific transcripts were amplified from cDNA obtained through an RT reaction using a reverse (CTTTAAGCACAGCTCCCTGAAT) and forward (ATGATTTTCGAGAGAAGGTCCA) primer set designed for the CD13 transcript. After the PCR reactions samples were either stored at -20°C for later analysis, or immediately processed for agarose gel or capillary electrophoresis and DNA fragment analysis on the ABI 31/30XL DNA genetic analyzer.
Cloning and sequencing. cDNAs for cloning and sequencing were obtained from the blasts of 8 AML patients all expressing CD13 splice variants. The CD13 transcripts were amplified in PCR reaction containing platinum Taq High Fidelity, cloned into the pCR4 TOPO TA cloning system and transformed into Mach1-T1 competent cells according to the manufacturer’s instructions. Positive colonies containing CD13 plasmids were identified through testing individual bacterial colonies by colony PCR. Positive CD13-TOPO-TA colonies were grown overnight in LB medium. Plasmids were prepared using the QIAprep Miniprep Kit (Qiagen). Then subclones from each cloned RT-PCR product from individual patients were sequenced, in both reverse and forward directions.

Flow cytometry. CD13 full length (FL) and splice variant transcripts were engineered according to the splicing modes detected by cloning and sequencing analysis of AML patient samples. Plasmids were subcloned into pCMV6-AC-GFP (OriGene) and transfected into The HEK293T cells using DMRIE-C transfection reagent according the manufacturer’s recommendations (Life Technologies). Alternatively, cells were transfected via nucleofection using the Cell Line Nucleofector Kit V (Lonza, USA). Three days after transfection HEK293T cells were stained with 0.5 ug anti-CD13 antibody (My7, Beckman Coulter), followed by staining with Alexa Fluor 647 labeled anti-mouse IgG. As a control we used cells stained with anti-rabbit IgG-Alexa fluor 647 only or untransfected HEK293T cells. Stained cells were analyzed on a FACS Canto II.

Results

Genome-wide splicing profiles for patients with AML
The Affymetrix GeneChip Human Exon 1.0 ST array was used to identify global changes in splicing profiles associated with AML. Two independent AML cohorts were studied, which included the DFCI cohort of 31 patients and the UHN cohort of 35 patients, for a total of 66 samples. No significant differences were found between these two cohorts with respect to any common clinical characteristics other than hemoglobin levels (Supplementary Table S1a).

**Splicing events associated with the DFCI patient cohort**

We compared splicing events in the DFCI cohort with those detected in NDs. After normalization and filtering procedures, using XRay software, 101,792 probes identifying 7,771 transcripts were scored as having above-background expression levels in the DFCI cohort, or in NDs. This analysis identified 29% (2,251 of 7,771) transcripts differentially spliced in patients compared to NDs, with a significance expression level of nominal P<0.05.

Next, we performed annotation analysis through NetAffx Analysis Center (affymetrix.com) to identify transcriptome regions of the DFCI cohort that were affected by aberrant splicing. This analysis showed that the 2,251 differentially spliced transcripts detected in the DFCI cohort correspond to 8,648 probe-sets. Of these, 6,573 (76%) mapped to known translated regions of the genome, 1,785 (21%) mapped to untranslated regions (introns, 5’UTRs or 3’UTRs) and 290 (3%) mapped to partial transcripts/cDNAs (Table 1).

Using Ingenuity software, we identified gene categories affected by aberrant splicing in the DFCI cohort. We determined that 2,201 of the 2,251 differentially spliced transcripts mapped to the annotated genes in public databases. Mapped genes were categorized based on their location as well
as the functions of proteins they encode. Of the differentially spliced genes, 41% encoded cytoplasmic proteins, 31% encoded nuclear proteins and 11% encoded plasma membrane proteins. Results of these analyses are reported in Supplementary Table S2.

**Splicing events associated with the UHN patient cohort**

To ensure that the genome-wide splicing profiles identified in the DFCI cohort were reproducible, we used a “biological repeats” approach for the exon array analyses. We performed identical exon array experiments for the independent UHN cohort. In this sample-set, 91,673 probes corresponding to 7,212 transcripts were expressed above background in patients or NDs. As in the DFCI cohort, 28% (2,004 of 7,212) of the UHN cohort transcripts were differentially spliced in patients compared to NDs at a significance level of nominal P<0.05. Next, we identified regions of the UHN cohort genome affected by the aberrant splicing. The differentially spliced transcripts in the UHN cohort correspond to 10,932 probe-sets, of which 8,794 (80%) mapped to translated regions, 1,842 (17%) mapped to untranslated regions such as introns, 5’UTRs or 3’UTRs, and 296 (3%) mapped to partial transcripts/cDNAs (Table 1).

We categorized differentially spliced transcripts in the UHN cohort and found that 41% of the significantly spliced genes encoded cytoplasmic proteins, 30% encoded nuclear proteins, and 13% encoded plasma membrane proteins. These fractions are similar to those noted in the DFCI sample set. In both cohorts, annotation analyses indicated similar proportions of differentially spliced genes encoding enzymes, kinases, transcriptional and translational regulators or transporter proteins, splicing factors, heterogeneous nuclear ribonucleoproteins, and proteins involved in spliceosome assembly (Supplementary Table S2a).
Evaluation aberrant splicing event frequencies in DFCI and UHM patient cohorts

To determine degree of splicing abnormalities in patients with AML, aberrant splicing event frequencies were evaluated in the DFCI and UHN cohorts and mapped to the genome. The graphic software, CIRCOS, was employed to illustrate the distribution status of each splicing event, which is represented as a dot on the dot plots. The distance of each dot from the outside circle of the CIRCOS plot represents aberrant splicing event frequencies of each splicing event (Figure 1, Table 2). On average, we detected 350 genes (range 135-786) differentially spliced in the DFCI cohort as compared to NDs, while an average of 506 genes were differentially spliced in the UHN cohort. We also determined the frequency of each spliced gene in individual patients of these two cohorts. In the DFCI cohort each gene was spliced on average in 14 patients (range 1-29), while in the UHN cohort each gene was spliced on average in 18 patients (range 1-34) (Figure 1a-b, Table 2).

We identified 47% (1,062 of 2,251) transcripts in the DFCI cohort and 46% (922 of 2,004) transcripts in the UHN cohort that had been previously reported as alternatively spliced in AML or other tumors and other cellular systems (Table 2). We identified 53% (1,189 of 2,251) novel aberrant splice variant transcripts in the DFCI cohort and 54% (1,082 of 2,004) in the UHN cohort that are not reported. Furthermore, these lists include novel splice variants of 52 oncogenes (e.g., RUNX1, PML, AKT1, ARAF, BRAF, FOS, JUN) and 50 tumor suppressor genes (e.g., ATM, P53, FLT3, HDAC2, STK3, SYK), or genes that regulate cellular processes involved in the maintenance of homeostasis (Supplementary Table S2b). These analyses show that genes mapping to certain chromosomes are more frequently spliced in AML patients than in NDs. Distribution of these aberrant splicing events genome-wide is shown in Figure 1a-c.
Among the large number of aberrantly spliced genes, we have identified some of known splice variants that are associated with AML or other cancers, e.g. splice variants of RUNX1, MLL and CD96 previously reported as spliced in AML, a Syk variant in breast cancer, an FGFR1 variant in breast and brain cancers, a caspase-8 in T-cell leukemia, and CD44 and MDM2 splice variants reported in many different types of cancers.

We evaluated the extent of similarities between the DFCI and UHN patient cohorts with respect to splicing event frequency, and found that 75% (262 of 348) of the transcripts that were recurrently spliced in 30% or more of DFCI patient cohort were also recurrently spliced in the UHN patient cohort (Figure 1c, Supplementary Table S2c). This analysis supports the conclusion that the genome-wide splicing profiles for AML are reproducible.

**Aberrant RNA splicing patterns and their frequencies in AML patients**

Using annotation analysis, we observed three different patterns of splicing in patients with AML: Pattern 1—full or partial exon(s) skipping (sequence exclusions), Pattern 2—full or partial cryptic exon or intron retentions (sequence inclusions), and Pattern 3—various combinations of both sequence exclusions and inclusions (Figure 1d). We evaluated the frequency of different patterns of gene splicing events in individual patients of the DFCI and UHN cohorts. We identified that collectively, in both cohorts, 486 transcripts were affected by full or partial exon skipping (Pattern 1), while 1148 transcripts were affected by full or partial intron retentions (pattern 2), and 2531 transcripts were affected by Patterns 1 and 2 in combinations. Also, an average we identified 7 transcripts per patient affected by pattern 1 splicing, 17 transcripts by Pattern 2 splicing and 38...
transcripts were subjected to Pattern 1 and 2 splicing in various combinations. Thus, complex combinations of splicing events are frequently detected in AML patients.

**Gene categories and pathways affected by aberrant splicing in AML**

To identify deregulated pathways affected by aberrant splicing events in patients with AML, we performed a functional enrichment analysis using DAVID (the Database for Annotation, Visualization and Integrated Discovery). This enrichment analysis identified a number of signaling pathways with significance $P < 0.005$ affected by aberrant splicing (Table 3). In the DFCI cohort we identified 26 signaling pathways affected by splicing events detected in AML patients. Among the most affected pathways in AML are: FLT3, CD13, c-Kit, NOTCH, PI3K and MAPK signaling, pathways involved in regulation of cell cycle, apoptosis, cell transformation, and splicing (Table 3). Similar results were obtained from the analyses of significantly spliced genes in the UHN cohort. We detected 14 pathways affected by aberrant splicing in the UHN cohort, 12 of which were also detected in the DFCI cohort (Table 3). Also, for both cohorts for any given signaling pathway, this analysis identified multiple genes that were differentially spliced in AML as compared to NDs (Data not shown). These findings suggest that the effects of AML-specific splicing on signaling networks are widespread and that the splicing mechanism itself is compromised in AML.

**Validation of genome-wide splicing studies in AML patients**

In order to demonstrate that the exon array data accurately detected transcripts with altered splicing, validation studies were performed on the CD13 gene encoding a myeloid cell surface protein. The CD13 transcripts are among those most often mis-spliced in both patient cohorts that were studied on the exon arrays.
**Evaluation splicing patterns of CD13**

cDNA representing CD13 transcripts was amplified by RT-PCR, separated by agarose gel electrophoresis, and then cloned and sequenced (Figure 2a-b). The analysis identified four CD13 splice variant transcripts: CD13Va, CD13Vb, CD13Vc and CD13Vd (Figure 2b). These CD13 novel splice variant transcripts were not detected in normal CD34+ cells, PB monocytes, or neutrophils (Figure 2a). Sequence alignment analyses showed that CD13 splice variants are results of exon skipping, partial deletion of an exon and/or partial retention of an intron (Figure 2b).

Alignment followed by bioinformatic analysis showed that some of these splicing events did not cause any frame shifts (CD13Vd), while other combinations of aberrant splicing led to frame-shifted transcripts (e.g CD13Va, CD13Vb and CD13Vc). All the CD13 splice variant transcripts retained their original start codons and conserved signal sequences. Also, alignment analysis showed that exons affected by aberrant splicing events mapped to extracellular domains of the CD13 protein sequences (data not shown).

**CD13 splice variant proteins**

To understand if novel splice variant transcripts are expressed as proteins, we engineered CD13-Va-GFP and CD13-Vb-GFP, transfected into HEK293T cells, and confirmed their protein expressions by flow cytometry. The HEK293T cells transfected with CD13-Va or CD13-Vb splice variant transcripts were harvested 72 hours after transfection and stained with anti-CD13-PE antibodies that detect cell surface CD13 proteins. The CD13-Va-GFP or CD13-Vb-GFP positive HEK293T cells showed high expression levels of CD13 compared to the untransfected cells or transfected with
plasmid backbone (Supplementary Figure S1) providing preliminary evidence that CD13-Va and CD13-Vb splice variants encode the proteins. More detailed analysis of splice variants at the protein level and their functional characteristics are underway.

**CD13 novel splice variant expression frequencies in a larger AML patient cohort**

To identify which of the CD13 splice variants were most frequently expressed in AML patients we analyzed expression of these novel variant transcripts by highly sensitive semi-quantitative RT-PCR capillary electrophoresis and DNA fragment analysis in the PB or BM samples from 193 AML patients, in CD34+ cells from BM of eight NDs, and in PB monocytes and neutrophils from four NDs. Examples of this assay is shown in Supplementary Figure S2.

The CD13 novel splice variants were expressed in the majority of AML patients (64.3%, 124 of 193) but were not detectable in the NDs (CD34+ BM cells, PB monocytes and neutrophils, Figure 2a). Four CD13 novel splice variants were expressed in patients with AML in various combinations with CD13 full-length (FL) transcripts, or with other CD13 variants (Figure 3a). Of the 193 patients, 124 patients expressed one or more splice variants in various combinations with CD13FL (117 patients) or with other splice variants of this gene (3 patients), while 47 patients did not express any CD13 transcripts (variants or FL) and 22 expressed only CD13FL transcript. Only four patients expressed CD13 variants alone (3 patients express CD13-Va and 1 patient CD13-Vb). As Figure 3a and Supplementary Table S3 show, expression of two or more splice variants in various combinations was the most common pattern in AML patients.
We also evaluated the expression consistency of novel splice variant transcripts over the course of disease in individual AML patients (Figure 3b-d). CD13-FL and their splice variant transcript expressions were examined in paired samples of 14 patients; samples were taken at diagnosis and remission (3 patients), over the course of refractory disease (4 patients) or diagnosis and relapse (7 patients). A total of 29 samples were collected. This analysis showed specific expression patterns and levels of the novel splice variants at different time points during disease development (Figure 3b-d). At diagnosis, the majority (>80%) of patients expressed CD13 splice variants in various combinations; at remission, expression frequencies of these splice variants were decreased; and at relapse, expression of all splice variants was dramatically increased at which point each patient expressed two of more CD13 variants in various combinations (Figure 3c-d). Similarly, in a group of patients with persistent AML, novel splice variant expression frequencies remained constantly high (Figure 3c).

Discussion

Based on genome-wide AS analyses of AML patients and NDs, in this manuscript we report the presence of widespread splicing changes in AML patients compared to NDs. These splicing events were considered aberrant and specific to AML, because they were not detected in CD34+ bone marrow cells, monocytes, or neutrophils from NDs. Thus, these events may provide potential markers for leukemic stem cells.

Since these AML-specific splicing abnormalities were found not only in the coding region, but also in the 5’- and 3’- untranslated regions (UTR) of the AML transcriptome, we argue that such widespread aberrant splicing in AML leads to altered expression of the genes, which may be
necessary to modulate epigenetic regulatory “circuits” in this disease. Those aberrant splicing events that occur in the 3’UTR of the AML transcriptome will compromise miRNA-dependent gene regulation, whereas splicing events in the 5’UTR will change the composition of translational regulatory elements, which in turn can alter protein expression. Furthermore, aberrant splicing events in the coding regions will modulate a core part of the AML proteome and generate altered proteins. Some of these proteins might act in a dominant negative manner and compromise normal functions of their wild type counterparts. Similar effects have been reported for MDM2, p53, CD44, STAT3 and other genes (32-36).

Similar to our findings, several related studies have reported widespread defects in AS in different types of cancer, including lung, breast, colon, prostate, and brain using the Affymetrix Human Exon 1.0ST Arrays (37-40). This well-adopted platform appears to be more useful than direct high-throughput sequencing using RNA-Seq when limited amounts of tissue are available from patients. Exon array analysis requires only 50 ng of total RNA whereas RNA-Seq requires 1.5 ug RNA(41). Also, even though the dynamic range of RNA-Seq analysis is unparalleled, evaluation of the technical reproducibility of the exon array and RNA-Seq platforms showed that the coefficient of variation for the microarray is lower than that of RNA-Seq (41, 42). Furthermore, human exon array analysis tools are well standardized, short analysis times are achieved, and few computational resources are required as compared to RNA-seq; thus use of exon arrays for screening large patient cohorts can be more practical. In addition to identifying novel transcripts, RNA-Seq analysis detects single nucleotide genetic variants in screened samples, which is one of the powerful options of this technology; however, the tools for RNA-Seq data analysis are not fully mature and capturing low abundance transcripts is compromised. Thus, microarrays remain a more useful and accurate
platform for screening large number of patient samples at this point. More details regarding the pros and cons of microarrays and RNA-Seq have been reported recently (41, 42). The combined use of both technologies holds much promise to identify genome-wide disease related aberrations.

Our study identified that a majority of the transcripts in AML patients are subject to complex splicing events that include partial exon deletions and partial or complete intron retentions in various combinations. These events are detected on single transcripts and are highly recurrent as demonstrated for the CD13 gene in this manuscript, and for the NOTCH2 and FLT3 genes elsewhere, in 193 AML patients. Similar patterns of aberrant splicing have been observed for oncogenes or genes associated with other malignant phenotypes; in these malignancies, splicing alterations produced novel transcripts encoding functional proteins (17, 19, 31, 43, 44).

Our results suggests that splicing is a global phenomenon in AML, raising the question as to whether modulated splicing in AML is a sporadic process or is organized/associated with the disease status. In longitudinal studies carried on limited number of AML patient samples we noted significant differences with respect to expression frequencies of the splice variants at different time points of disease progression. Our findings suggest that splicing events are increased at diagnosis and relapse and “normalized” during remission. Extended studies are in progress to confirm our findings, which well allow us identification novel biomarker reflecting disease status, and consequently identify selective druggable targets.

Finally, the splicing aberrations we report in this manuscript might alter mechanisms of splicing via modulation of splicing factors and/or proteins. In this context, it is noteworthy that we identified
novel splice variants of multiple splicing factors and proteins in our study. Detailed studies are underway to elucidate functional consequences of these aberrations in patients with AML. Overall our findings are consistent with published studies and underscore a key role for “cancer-specific” alternative splicing in pathogenesis in AML.

**Acknowledgments.** We thank Zachary Hunter (DFCI) for assistance with CIRCOS tool.

**References**


Adamia et al. 23 Aberrant Splicing in AML


Table 1. Annotation analyses of aberrantly spliced transcripts identified in AML

<table>
<thead>
<tr>
<th>AML patient cohort</th>
<th>Splicing events mapped to the translated regions of the genome (probe-sets)</th>
<th>Splicing events mapped to the untranslated regions (5’UTR, 3’UTR) of the genome (probe-sets)</th>
<th>Splicing events mapped to partial transcripts/cDNAs genome-wide (probe-sets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFCI</td>
<td>6,573 of 8,648 (76 %)</td>
<td>1,785 of 8,648 (20.6 %)</td>
<td>290 of 8,648 (3.4 %)</td>
</tr>
<tr>
<td>UHN</td>
<td>8,794 of 10,932 (80 %)</td>
<td>1,842 of 10,932 (17 %)</td>
<td>296 of 10,932 (3 %)</td>
</tr>
</tbody>
</table>

This table is a summary of Affymetrix GeneChip Human Exon 1.0 ST array results. Reported study includes two independent cohorts (DFCI and UHN) of patients with AML. Results were obtained as compared to normal donor samples (P<0.05).

Table 2. Splicing event frequencies detected genome-wide in patients with AML

<table>
<thead>
<tr>
<th>Splicing events</th>
<th>DFCI cohort (n=31)</th>
<th>UHN cohort (n=35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of genes in each patients</td>
<td>350 of 2,251 (range 135-786)</td>
<td>506 of 2,004 (range 111-1152)</td>
</tr>
<tr>
<td>Any give gene is spliced in average number of patients</td>
<td>14 (range 1-29)</td>
<td>18 (range 1-34)</td>
</tr>
<tr>
<td>Splice variants previously reported</td>
<td>47% (1,062 of 2,251)</td>
<td>46% (922 of 2,004)</td>
</tr>
<tr>
<td>Novel splice variants</td>
<td>53% (1,189 of 2,251)</td>
<td>54% (1,082 of 2,004)</td>
</tr>
</tbody>
</table>

In this table splicing events frequencies detected in the DFCI and UHN cohorts were evaluated as compared to normal donors (P<0.05).
Table 3. Signalling pathways affected by splicing enriched in AML patients of the DFCI and UHN cohorts

<table>
<thead>
<tr>
<th>Term-Pathways</th>
<th>DFCI cohort</th>
<th>UHN cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P&lt;0.005</td>
<td>Fold Enrichment</td>
</tr>
<tr>
<td>Lysine degradation</td>
<td>6.69E-05</td>
<td>2.7</td>
</tr>
<tr>
<td>AML-FLT3 &amp; c-Kit</td>
<td>1.40E-04</td>
<td>2.4</td>
</tr>
<tr>
<td>Chronic myeloid leukemia</td>
<td>1.82E-05</td>
<td>2.3</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>5.63E-06</td>
<td>2.3</td>
</tr>
<tr>
<td>Lysosome</td>
<td>9.66E-08</td>
<td>2.3</td>
</tr>
<tr>
<td>Aminoacyl-tRNA biosynthesis</td>
<td>3.38E-03</td>
<td>2.3</td>
</tr>
<tr>
<td>Notch signaling</td>
<td>1.84E-03</td>
<td>2.3</td>
</tr>
<tr>
<td>B cell receptor</td>
<td>5.50E-05</td>
<td>2.2</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>6.43E-05</td>
<td>2.1</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>5.71E-04</td>
<td>2.1</td>
</tr>
<tr>
<td>Phosphatidylinositol signaling</td>
<td>8.78E-04</td>
<td>2.0</td>
</tr>
<tr>
<td>Glycolysis / Gluconeogenesis</td>
<td>4.74E-03</td>
<td>2.0</td>
</tr>
<tr>
<td>Renal cell carcinoma</td>
<td>2.37E-03</td>
<td>2.0</td>
</tr>
<tr>
<td>Adipocytokine signaling</td>
<td>3.19E-03</td>
<td>2.0</td>
</tr>
<tr>
<td>Adherens junction</td>
<td>1.60E-03</td>
<td>1.9</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>1.18E-03</td>
<td>1.9</td>
</tr>
<tr>
<td>Ubiquitin mediated proteolysis</td>
<td>6.30E-05</td>
<td>1.9</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>6.38E-06</td>
<td>1.8</td>
</tr>
<tr>
<td>Leukocyte transendothelial migration</td>
<td>6.44E-04</td>
<td>1.8</td>
</tr>
<tr>
<td>Fc gamma R-mediated phagocytosis</td>
<td>3.25E-03</td>
<td>1.8</td>
</tr>
<tr>
<td>Neurotrophin signaling</td>
<td>7.91E-04</td>
<td>1.8</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>1.89E-03</td>
<td>1.7</td>
</tr>
<tr>
<td>Regulation of actin cytoskeleton</td>
<td>8.58E-05</td>
<td>1.7</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>1.21E-03</td>
<td>1.5</td>
</tr>
<tr>
<td>Pathways in cancer</td>
<td>6.33E-05</td>
<td>1.5</td>
</tr>
<tr>
<td>MAPK signaling pathway</td>
<td>3.00E-03</td>
<td>1.4</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spliceosome</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the Tables pathways in bold letters have been identified as enriched in both, DFCI and UHN cohorts.
Figure legends

**Figure 1a-c. Aberrant splicing event frequencies detected in patients with AML**; CIRCOS plots (Figure 1a-c) describe distributions of the genome-wide splicing events in patients with AML. Outer circles represent the human reference genomes; spliced gene frequencies in patients are shown as dot plots, with each dot representing a single spliced gene, which is spliced in patients recurrently but is absent in NDs. The greater the distance of a given dot from the outer circles, more frequently the gene is spliced in AML patients. The inner circle represents chromosomes. More than a hundred differentially and recurrently spliced genes are detected in chromosomes 1, 2 and 17 of the DFCI cohort (a), and in chromosomes 1-3, 11, 12, 17 and 19 of the UHN cohort (b). CIRCOS plot was generated based on the list of the genes we have identified as spliced in at least 30% or more patients from DFCI cohort; 75% (262 of 348) of these spliced genes were frequently (range 2-34 patients) spliced in the UHN cohort (c).

**1d. Genome-wide splicing patterns detected in patients with AML.** Based on probe-set expression levels a complex splicing pattern as detected in patients with AML compared to NDs, summarized in this figure as “Pattern 1”— full or partial exon(s) skipping (sequence exclusions), and “Pattern 2”—full or partial cryptic exon or intron retentions (sequence inclusions). On these diagrams dark black boxes represents exons and solid lines introns, on the exons partial deletion are marked by the arrows, and intron retentions are shown as transparent gray boxes.
Figure 2. Exon array validation studies

Figure 2a. CD13-FL and splice variant transcript amplifications for cloning and sequencing analyses. CD13 novel splice variant transcripts were amplified from AML patients and normal donor samples. After cycling RT-PCR products were subjected to 1% agarose gel electrophoresis. Bands representing different splice variant transcripts were cut out and cloned and sequenced to identify modes of splicing.

2b. Splicing modes of CD13 identified in patients with AML. CD13 splicing pattern was identified by cloning and sequencing analyses carried out in patients with AML. Obtained CD13 novel variant sequences were aligned with the published sequence for human CD13 mRNA (gi:178535 NCBI). The CD13Va and CD13-Vb splice variants are the result of complete deletions of exon 4 (140bp) or exons 4 (140bp) and 6 (155bp), respectively, while CD13Vc is a result of a partial deletion of exon 4 (44bp) and a complete deletion of exon 6 (155bp). The CD13-Vd splice variant transcript is result of activation of a cryptic 3’ splicing site on intron 4. This splicing aberration causes the partial retention of intron 4 (132bps). The splicing event does not lead to a frameshift, and the native stop codon is retained. However, as a result of these aberrations 140bp is spliced out from CD13Va, 295bp from CD13Vb and 199bp from CD13Vc transcripts. Thus, all these transcripts are subjected to frameshift.

Figure 3. CD13 splice variants are commonly expressed in patients with AML and are associated with disease status. This figure display overall expression patterns of CD13-FL and their splice variants transcripts. On the figure the x-axes display 193 patient samples and the y-axes Relative Fluorescent units (RFU). PCR product RFU=Log _2RFU. RFU is a unit of measurement calculated relative to the size standards included in each reaction. For relative level determination,
product levels were kept below 3500 RFU and size standard levels were within 500-800 units as recommended by the manufacturer. All calculations and instrument calibration was done according to Applied Biosystems recommendations. a) CDF13 and splice variant expression profiles in 193 AML patients. b-d) display overall expression patterns of CD13 FL and novel splice variants transcripts over the course of the disease in 14 patients. Patient samples were collected and expression profiles CD13 FL and novel splice variant transcripts were evaluated: b) at diagnosis (Dx) and remission (REM), c) at relapse (REL), diagnosis (Dx) and during persistent disease (PER), d) at diagnosis (Dx) and relapse (REL). A total 29 AML patient samples were collected including: ten samples obtained at diagnosis (Dx 1- Dx 3 and Dx 8 - Dx 14); three samples obtained at remission (REM 1- REM 3); six samples obtained during regular visits over the course of refractory disease (PER 4, 5, 7); nine samples obtained at relapse (REL 6, REL 8- REL 14). Samples take at 1st and 2nd relapse or visit are marked as -1 or -2.
Figure 1. Alternative splicing events detected in patients with AML compared to NDs

a. DFCI cohort

29% (2,251) genes are differentially spliced (P<0.05)

b. UHN cohort

28% (2,004) genes are differentially spliced (P<0.05)

c. Reproducibility of the exon array results

d. Splicing patterns detected in AML patients

Pattern 1- sequence exclusion

Pattern 2- sequence inclusion
Figure 2a. CD13-FL and splice variant transcript amplification for cloning and sequence analysis

AML patient samples

Normal donor samples

2b. Splicing modes of CD13 in patients with AML

CD13Va

CD13Vb

CD13Vc

CD13Vd
Figure 3. CD13 splice variants are commonly expressed in patients with AML and are associated with disease status

a) CDF13 and splice variant expression profiles in 193 AML patients

b) Expression profiles at diagnosis (Dx) and remission (REM)

c) Expression profiles at relapse (REL), diagnosis (Dx) and during persistent disease (PER)

d) Expression profiles at diagnosis (Dx) and relapse (REL)
A Genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets

Sophia Adamia, Benjamin Haibe-Kains, Patrick M Pilarski, et al.

Clin Cancer Res  Published OnlineFirst November 27, 2013.