RPPA-profiling in pediatric acute leukemia: pattern recognition and therapeutic guidance

Hoff, Fieke

DOI: 10.33612/diss.168765254

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2021

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 11-12-2022
Chapter 7

Clinical relevance of proteomic profiling in de novo pediatric acute myeloid leukemia: A Children’s Oncology Group study

Hoff FW¹, van Dijk AD¹, Qiu Y², Hu CW³, Ries RE⁴, Ligeralde A⁵, Jenkins GN⁶, Gerbing RB⁷, Gamis AS⁸, Aplenc R⁹, Kolb EA¹⁰, Alonzo TA⁷, Meshinchi S⁴, Qutub AA¹¹, de Bont ESM¹, Horton TM⁶, Kornblau SM²,†

1. Department of Pediatric Oncology/Hematology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands;
2. Department of Leukemia, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA;
3. Department of Bioengineering, Rice University, Houston, TX, USA;
4. Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA;
5. Biophysics, University of California, Berkeley, CA, USA;
6. Department of Pediatrics, Texas Children’s Cancer Center, Baylor College of Medicine, Houston, TX, USA;
7. University of Southern California, Los Angeles, CA, USA;
8. Department of Hematology-Oncology, Children’s Mercy Hospitals and Clinics, Kansas City, MO, USA;
9. Division of Pediatric Oncology/Stem Cell Transplant, Children’s Hospital of Philadelphia, Philadelphia, PA, USA;
10. Nemours Center for Cancer and Blood Disorders,
11. Department of Biomedical Engineering, The University of Texas at San Antonio, USA.

†Senior co-author

Under review
Key points

- Pediatric acute myeloid leukemia could be classified into nine subgroups based on protein expression signatures;
- Proteomic classification could enhance outcome risk stratification when combined with known risk-stratifiers.
Abstract

Pediatric acute myeloid leukemia (AML) remains a fatal disease for at least 30% of patients, stressing the need for improved therapies and better risk stratification. As proteins are the unifying feature of (epi)genetic and environmental alterations, and are often targeted by novel chemotherapeutic agents, we studied the proteomic landscape of pediatric AML. Protein expression and activation were assessed in 500 patient and 30 healthy CD34+ samples, using the reverse phase protein arrays with 296 strictly validated antibodies. The multi-step “MetaGalaxy” analysis methodology was applied and identified nine protein expression signatures (SIG). SIGs were formed based on strong recurrent protein expression patterns. SIGs were associated with cytogenetics and mutational state, and with both favorable or unfavorable prognosis. Analysis based on treatment (i.e., ADE vs. ADE plus bortezomib (AEB)) identified three SIGs that did better with AEB vs. ADE. When SIGs were studied in the context of genetic subgroups (i.e., CEBPA mutation, AAML1031 low- or high-risk groups), SIGs were independently prognostic after multivariate analysis, suggesting a potential value for proteomics in combination with current classification systems. Universally increased (n = 7) or decreased (n = 17) protein expression was observed across all SIGs. Post-treatment protein expression at 10 and 24 hours revealed predominantly changes in DNA damage and cell cycle regulators, but were not specific to treatment arm. Within each SIG, expression of certain proteins significantly differentially expressed from normal could be identified, forming a hypothetical platform for personalized medicine.
Introduction

Pediatric acute myeloid leukemia (AML) is a heterogeneous disease resulting from clonal expansion of myeloid precursors that have lost the ability to normally differentiate.\(^1\) Despite improvements in outcome, 5-year overall survival (OS) approximates 70% and serious long-term complications are common among survivors.\(^2\) With the exception of acute promyelocytic leukemia, FLT3-ITD mutations, and mixed phenotype acute leukemia, pediatric AML has been treated as a homogeneous disease, as therapy does not differ based on the underlying mutations.

Many genetic “drivers” have been implicated in pediatric AML disease pathology and risk stratification. However, only a minority of these drivers have been exploited by targeted therapeutic interventions.\(^3\) Current risk-stratification considers genetic abnormalities (e.g., inv (16), monosomy 7), but otherwise relies on early response to therapy (minimal residual disease status). Since many drivers may prove to be directly “undruggable”, targeting downstream proteins may be efficacious. Doing so requires knowledge of the proteomic landscape that emerges from the combined “net” consequences of genetic and epigenetic events. Little is known, however, about the proteomic landscape in pediatric AML. Furthermore, most targeted therapies act by interfering with protein function, rather than by affecting gene expression. Improved understanding of the pediatric AML proteomic landscape might enhance risk stratification prior to therapy commencement and ultimately guide the selection of therapies against targetable molecular lesions.

Genomic mutations influence cellular physiology via altered protein abundance or activity, but several factors diminish the correlation between genetic alterations and protein effects, including the general lack of correlation between cellular messenger RNA (mRNA) abundance and protein expression,\(^4,5\) and the inability to assess protein post-translational modifications (PTM) with genomic techniques. In other tumor types, protein quantification commonly influences diagnosis, classification and therapy (e.g., estrogen and progestin receptors, programmed cell death ligand 1).\(^7,8\) Thus far, proteomics has not been used to guide AML therapy.

We previously performed a pilot study of the proteomic landscape examining 194 proteins in 95 \textit{de novo} pediatric AML patients using an approach that recognized protein expression patterns within protein functional groups (PFG). After determining the characterization of each
patient PFGs, we built higher order structures based on strong correlations between PFG-patterns, recognizing eight protein expression signatures (SIG) that were prognostic.\textsuperscript{9} Here, we use this same approach to prospectively examine 500 pediatric AML patients treated on a Children’s Oncology Group (COG) randomized phase 3 clinical trial (AAML1031). This trial hypothesized that the addition of the proteasome inhibitor (PI) bortezomib could improve therapy with ADE-based therapy (Ara-C (cytarabine), daunorubicin, and etoposide). This trial closed early when it was determined that adding bortezomib to ADE (ADEB) did not improve 3-year event-free (EFS) or 3-year OS across the entire patient group. We aimed to 1) validate the ability to classify pediatric AML patients based on the proteomics in a larger cohort, with significantly more protein targets, 2) determine if protein classification could enhance risk stratification, 3) identify patients that could benefit from ADEB, and 4) identify additional targets for potential combination therapy.

Methods

Patient’s samples
Peripheral blood (PB) samples from 500 \textit{de novo} pediatric AML patients participating in the COG AAML1031 (#NCT01371981) phase 3 clinical trial \((n = 483)\) or in older clinical trials \((n = 17)\), as well as 30 control CD34\textsuperscript{+} bone marrow (BM) samples from healthy donors (pediatric \((n = 20)\) and adults \((n = 10)\)) were collected during routine diagnostic assessments between July 2011 and February 2017. The 483 samples were statistically representative of the 1097 patients enrolled in the AAML1031 trial.\textsuperscript{10} Samples were collected in Cell Save preservation tubes, or in heparin tubes when Cell Save tubes were not available.\textsuperscript{11} Samples were collected before the start of chemotherapy \((n = 500)\), 10 hours (hr) \((n = 463)\) and 24hr \((n = 466)\) after the start of induction chemotherapy. The 10 and 24hr samples were called “post treatment”. Samples from older trials were included to enable comparison with previous RPPAs.\textsuperscript{9} Written informed consent was obtained in accordance with \textit{Declaration of Helsinki} and local institutional policies.

Outcome data was restricted to 410 of the 483 patients enrolled on the AAML1031 trial. Outcome for 69 patients treated with ADE was not determined after DSMB trial closure, and four patients that did not meet eligibility criteria. Of the 410 patients, 200 received standard ADE induction therapy, including 36 with known FLT3-ITD mutations that also received sorafenib (ADES), while 210 patients received ADEB. Outcomes of ADE and ADES treated
patients were statistically similar, and were therefore combined for this analysis. Induction therapy produced complete remission (CR) by the end of induction II in 348 (85%) patients, 31 (7.6%) were refractory or died early. Relapse occurred in 156 (45%) patients, and 286 (70%) were still alive with a median follow-up of 4.4 years (range 0.3 to 7.5 years). Outcome data was calculated as published previously.10 As part of the routine COG clinical work-up mutation data were available for CEBPα, NPM1, KIT (exon 8 and 17) and FLT3-ITD.12

RPPA methodology
The methodology and validation of the RPPA technique, including antibody validation, are fully described elsewhere.13-15 Slides were probed with 301 validated antibodies, including 229 antibodies to total proteins, 63 targeting phosphorylation sites, five targeting histone methylation sites and four targeting cleaved forms of caspases or Parp1. A “Rosetta Stone” of the protein nomenclature used, antibody names, primary and secondary dilutions, and source is shown in supplemental Table S1. Stained slides were analyzed using Microvigene® Software version 3.0 (VigeneTech, Inc., Carlisle, MA) to produce quantified data.

Data processing and normalization
SuperCurve algorithms were used to generate a single protein concentration value from the five serial dilutions in log2 format.16 Loading control17 and topographical normalization18 procedures were performed to account for protein concentration and background staining variations. Replicates-based normalization19 was used to align samples from two different slides. The median Pearson’s correlation coefficient between the two slides was 0.89, although two antibodies (CDX2, PSMB9) were excluded for poor correlation. Median expression levels of the normal BM CD34+ samples were subtracted from the expression in the patient samples to set the median of the control samples to zero. Protein expression differences between paired samples collected in different collection tubes or between different cell types (i.e., bulk vs. CD3-/CD19-) were assessed and three antibodies were excluded (H2AX-pSer140, MYH9-pSer1943, PRKAA1_2-pTyr172) for differences. This yielded a final of 296 antibodies used for analysis.

Quality control
Samples with a highly significantly greater number (>20%) of proteins with extremely high or low expression (≥ 3SD from the median) might have problematic “handling” effects.11 This metric identified six pre-treatment samples with quality control issues which were excluded.
Gene expression profiling
Ribodepleted RNA-sequencing expression data was generated for 390 of the 500 RPPA-patients as described.\textsuperscript{20}

Transcriptome mutation calls
Singe-nucleotide variants and indels were analyzed utilizing a custom pipeline (British Columbia Cancer Agency, British Columbia, Canada). Targeted alignment of custom in silico probes were produced for 1160 physical positions across 166 genes to detect a “hit count” of variant reads. Variant “hit counts” were compared with positive controls (patient samples with validated mutations) and negative controls (normal BM). A cut-off of seven variant hit counts was used.

Computational analysis
The “MetaGalaxy” computational analysis was done as published previously (Figure 1).\textsuperscript{9,21,22} This analysis takes a multi-step approach, first separating proteins into 31 PFGs, based on prior knowledge from the literature or strong correlations within this dataset. A list of proteins in each PFG is provided in supplemental Table S1. To identify the optimal number of patient subgroups with similar correlated protein expression in each PFG, defined as a “Protein Cluster” (PC), k-means\textsuperscript{23} coupled with the Progeny Clustering algorithm\textsuperscript{24} was applied. Linear discriminant analysis was performed to determine which cluster (hereafter called PC-1) was statistically most similar to normal CD34\textsuperscript{+} samples. Collective PC-memberships for each patient were combined into a binary matrix. Block clustering\textsuperscript{25} was used to search for correlations between PCs ("Protein Expression Constellation" (CON)), and to cluster patients with similar CON-membership into a SIG. An optimization calculation\textsuperscript{9,22} was used to determine the optimal number of CONs and SIGs. Robustness of the protein CONs and SIGs were tested on a training set ($n = 355$) and test set ($n = 145$). Sets were created by using random sampling.\textsuperscript{26}

Proportions between protein SIGs and categorical variables were compared using Pearson’s Chi-square test. Continuous variables were compared using the Kruskal-Wallis test. Networks were constructed based on known protein-protein interactions from the STRING database (combined score $> 0.9$)\textsuperscript{27} and computationally reconstructed interactions from the RPPA data using graphical lasso\textsuperscript{28} and StARS.\textsuperscript{29} Networks were visualized using Cytoscape (version 3.8.0).\textsuperscript{30,31} Proteins significantly up or down regulated compared to the normal CD34\textsuperscript{+} samples
were identified using the Wilcoxon signed-rank test with a Bonferroni adjusted $p$-value. All statistical analyses were performed in R version 1.3.959 2009-2020 (RStudio, Inc., Boston, MA) or SAS version 9.4 (SAS Institute, Inc., Cary, NC).

**Figure 1. Computational MetaGalaxy work flow.** Multi-step analysis that starts with relative expression of 296 protein targets. Proteins were allocated into 31 PFGs based on known their known functionality or strong correlation from the data set. Progeny clustering algorithm identified an optimal number of PCs in each PFG. Block clustering was applied to a binary matrix indicating PC-membership for all patients, and identified the existence of CONs (i.e., strongly correlated PCs from various PFGs). Patients that expressed similar combinations of CONs were defined as SIG.

### Results

**Correlation of protein expression identifies functional protein patterns within a PFG**

The 296 proteins that were analyzed in this study were allocated into 31 protein functional groups (PFG) (e.g., autophagy, cell cycle, metabolism, etc.). Clustering analysis identified an optimal number of patterns (protein clusters (PC)) of similar correlated protein expression (negatively or positively) between patients within each PFG. PC numbers ranged from 3 to 5 per PFG (Figure 2A) and the total number of PCs was 116.

PCA was applied to graphically compare patient PC expression patterns to non-malignant CD34$^+$ cells. Although the overall proteomic profiles of the pediatric AML patients were distinct from normal CD34$^+$ cells (supplemental Figure S1), we found overlapping “normal-like” expression patterns for 31 (27%) of the PCs (Figure 2A). In four PFGs, more than one cluster was defined as “normal-like”, and in five PFGs no “normal-like” pattern was found. PCs without dominant co-localization to CD34$^+$ samples on the PCA plot were defined as “leukemia-specific”.

---

**Proteins**

(n=296)

**Protein Functional Group**

(n=31)

**Protein Cluster**

(n=116)

**Protein Constellation**

(n=12)

**Protein Signature**

(n=9)
Figure 2. PFG classification and similarity to normal CD34+ cells. (A) Progeny clustering algorithm was applied to the 31 PFG and identified an optimal number of PCs. PCs were compared with normal CD34+ cells using PCA and classified as either “normal-like” or “leukemia-like”. “Normal-like” patterns were denoted with a hatch pattern, “leukemia-specific” patterns with fill pattern. (B, C) PCA analysis with each PC being assigned to a color within the PFG to note its similarity, or difference, to the normal CD34+. Two examples of PCA mapping include (B) cell cycle and (C) mTOR-signaling. Normal CD34+ samples are noted by small black squares and the large black circles. There was no colocalization with CD34+ cells for cell cycle.

To visually map how proteins interact with other PFG core-members and RPPA dataset proteins, networks were generated for each PC. Proteins were connected if they were known to interact with other proteins based on the STRING database or correlation in our data set. Median protein expression was calculated for each PC and overlaid onto the networks to visualize relative expression. Networks can be viewed online at: http://leukemiaproteinatlas.org/pediatric-aml/.
As an example, the PFG “Heat shock” comprises five antibodies recognizing total proteins and two recognizing phosphorylated proteins. We discerned four “Heat shock” PCs (Figure 3). Expression levels in PC-1 were identified as most “normal-like”. Protein networks were generated for the seven heat shock protein members. For PC-2 to PC-4, expression of HSPA1A_L and HSBP1-pSer\textsuperscript{82} changed between the four PCs, as well as AKT1S1 (connected to HSF1 and HSF1-pSer\textsuperscript{326}) and CAV1 (connected to HSP90AA1_B1 and HSPA1_L1), showing that associated nodes correlate with core-protein PFG-members. It is important to note that when heatmaps are typically presented in other analyses they have been median normalized to 0 with the variance set from -1 to +1, so that all variables are shown as ranging from the minimum to the maximum of the scale (color range). In contrast, our expression levels are shown relative to normal, and therefore may only use a portion of the scale (color) range.

**PCs correlate with clinical outcome**

To evaluate the effect of protein expression alone on prognosis, PCs were correlated to outcome. Seven (23%) PFGs were found to be significantly associated with outcome (supplemental Figure S2). For instance, overall, heat shock PCs were prognostic for OS ($p = 0.004$), EFS ($p = 0.0009$), and RR ($p = 0.0016$), in patients treated with either ADE or ADEB (ADE: OS, $p = 0.0035$; EFS, $p = 0.0097$; RR, $p = 0.021$; ADEB: OS, $p = 0.0002$; EFS, $p < 0.0001$; RR, $p = 0.0009$). When we compared outcome after ADE to ADEB, PC-2 significantly benefitted from ADEB ($n = 131$, 5-year OS, 54% vs. 81%, $p = 0.00087$), whereas PC4 did worse ($n = 35$, 100% vs. 67% at 5-year OS, $p = 0.019$). Bortezomib had no effect on PC-1 ($n = 91$), which was an unfavorable prognostic indicator after both ADE and ADEB; this cluster was characterized by high-HSF1-pSer\textsuperscript{326} and HSB1-pSer\textsuperscript{82}. Survival curves for EFS and RR are shown in supplemental Figure S3.

We previously published that low-HSF1-pSer\textsuperscript{326} was associated with a better outcome after ADEB.\textsuperscript{10} In the current data, this effect was true for PC-2 and PC-3 (characterized by low-HSF1-pSer\textsuperscript{326}), but was absent in PC4, which also had concomitant increased expression of HSPA1A_L, emphasizing that a simultaneous integrated analysis of multiple proteins, rather than a single protein, could identify more detailed protein expression patterns and better characterize subpopulations that could benefit from the addition of novel agents.
Figure 3. Analysis of the Heat Shock PFG. (A) Optimal number of four PCs was identified as shown by the heat map (annotation bar: PC-1 (red), PC-2 (pink), PC-3 (yellow) and PC-4 (light green). (B) PCA-analysis shows relative expression of the four PCs in relation to the normal CD34+ cells (black squares). (C) Protein networks visualizing median expression of PFG core-protein members (large nodes) in relation to associated proteins (small nodes) for the four PCs. Interactions between nodes are based on the literature (…), RPPA data (−−−) or both (−). Colors represent the median expression of each individual protein within a given PC. Green circles indicate change in HSPA1A_L expression and blue circles the change in HSBI-pSer326 across the four PCs. Red circles show the negative correlation between AKT1S1 and HSFI-pSer326, and the positive correlation between HSPA1A_L and CAV1. (D) Outcome data stratified by PC. Outcome for all patients (upper left panel), ADE treated (middle left panel) and ADEB (middle right panel). Lower panels show outcome data in PC-2 (left, pink) and PC-4 (right, green) for patients treated with ADE (dark blue) vs. ADEB (red).
Figure 4. “MetaGalaxy” analysis identified the existence of 12 CONs and 9 SIGs. Block clustering was applied to a binary matrix constructed from 116 PCs from 31 PFGs. Each column indicates one patient (n = 500) and its PC-membership. This identified the existence of 12 CONs (horizontally); i.e., PCs that strongly correlated with each other. Patients with a similar pattern of CONs were defined as SIG (vertically). Annotation bar shows 9 SIGs, cytogenetics (t(8;21) (red), inv(16) (yellow), normal karyotype (green), MLL (11q23) rearrangement (black), -5,-7,+8 abnormalities (light blue), other (gray)), and mutational status for CEBPA, FLT3-ITD, FLT3-ITD high allelic ratio (≥ 0.1), NPM1 and KIT (exon 8 and 17) (wildtype (blue), mutated (yellow)). Those with unknown mutational status are noted in grey.

Recurrences in protein patterns classify patients into 9 SIGs
To obtain a more systemic understanding of the 116 identified PC, pattern recognition of the relations between PCs from various PFGs was done using co-clustering. PC-memberships for the 31 PFG’s were assigned to all 500 pediatric AML patients and compiled in a binary matrix (“MetaGalaxy”) (Figure 4). Optimization calculations identified twelve patterns of recurrent (i.e., correlated) PC, defined as Protein Constellations (CON). From this, nine protein expression signatures (SIGs) were defined as clusters of patients that expressed similar combinations of CON. Robustness was tested and showed high reproducibility (p < 0.0001)
(supplemental Figure S4). None of the CONs were predominantly associated with the “normal-like” clusters ($p = 0.200$). The PCs in each CON are listed in supplemental Table S2.

**Protein expression partially correlates with cytogenetics and mutational state, but not with gene expression**

Correlation between mRNA expression from RNA-sequencing and RPPA protein abundance was determined for 205 total-proteins in 390 samples, with a mean correlation of 0.17. This low correlation is very similar to that seen in several previous studies.$^{5,32}$ Thirty-four (17%) proteins were negatively correlated, while 83% were positively correlated (supplemental Figure S5).

We found associations between SIGs and cytogenetics and mutation states. Data was available for CEBPA, NPMI, FLT3-ITD, KIT (exon 8 and 17), KRAS, NRAS, GATA2, PTPN11, MYH11 IDH1/2. Mutations present in $\leq 10$ of the patients were not analyzed (Table 1). Translocation $t(8;21)$ was more frequently detected in SIG-4 (35% vs. 6% overall) ($p = 0.001$). Inversion (16) was associated with SIG-1, 6 and 8 (25%, 25%, 30% vs. 14% overall), but scarcely seen in SIG-2, 3, 5 and 7 (2%, 5%, 0% and 3%) ($p < 0.001$). Normal karyotype (diploid) was enriched in SIG-3, 5 and 6 (59%, 42%, 38% vs. 28% overall) that shared CON-3. Those three also had the highest frequencies of CEBPA mutation (SIG-3) and FLT3-ITD (SIG-3, 5 and 6). While the MLL-rearrangement (11q23) did not exclusively result in the SIG-7 protein expression pattern, 85% of the patients harbored the MLL-rearrangement (vs. 18% overall). KIT mutations were mostly in SIG-4 and 6 ($p = 0.004$), and NRAS and MYH11 in SIG-1 ($p = 0.024$, $p = 0.037$, respectively). Although only 3% ($n = 12$) of the patients were GATA2 mutated, 19% of SIG-3 had this mutation. Patients with fusion gene NUP98-KDM5A ($n = 4$) were all present in SIG-4 ($p = 0.007$). NPM1, KRAS, PTPN11 and IDH were not associated with specific SIGs.

**Correlation with patient characteristics and clinical variables**

Patients aged $\leq 1$ year at time of diagnosis were most frequently clustered in SIG-7 to 9, which are associated with CON-5. Low WBC ($\leq 100,000$ cells/uL) strongly correlated with CON-7, with 81-89% of the patients in SIG-1 to 4 and 9 having a low WBC vs. 76% overall ($p = 0.001$). Gender, ethnicity, race and CNS-status at time of diagnosis were not associated with any SIG or CON (supplemental Figure S6).
**SIGs provide prognostic information**

SIGs were associated with response to therapy (Figure 5) with greater spread in 5-year OS in ADEB vs. ADE treated cases (supplemental Figure S7). Similar to what was previously done with cytogenetics prognostication, we identified SIGs associated with favorable (SIG-3), intermediate (SIG-1, 4, 6, 8, 9) and unfavorable (SIG-2, 5, 7) overall prognosis (OS, \( p = 0.007; \text{EFS, } p = 0.046; \text{RR, } p = 0.045 \)). Unfavorable SIGs remained an independent prognostic factor using multivariate cox regression analysis in OS, EFS and RR (Table 2A). SIGs with poorest prognosis in ADE were SIG-5 to 8, all characterized by CON-4-membership. Addition of bortezomib was beneficial for 5-year OS in SIG-6 (62% vs. 84%, \( n = 41, p = 0.07 \)) and SIG-8 (56% to 79%, \( n = 72, p = 0.05 \)), and a similar trend was observed in SIG-7. As SIG-6 to 8 were most strongly associated to CON-11, we compared ADE vs. ADEB in the CON-11 SIGs. OS at 5-years increased from 58% to 78% (\( p = 0.011 \)). Across the 9 SIGs, RR showed a significant dispersion, ranging from 24% to 63% at 5-years CR (\( p = 0.03 \)). In SIG-3, RR decreased from 45% to 11% with ADEB (\( n = 21, p = 0.08 \)), while SIG-6 did worse with ADEB (36% vs. 12%, \( n = 34, p = 0.10 \)). While analysis of CEBPA-mutated patients as a group did not show benefit from bortezomib,\(^1\) none of the patients in SIG-3 relapsed or died after ADEB vs. a 60%-event and relapse-rate (\( n = 3/5, p = 0.039, p = 0.037 \), respectively) with ADE.

**Proteomics augment cytogenetic risk stratification**

We analyzed AAML1031 low-risk patients (defined by inv(16)/t(16;16), t(8;21), NPM1 or CEBPA mutations) separately to determine if proteomics were informative for outcome. We found stratification for EFS and RR by SIG with favorable prognosis in SIG-1, 3, 6 and 9 and unfavorable prognosis in SIG-5 and 7 (OS, \( p = 0.071; \text{EFS, } p = 0.027; \text{RR, } p = 0.014 \)) (supplemental Figure S8). Cox proportional hazards regression models identified unfavorable proteomic-SIGs as a significant independent prognostic factor in multivariate analysis (Table 2B). Within the AAML1031 high-risk patients (i.e., FLT3/ITD+ high allelic ratio, monosomy 5 or 7 or del5q, or MRD > 0.1% at end of induction 1) SIGs were also significantly prognostic.

**Proteins significantly different from normal can be identified for each SIG and CON**

The majority of targeted drugs in development act on protein function. Recognition of proteins with an abnormal expression could identify targets for therapy across AML subgroups. We identified proteins significantly different from normal for each SIG/CON (Figure 6A and 6B, available online at: http://leukemiaproteinatlas.org/pediatric-aml/). As example, two proteins, VEGFR and PARP1, were shown in particular, as they may also function as potential...
Figure 5. Kaplan-Meier survival curves for OS and RR. Left panels: overall survival curves (logrank), right panels: relapse risk (Gray’s statistics). (A) OS and RR stratified by the 9 SIGs. (B) Proteomic risk groups defined as “favorable”, “intermediate”, “unfavorable”. (C) Comparison ADE vs. ADEB in SIG-8 (left) and SIG-3 (right). (D) EFS and RR for CEBPA-mutated patients in SIG-3 in ADE (blue) and ADEB (red) treated patients.
Figure 6. Protein identification significantly different from normal. (A) Protein expression significantly deviated from normal in SIG-3 ($p < 0.05$, and log2 change $\geq 0.50 \log2$). Proteins indicated by red circles are potentially targetable. (B) Relative VEGFR (KDR) (left, potential target for SIG-2 to 4, and SIG-6 to 9) and PARP1 (right, target for patients in SIG-3 to 5) protein expression levels across the 9 SIGs. (C) Twenty-four universally higher ($n = 7$) and lower ($n = 17$) expressed proteins ($p < 0.05$, and log2 change $\geq 0.50$) compared to normal CD34$^+$. VIM (indicated by the red box), was most strongly upregulated across the nine SIGs.

therapeutic target for inhibitory drugs (Figure 6B). Twenty-four proteins were identified as universally higher ($n = 7$) or lower ($n = 17$) expressed across all SIGs with vimentin (VIM) most strongly expressed (Figure 6C).

Proteins involved in cell cycle regulation and DNA damage change following chemotherapy exposure

To assess whether cells would adapt their SIGs differentially following chemotherapy, unsupervised hierarchical clustering of pre-treatment and 10hr and 24hr post-treatment samples was performed. None of the expression patterns were specific to a given time point or treatment arm. Comparison of individual protein expression levels identified 87 (29%) proteins...
that had changed by 10hr post-treatment and 173 (58%) by 24hr. Sixty-seven (77%) proteins were changed at both timepoints (supplemental Table S3), predominantly involved in the TP53 pathway (TP53, MDM4), DNA damage response (ATM, Chek2) and cell cycle regulation (Wee1, CCND3, RB1-pSer).

**Pediatric AML web portal**

The complete RPPA dataset and all of the analyses performed in this study, including those not discussed in this paper are published online at: http://leukemiaproteinatlas.org/pediatric-aml.22

**Discussion**

In this study, to our knowledge the largest proteomic study in pediatric AML, we validated our central hypothesis that the genetic heterogeneity of pediatric AML coalesces into a finite number of recurrent protein expression patterns. Unique to this study is the use of a multi-step protein analysis, which moves beyond individual protein expression and activation, to a combined analysis in functionally related protein groups, and then into a system spanning structure based on strongly correlated PCs. We believe that this is a more holistic approach, utilizing known functional interactions, and is superior to conventional unsupervised clustering which weights all proteins equally and ignores known relationships.

Traditional risk stratification in pediatric AML considers selected cytogenetics and molecular features, and early response to induction chemotherapy, but predicts outcome for only 40% of patients.33 When prognostically similar SIGs were grouped as either “favorable”, “intermediate”, or “unfavorable”, we demonstrated an increase in the prognostic significance for the proteomic data when added to traditional risk stratifying factors in multivariate analysis. SIGs were more strongly predictive when combined with AAML1031 risk groups. This is exemplified by the outcome of patients in SIG-3 harboring the favorable CEBPA mutation: none of six treated with ADEB had an event compared to 3 of 5 patients treated with ADE, demonstrating that adding proteomics to genetic risk-stratification can direct therapy leading to improved outcome.

Importantly, we identified three patient SIGs (SIG-3, 6, 8; 34% of all patients) that benefited from PI therapy, a finding not appreciated by analysis of the entire cohort. This finding suggests that proteomic analysis is able to predict a-priori those who would benefit from a specific
therapy. SIG-6 and 8 both contained CON-11 and were characterized by upregulation of autophagy proteins (ATG3, ATG7, BECN1). We hypothesize that these autophagy effectors are required for bortezomib induced autophagy, given that the ubiquitin-proteasome system has active crosstalk with autophagy, and bortezomib stimulates this compensatory autophagy mechanism resulting in cell death. Notably, in SIG-7, which shares CON-11, a similar but less strong beneficial effect of bortezomib was observed. Unlike SIG-6 and 8, SIG-7 had higher phosphorylation levels of heat shock binding protein 1 (HSPB1-pSer^{82}), a strongly unfavorable prognostic factor in our data, but without a PI benefit. This protein could possibly work by preventing the toxic build-up of misfolded proteins due to bortezomib. SIG-7 had a protein expression profile suggesting of high cell cycle turnover (high CCND3, CDKN1B-pThr^{198}, and RB1-phospho), high WBC, and a high frequency of MLL-rearrangements. This highlights the need to apply a holistic system approach to be able to predict drug responsiveness.

The importance of studying protein expression and activation is highlighted by the low correlation ($r = 0.17$) between protein and mRNA expression and the inability of mRNA data to replicate protein SIGs. The lack of correlation was expected, as the presence of mRNA does not imply that translation is occurring (microRNA could impede it); nor dictate the rate of translation, nor protein longevity post translation. Nonetheless, SIGs were partially correlated with cytogenetics and mutational state. “Driver” mutations would be expected to have a defining effect on the biology/protein expression even though combinations of events might further alter these signals. For instance, the majority of SIG-7 (85%) had MLL-rearrangement and, similar the existence of Ph+-like ALL, this may point toward the existence of MLL-like patients based on proteins.

This study demonstrates that the “hallmarks of cancer” are achieved via different patterns of protein utilization within the defined CONs. As an example of the five apoptosis PFG used in this study, CON-6 demonstrated high BH3 pro-apoptotic member activation (BAX, BBC3, and BCL2L11) and simultaneously high expression of anti-apoptotic BH3-member MCL1, a protein associated with chemotheraphy resistance. CON-11 demonstrated a different pathway with high expression of autophagy proteins, but no associations with the other apoptotic PFG. CON-9 had modest upregulation of autophagy proteins and evidence of increased spontaneous apoptosis proteins (high PARP1-cleavage, apoptosis-occurring PC2, and PC3).
Targeted therapies offer the promise of improved outcome, often with less toxicity, compared to current chemotherapy, but an effective means of matching the appropriate patient to the correct agent hampers implementation. To rationally select drugs for specific patient subgroups, we identified proteins with expression significantly different from normal, hypothesizing that those could be targeted by inhibitory/replacement or (re)activation agents. Supplemental Table S4 lists matched proteins and drugs used in the clinical setting. If validated, rapid real-time classification could enable SIG determination to facilitate initial therapy selection.

We identified 24 universally altered proteins, identifying novel potential targets for all patients. The most highly overexpressed protein was VIM, a protein involved in the epithelial-to-mesenchymal transition (EMT). The role for VIM in AML is uncertain, but there is evidence that EMT occurs in hematological malignancies.\textsuperscript{40,41} Fluvastatin targets VIM via caspase-3 mediated proteolysis,\textsuperscript{42,43} and prior trials in AML demonstrated that the addition of pravastatin to idarubicin and high-dose cytarabine counteracted the chemoresistance associated with the defensive adaptation of increasing cellular cholesterol.\textsuperscript{44,45} MCL1 was also universally highly expressed, most prominently in SIG-7 to 9, which had the highest frequency of infants, an historically chemoresistant population.\textsuperscript{9,46} MCL1 is also strongly associated with chemoresistance to venetoclax. Although clinical trials have evaluated the benefit of (combinational) treatment with venetoclax in adults and relapsed pediatric AML, no studies have yet evaluated the effect of venetoclax in de novo pediatric AML. Our finding predicts for lower efficacy in pediatric de novo AML, and suggest that venetoclax should be combined with MCL1 inhibition agents.

A final feature of this RPPA study was to measure therapy effects on protein expression over time. We expected to find specific treatment-induced changes in protein expression, and different proteins adaptation across the SIGs; however, changes were limited to DNA damage, cell cycle regulation, protein translation and histone modification pathways. This likely reflects the presence of many pre-apoptotic cells trying to repair DNA damage. Single cell proteomics might better profile post-treatment AML heterogeneity and predict which changes are associated with resistance or relapse by identifying “survivor cells”, which cannot be identified in studies of bulk cells.\textsuperscript{47} Proteomics may also enable identification and analysis of stem cell proteomics, which differ from that of bulk cells.\textsuperscript{48-50}
In summary, we confirmed the existence of recurrent protein patterns in pediatric AML that enabled separation of AML patients into recurrent protein SIGs that were prognostic, particularly when combined with known pediatric AML risk factors. We identified SIGs that benefitted from ADEB, and postulate that recognition of abnormal proteins can aid in risk stratification and therapy selection in pediatric, and perhaps adult AML.

Acknowledgements

TMH, RBG, ASG, RA, EAK, and RAA were funded by the NIH COG Grants U10 CA98543, U10 CA98413, U10 CA180886, 1U24 CA196173 and U10 CA180899. TMH was funded by the NCI R01-CA164024, a grant from Hope on Wheels, Hyundai Foundation, the St. Baldrick’s Foundation, and a grant from Takeda Pharmaceuticals. FWH was funded by the Junior Scientific Masterclass (Groningen, the Netherlands).

Conflict of interest

TMH receives research funding from Takeda Pharmaceuticals.
Table 1. Patient characteristics (n = 500)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Variable</th>
<th>Total</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number (n=500)</td>
<td>Count</td>
<td>500</td>
<td>48</td>
<td>55</td>
<td>39</td>
<td>53</td>
<td>85</td>
<td>54</td>
<td>40</td>
<td>83</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Gender (n=498)</td>
<td>Female</td>
<td>49%</td>
<td>56%</td>
<td>48%</td>
<td>46%</td>
<td>49%</td>
<td>56%</td>
<td>37%</td>
<td>48%</td>
<td>52%</td>
<td>47%</td>
<td>0.620</td>
</tr>
<tr>
<td>Age (yr.)</td>
<td>0-1</td>
<td>12%</td>
<td>8%</td>
<td>11%</td>
<td>3%</td>
<td>11%</td>
<td>4%</td>
<td>7%</td>
<td>18%</td>
<td>23%</td>
<td>21%</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>2-10</td>
<td>55%</td>
<td>54%</td>
<td>56%</td>
<td>72%</td>
<td>60%</td>
<td>56%</td>
<td>69%</td>
<td>45%</td>
<td>43%</td>
<td>49%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11+</td>
<td>33%</td>
<td>38%</td>
<td>33%</td>
<td>26%</td>
<td>28%</td>
<td>40%</td>
<td>24%</td>
<td>38%</td>
<td>34%</td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>Ethnicity (n=497)</td>
<td>Hispanic</td>
<td>20%</td>
<td>17%</td>
<td>26%</td>
<td>21%</td>
<td>30%</td>
<td>21%</td>
<td>20%</td>
<td>15%</td>
<td>14%</td>
<td>17%</td>
<td>0.479</td>
</tr>
<tr>
<td>CNS status (at Dx) (n=497)</td>
<td>Positive</td>
<td>40%</td>
<td>42%</td>
<td>28%</td>
<td>21%</td>
<td>42%</td>
<td>47%</td>
<td>47%</td>
<td>33%</td>
<td>47%</td>
<td>36%</td>
<td>0.050</td>
</tr>
<tr>
<td>WBC (at study entry)</td>
<td>&gt;100,000</td>
<td>24%</td>
<td>19%</td>
<td>15%</td>
<td>13%</td>
<td>11%</td>
<td>33%</td>
<td>30%</td>
<td>43%</td>
<td>31%</td>
<td>16%</td>
<td>0.001</td>
</tr>
<tr>
<td>FAB classification (n=339)</td>
<td>M0</td>
<td>2%</td>
<td>0%</td>
<td>3%</td>
<td>4%</td>
<td>7%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>22%</td>
<td>29%</td>
<td>31%</td>
<td>44%</td>
<td>36%</td>
<td>40%</td>
<td>40%</td>
<td>14%</td>
<td>3%</td>
<td>3%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>19%</td>
<td>14%</td>
<td>23%</td>
<td>41%</td>
<td>25%</td>
<td>25%</td>
<td>17%</td>
<td>3%</td>
<td>8%</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>24%</td>
<td>26%</td>
<td>17%</td>
<td>7%</td>
<td>18%</td>
<td>16%</td>
<td>40%</td>
<td>24%</td>
<td>38%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>27%</td>
<td>17%</td>
<td>23%</td>
<td>0%</td>
<td>11%</td>
<td>11%</td>
<td>26%</td>
<td>69%</td>
<td>48%</td>
<td>31%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M6</td>
<td>1%</td>
<td>3%</td>
<td>0%</td>
<td>4%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M7</td>
<td>4%</td>
<td>11%</td>
<td>3%</td>
<td>0%</td>
<td>4%</td>
<td>4%</td>
<td>3%</td>
<td>0%</td>
<td>3%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>AAML1031 Risk Group† (n=485)</td>
<td>High risk</td>
<td>28%</td>
<td>29%</td>
<td>31%</td>
<td>28%</td>
<td>22%</td>
<td>41%</td>
<td>40%</td>
<td>10%</td>
<td>16%</td>
<td>26%</td>
<td>0.002</td>
</tr>
<tr>
<td>Cytogenetics (n=492)</td>
<td>t(8;21)</td>
<td>16%</td>
<td>19%</td>
<td>21%</td>
<td>13%</td>
<td>35%</td>
<td>16%</td>
<td>8%</td>
<td>3%</td>
<td>10%</td>
<td>21%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>inv16</td>
<td>13%</td>
<td>25%</td>
<td>2%</td>
<td>5%</td>
<td>12%</td>
<td>0%</td>
<td>25%</td>
<td>3%</td>
<td>30%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td>Characteristic</td>
<td>Variable</td>
<td>Total</td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
<td>S4</td>
<td>S5</td>
<td>S6</td>
<td>S7</td>
<td>S8</td>
<td>S9</td>
<td>p</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------------------------</td>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>Cytogenetics (n=492)</td>
<td>Normal karyotype</td>
<td>28%</td>
<td>27%</td>
<td>23%</td>
<td>59%</td>
<td>21%</td>
<td>42%</td>
<td>38%</td>
<td>5%</td>
<td>16%</td>
<td>16%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t(9;11)(p22;q23)</td>
<td>18%</td>
<td>4%</td>
<td>23%</td>
<td>3%</td>
<td>10%</td>
<td>8%</td>
<td>6%</td>
<td>85%</td>
<td>28%</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-5, -7, or +8</td>
<td>9%</td>
<td>4%</td>
<td>15%</td>
<td>3%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
<td>3%</td>
<td>9%</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>15%</td>
<td>21%</td>
<td>17%</td>
<td>18%</td>
<td>13%</td>
<td>22%</td>
<td>12%</td>
<td>3%</td>
<td>7%</td>
<td>23%</td>
<td></td>
</tr>
<tr>
<td>FLT3-ITD (n=489)</td>
<td>Mutated</td>
<td>21%</td>
<td>10%</td>
<td>13%</td>
<td>10%</td>
<td>11%</td>
<td>4%</td>
<td>16%</td>
<td>0%</td>
<td>11%</td>
<td>7%</td>
<td>0.192</td>
</tr>
<tr>
<td>NPM1 (n=483)</td>
<td>Mutated</td>
<td>10%</td>
<td>13%</td>
<td>10%</td>
<td>11%</td>
<td>4%</td>
<td>16%</td>
<td>0%</td>
<td>11%</td>
<td>7%</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CEBPA (n=483)</td>
<td>Mutated</td>
<td>9%</td>
<td>4%</td>
<td>4%</td>
<td>33%</td>
<td>6%</td>
<td>21%</td>
<td>4%</td>
<td>0%</td>
<td>4%</td>
<td>2%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>c-KIT (exon 8) (n=399)</td>
<td>Mutated</td>
<td>4%</td>
<td>7%</td>
<td>4%</td>
<td>0%</td>
<td>5%</td>
<td>2%</td>
<td>13%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>0.125</td>
</tr>
<tr>
<td>c-KIT (exon 17) (n=399)</td>
<td>Mutated</td>
<td>8%</td>
<td>13%</td>
<td>9%</td>
<td>4%</td>
<td>20%</td>
<td>6%</td>
<td>10%</td>
<td>3%</td>
<td>3%</td>
<td>6%</td>
<td>0.068</td>
</tr>
<tr>
<td>c-KIT (combined) (n=399)</td>
<td>Mutated</td>
<td>12%</td>
<td>20%</td>
<td>11%</td>
<td>4%</td>
<td>25%</td>
<td>8%</td>
<td>23%</td>
<td>3%</td>
<td>6%</td>
<td>9%</td>
<td>0.004</td>
</tr>
<tr>
<td>KRAS (n=390)</td>
<td>Mutated</td>
<td>7%</td>
<td>2%</td>
<td>7%</td>
<td>0%</td>
<td>3%</td>
<td>8%</td>
<td>8%</td>
<td>13%</td>
<td>13%</td>
<td>9%</td>
<td>0.291</td>
</tr>
<tr>
<td>NRAS (n=390)</td>
<td>Mutated</td>
<td>25%</td>
<td>42%</td>
<td>23%</td>
<td>12%</td>
<td>16%</td>
<td>16%</td>
<td>23%</td>
<td>32%</td>
<td>33%</td>
<td>27%</td>
<td>0.024</td>
</tr>
<tr>
<td>KRAS and/or NRAS (n=390)</td>
<td>Mutated</td>
<td>31%</td>
<td>44%</td>
<td>27%</td>
<td>12%</td>
<td>18%</td>
<td>22%</td>
<td>28%</td>
<td>42%</td>
<td>41%</td>
<td>30%</td>
<td>0.012</td>
</tr>
<tr>
<td>PTPN11 (n=390)</td>
<td>Mutated</td>
<td>7%</td>
<td>2%</td>
<td>16%</td>
<td>0%</td>
<td>5%</td>
<td>5%</td>
<td>10%</td>
<td>10%</td>
<td>11%</td>
<td>0%</td>
<td>0.063</td>
</tr>
<tr>
<td>MYH11 (n=390)</td>
<td>Mutated</td>
<td>4%</td>
<td>16%</td>
<td>5%</td>
<td>4%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>0.037</td>
</tr>
<tr>
<td>GATA2 (n=390)</td>
<td>Mutated</td>
<td>3%</td>
<td>0%</td>
<td>2%</td>
<td>19%</td>
<td>3%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>3%</td>
<td>0.001</td>
</tr>
<tr>
<td>IDH1/2 (combined) (n=390)</td>
<td>Mutated</td>
<td>4%</td>
<td>2%</td>
<td>0%</td>
<td>8%</td>
<td>5%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
<td>6%</td>
<td>6%</td>
<td>0.455</td>
</tr>
<tr>
<td>NUP98-KDM5A (n=352)</td>
<td>Positive</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>11%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0.007</td>
</tr>
</tbody>
</table>
180 | Chapter 7

\(^1\)AAML1031 protocol risk group definition:

- Low risk: inv(16)/t(16;16) or t(8;21), or NPM or CEBPa mutation;
- High risk: FLT3/ITD+ with high allelic ratio \(\geq 0.4\), or monosomy 5/del5q or 7, without low-risk features;

*NA* unknown values not considered in \(p\)-value calculations and are excluded from the results;
Table 2. Multivariable analysis for OS, EFS and RR, including proteomic-based SIGs

A. Multivariate analysis, all patients (n = 400) 

<table>
<thead>
<tr>
<th></th>
<th>OS from study entry</th>
<th>EFS from study entry</th>
<th>RR from end of induction II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Proteomic-SIGs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>226</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>FAVORABLE</td>
<td>27</td>
<td>0.3</td>
<td>0.08 - 1.27</td>
</tr>
<tr>
<td>UNFAVORABLE</td>
<td>147</td>
<td>1.6</td>
<td>1.07 - 2.25</td>
</tr>
<tr>
<td>AAML1031 risk group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>281</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HIGH</td>
<td>119</td>
<td>2.6</td>
<td>1.75 - 3.73</td>
</tr>
<tr>
<td>AGE (yr.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥2</td>
<td>354</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>46</td>
<td>2.3</td>
<td>1.43 - 3.77</td>
</tr>
<tr>
<td>NPMI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEG</td>
<td>355</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>45</td>
<td>0.4</td>
<td>0.15 - 0.88</td>
</tr>
</tbody>
</table>

† 10 patients have unknown classification for protocol risk group and are excluded from analyses because all patients must have complete data for a multivariable analysis
Reference list


