Published in final edited form as:

Br J Haematol. 2013 July ; 162(2): 250–262. doi:10.1111/bjh.12370.

Development and Validation of a Single-Cell Network Profiling (SCNP) Assay-based Classifier to Predict Response to Induction Therapy in Paediatric Patients with *de novo* Acute Myeloid Leukaemia (AML): A Report from the Children's Oncology Group

Norman J. Lacayo^{1,2}, Todd A. Alonzo^{1,3}, Urte Gayko⁴, David B. Rosen⁴, Matt Westfall⁴, Norman Purvis⁴, Santosh Putta⁴, Brent Louie⁴, James Hackett⁴, Aileen Cleary Cohen⁴, Alessandra Cesano⁴, Robert Gerbing⁴, Yaddanapudi Ravindranath^{1,5}, Gary V. Dahl^{1,2}, Alan Gamis^{1,6}, and Soheil Meshinchi^{1,7}

¹Children's Oncology Group, Arcadia, CA

²Stanford University School of Medicine and Stanford Cancer Institute, Palo Alto, CA

³University of Southern California, Los Angeles, CA

⁴Nodality, Inc., South San Francisco, CA

⁵Wayne State University, Detroit, MI

⁶Children's Mercy Hospitals & Clinics, Kansas City, MO

⁷Fred Hutchinson Comprehensive Cancer Research Center, Seattle, WA

Summary

Single cell network profiling (SCNP) is a multi-parameter flow cytometry technique for simultaneous interrogation of intracellular signalling pathways. Diagnostic paediatric acute myeloid leukaemia (AML) bone marrow samples were used to develop a classifier for response to induction therapy in 53 samples and validated in an independent set of 68 samples. The area under the curve of a receiver operating characteristic curve (AUC_{ROC}) was calculated to be 0.85 in the

Corresponding Author: **Name:** Norman J. Lacayo, MD, **Address:** Division of Pediatric Hematology-Oncology-Stem Cell Transplantation and Cancer Biology, Stanford University School of Medicine and Stanford Cancer Institute, 1000 Welch Road, Suite 300, Palo Alto, CA 94304-1812, lacayon@stanford.edu, **Phone:** 650-723-5535, **Fax:** 650-723-5231.

Authorships and Disclosures:

N.J.L. designed the study, analysed the data and wrote the paper. He has no conflict of interest. T.A.A. analysed the data, performed statistical analysis, and wrote the paper. He has no conflict of interest. U.G. designed the study, analysed the data and wrote the paper. At the time of this work UG was an employee and stockholder of Nodality. D.B.R. designed the study, performed the study, analysed the data and wrote the paper. He is an employee and stockholder of Nodality. M.W. performed the study, analysed the data and wrote the paper. He is an employee and stockholder of Nodality. N.P. performed the study, analysed the data and wrote the paper. He is an employee and stockholder of Nodality. S.P. designed the study, analysed the data, performed statistical analysis and wrote the paper. He is an employee and stockholder of Nodality. B.L. analysed the data and wrote the paper. He is an employee and stockholder of Nodality. J.H. designed the study, analysed the data and wrote the paper. He is an employee and stockholder of Nodality. A.C.C. designed the study, analysed the data and wrote the paper. She is an employee and stockholder of Nodality. A.C. designed the study, analysed the data and wrote the paper. He is an employee and stockholder of Nodality. R.G. designed the study, analysed the data and wrote the paper. He has no conflict of interest. Y.R. designed the study, analysed the data and wrote the paper. He has no conflict of interest. G.V.D. designed the study, analysed the data and wrote the paper. He has no conflict of interest. A.G. designed the study, analysed the data and wrote the paper. He has no conflict of interest. S.M. designed the study, analysed the data and wrote the paper. He has no conflict of interest.

training set and after exclusion of induction deaths, the AUC_{ROC} of the classifier was 0.70 ($p=0.02$) and 0.67 ($p=0.04$) in the validation set when induction deaths (intent to treat) were included. The highest predictive accuracy was noted in the cytogenetic intermediate risk patients (AUC_{ROC} 0.88, $p=0.002$), a subgroup that lacks prognostic/predictive biomarkers for induction response. Only white blood cell count and cytogenetic risk were associated with response to induction therapy in the validation set. After controlling for these variables, the SCNP classifier score was associated with complete remission ($p=0.017$), indicating that the classifier provides information independent of other clinical variables that were jointly associated with response. This is the first validation of an SCNP classifier to predict response to induction chemotherapy. Herein we demonstrate the usefulness of quantitative SCNP under modulated conditions to provide independent information on AML disease biology and induction response.

Keywords

paediatric AML; SCNP; Induction response; intracellular signaling

Introduction

In diagnostic acute myeloid leukaemia (AML) samples, the bulk leukaemia blast cells and the smaller self-renewing AML cell population harbour different gene expression programs (Gentles, *et al* 2010, Gibbs, *et al* 2011, Jan, *et al* 2011, Majeti, *et al* 2009, Majeti and Weissman 2011). Preliminary data (Bendall, *et al* 2011, Gibbs, *et al* 2011, Irish, *et al* 2004) demonstrated that both gene expression and cell signalling profiles vary in bone marrow leukaemia cell populations. Moreover, abnormal cell signalling patterns identified by phosphoprotein profiles also vary among patients with AML and are probably modulated by several genetic and epigenetic characteristics in leukaemia (Kornblau, *et al* 2010a, Rosen, *et al* 2010a, Rosen, *et al* 2010b) and host cells. Although gene expression signatures may identify relevant signalling pathways, these cannot be probed in real-time to assess aberrant signalling in a single cell, cell subpopulations, or bulk leukaemia cells in blood or bone marrow samples. Similarly, detectable genetic mutations vary among cell subpopulations and therefore its association with outcome may be equivocal regardless of sample set size. It is unknown how remodelling of multiple canonical or non-canonical cellular signalling networks in leukaemia drive proliferation, cause de-differentiation, suppress apoptosis, or modulate senescence programmes at disease presentation or while on therapy. Moreover, how these signalling patterns relate to relevant genetic and epigenetic changes remains unidentified. Therefore, to initiate evaluation of these hypotheses regarding the intra and inter-patient heterogeneity of paediatric AML in the clinical setting, we applied a novel technology called single cell network profiling (SCNP) to unsorted diagnostic bone marrow samples.

SCNP uses a multi-parameter flow cytometry platform as a distinct proteomic assay for the analysis and interpretation of protein activity under baseline and modulated conditions. Pathway responses revealed by SCNP assays include activation failure, hypersensitivity/hyposensitivity of a pathway to modulators, altered response kinetics and rewiring of canonical pathways (Irish, *et al* 2006a, Kornblau, *et al* 2010a, Kotecha, *et al* 2008, Rosen, *et*

al 2010a, Rosen, *et al* 2010b). While this method of mapping signalling networks has potential applications in several facets of clinical medicine(Hotson, *et al* 2009, Irish, *et al* 2004, Irish, *et al* 2006b, Irish, *et al* 2010, Krutzik, *et al* 2004, Krutzik and Nolan 2006, Perez and Nolan 2006, Sachs, *et al* 2005) herein it was used to develop and validate a novel classifier associated with response to anthracycline/cytarabine-based induction chemotherapy in banked diagnostic samples.

Anthracycline/cytarabine-based combination chemotherapy regimens result in complete remission (CR) in over 85% of paediatric AML patients(Pui, *et al* 2011). Achieving a CR after two cycles of induction is associated with cytogenetic risk-group, FMS-like tyrosine kinase 3 receptor internal tandem duplication (*FLT3*-ITD) status, age, and white blood cell (WBC) count at diagnosis(Berman, *et al* 2011, Brown, *et al* 2004, Ho, *et al* 2009, Ho, *et al* 2010, Ho, *et al* 2011, Lacayo, *et al* 2004, Meshinchi, *et al* 2001, Pollard, *et al* 2010, Razzouk, *et al* 2006, Stirewalt, *et al* 2001, Zwaan, *et al* 2003). After initial Induction therapy, minimal residual disease (MRD) and cytogenetic and molecular assessments - available within one to four weeks of starting therapy - are used to inform second induction cycle or consolidation therapy intensification for patients with high-risk features. However, we still lack at diagnosis a validated predictor of induction response prior to initiation of therapy. Although MRD in paediatric AML is a robust predictor of treatment response and relative risk of relapse, it is usually based on a measurement obtained between days 22–28 of induction therapy(Loken, *et al* 2012, Rubnitz, *et al* 2010). The availability of a classifier soon after diagnosis may allow early institution of alternative therapies prior to or within the first course of induction therapy.

In the first phase of this study, SCNPs data from 53 banked and available diagnostic paediatric AML bone marrow (BM) samples were used to develop classifiers associated with response to standard induction chemotherapy. In the second phase a pre-specified classifier, chosen based on its accuracy in the training set, was validated in an independent set of 68 diagnostic BM samples available from a different Children's Oncology Group (COG) clinical trial.

In summary, by examining banked diagnostic samples we identified SCNPs profiles that were associated with response to induction therapy; subsequent validation of the SCNPs response characteristics in a distinct set of diagnostic samples confirmed a classifier predictive of response to induction therapy in non-M3, non-Down Syndrome (non-DS) paediatric AML patients. A prospective validation of this approach using fresh diagnostic AML samples is warranted.

Methods

Patient Samples and Study Inclusion Criteria

In accordance with the Declaration of Helsinki, all patients provided informed consent for use of their samples in research; each study was approved by the Institutional Review Board of all participating clinical centres; and clinical data was de-identified in compliance with US Health Insurance Portability and Accountability Act (HIPPA) regulations.

Inclusion criteria were diagnosis of paediatric (age 0–21 years) *de novo*, non-M3, non-DS AML, and availability of diagnostic cryopreserved BM sample with clinical information on response to induction chemotherapy. Criteria for evaluable samples included: 1) adequate cell health defined as greater than 25% cleaved-PARP (cPARP) negative (non-apoptotic, live leukaemic cells) after thaw, 2) sufficient cell numbers (>500 viable cells in the leukaemic cell population per condition/well at the end of the assay), and 3) absence of a technical assay deviation. Evaluable and non-evaluable patient clinical characteristics are shown in Table SI. For the training and validation study, each sample had at least 10×10^6 cells and 5×10^6 cells respectively at the time of original cryopreservation.

The training sample set consisted of 77 (53 evaluable) samples from 560 patients enrolled on Pediatric Oncology Group (POG)-9421 study, collected between 1995–1999 –a period when MRD testing was not available. The validation set consisted of 95 (68 evaluable) samples from 340 and 968 patients enrolled on COG trials AAML03P1 and AAML0531, collected between 2004–2005 and 2006–2010 respectively. Both sets were enriched with samples from patients who did not achieve remission in order to provide sufficient power to assess the primary study objective. There was limited availability of contemporary adult prognostic factors such as nucleophosmin (*NPM1*), CCAAT/ enhancer-binding protein alpha (*CEBPA*) mutation, Wilms Tumor (*WT1*) mutations, or DNA methyl-transferase (*DNMTA*) mutations in samples banked over the last decade, and additional samples were not tested because these factors are significantly less frequent in paediatric AML and thus rare in the cohorts tested (Ho, *et al* 2011) as well as lack of cryopreserved DNA and RNA for some samples. There was limited availability of MRD testing in the banked samples from 2006–2010.

Protocol Therapy

Induction chemotherapy used in the POG-9421 study (training set) consisted of one cycle of ADT (cytarabine randomized to 100 mg/m²/day or 1 g/m²/day $\times$ 7 days; daunorubicin 45 mg/m²/day $\times$ 3 days; and thioguanine 100 mg/m²/day $\times$ 7 days), followed by a second cycle of high dose cytarabine (1 g/m² every 12 h $\times$ 5 days). The COG studies AAML03P1 and AAML0531 (validationCrie set) induction regimens consisted of two cycles of ADE (+/-gemtuzumab ozogamicin [GMTZ] cycle 1 only) (cytarabine 100 mg/m²/q12 h/day $\times$ 10 days (cycle 1) and 8 days (cycle 2); daunorubicin 50 mg/m²/day on days 1, 3 and 5 for both cycles and etoposide 100 mg/m²/day on days 1–5 for both cycles; +/-GMTZ 3 mg/m² on day 6 of cycle 1 only).

Clinical response criteria specified in the clinical protocols were identical across all 3 studies. CR was defined as <5% BM blasts, no extramedullary disease and recovery of peripheral counts in a setting of a cellular marrow (Cheson, *et al* 2003). For this study, the best clinical response at the end of induction 1 and 2 was used for all analyses. Patients whose disease did not meet the definition of CR at the end of induction therapy were assigned an outcome of non-responder (NR), these were patients with resistant disease; no incomplete CR (CRi) cases were included in this analysis. Patients who died in induction therapy (early deaths, ED) or prior to assignment of response were classified in the COG database as an induction death and were considered as NR for the primary analysis. A

sensitivity analysis removing the induction-death patients was pre-specified in the validation phase of the study. This pre-specification was performed because the goal of the study involved the prediction of leukaemic cell killing (remission induction) which is probably different from the biology underlying patient death during induction related to infection or other complications of high dose therapy. Furthermore, induction failures are both less frequent in paediatric AML and associated with a higher risk of death.

Study Design—This study consisted of training and validation phases designed as prospective investigations but using retrospective samples (Figure 1). The training phase resulted in the development of a multivariate SCNP assay-based classifier that was then tested for accuracy in an independent, blinded set of validation samples. Limited availability of banked samples precluded identical proportion of known AML subtypes in the test and validation set. Moreover, the heterogeneity of AML with respect to the taxonomy of signalling is expected to have phenotypes at variance from defined cytogenetic and molecular subtypes (Rosen, *et al* 2010a).

Sample size for the training study was driven by availability of samples in the COG tissue bank that met patient and sample requirements. The validation set was designed to provide at least 90% power to test the primary hypothesis, i.e., area under the curve of a receiver operating characteristic curve (AUC_{ROC}) for the SCNP classifier >0.5 at a two-sided significance level of 0.05, assuming the true AUC_{ROC} of the pre-specified induction response classifier was 0.7. To allow for appropriate training of a response classifier, samples were enriched in both studies for NR (~30%) to induction therapy as compared to 10–13% observed in the original treatment studies (Becton, *et al* 2006, Cooper, *et al* 2012).

SCNP Assay Terminology and Signalling Nodes Evaluated

The term “signalling node” is used to refer to proteomic readout in the presence or absence of a specific modulator. Modulators used were either endogenous growth factors (e.g., FLT3 ligand) or drugs (cytarabine, daunorubicin, etoposide) for the induction of apoptosis. Several metrics (normalized assay readouts defined in Figure SI) were applied to interpret the functionality and biology of each signalling node.

A total of 82 nodes were evaluated in the Training phase of the study (3 surface markers, 19 unmodulated and 60 modulated readouts) (Table SII, Figure 2A). Approximately 4×10^6 cells were required to test all 82 nodes (total of 15 modulators tested). The classifier evaluated in the Validation phase of the study included 3 signalling nodes (Figure 2B) (1.1×10^6 cells required for 3 modulator conditions).

SCNP Assay

SCNP assays were performed as described previously (Kornblau, *et al* 2010a, Kornblau, *et al* 2010b, Kornblau, *et al* 2009, Rosen, *et al* 2010a). In short, cryopreserved samples were thawed, processed over ficoll gradient to remove debris, washed and stained with Aqua Viability Dye. After rest at 37°C, cells were aliquoted on 96-well plates using a robotic system (Hamilton, Reno, NV) and incubated with modulators for pre-specified time (Table SIIIA and B). After exposure to modulators, cells were fixed, permeabilized and stained

against phenotypic markers for identification of myeloid blasts with three antibodies against intracellular signalling molecules or other surface markers for an 8-colour flow cytometry assay.

Flow Cytometry Data Acquisition and Analysis

Flow cytometry data was acquired on several CANTO II flow cytometers using the FACS DIVA software (BD Biosciences, San Jose, CA). All flow cytometry data were analysed with FlowJo (TreeStar Software, Ashland, OR) and WinList (Verity House Software, Topsham, ME). All analyses presented were based on leukaemic cells (43.6–98.2% [median 90%] for the training set and 15.7–97.3% [median 74.3%] for the validation set, of a given cell preparation) fitting characteristics consistent with myeloid leukaemia cells and lacking characteristics of mature lymphocytes (low side scatter, bright CD45⁺, and CD34 negative) (Cesano, *et al* 2012, Stelzer and Goodpasture 2000).

Sample Processing, Cell Recovery and Evaluation—For the Training study BM samples were previously archived in aliquots of a median of 40×10^6 cells. Seventy-seven patient samples were selected and processed for the SCNPN assay between March–June, 2010. Median cell number after thaw and ficoll separation was 5.4×10^6 . Only 53 of 77 samples from 38 US centres were evaluable, 24 samples were considered non-evaluable based on pre-specified criteria: 13 due to poor viability/insufficient cells, 11 due to technical assay error and potential bacterial contamination.

For the Validation study BM Samples had been archived at aliquots of a median of 15×10^6 cells. Ninety-five patient samples were selected and processed for the SCNPN assay between March and May 2011. Median cell numbers after thaw and ficoll separation were 4.2×10^6 . Only 68 of 95 samples from 53 US centres were evaluable; 27 samples were considered non-evaluable based on pre-specified criteria. Specifically, 19 samples had insufficient cells to generate an SCNPN score, 4 failed the cell health criteria and 4 samples had an assay deviation leading to exclusion.

Controls and Reproducibility

Samples were divided in aliquots of 100,000 cells and run on a 96-well plate. On each plate, one row of control cell lines and one column of rainbow beads were run for assay quality control (QC) and normalization of mean fluorescence intensity (MFI) values, respectively. Each cytometer had to pass daily QC in order to allow processing of cell preparations for data acquisition. To assess assay reproducibility, the assay was repeated for a subset of samples in both Training and Validation phases.

Statistical Analysis

Tests of differences in patient subsets were performed using Pearson's chi-square test or Fisher's exact test when assumptions of the chi-square test were not met. The consistency of assay results was measured by a multiple correlation coefficient (R^2). The AUC_{ROC} was used as a measure of predictive accuracy. During the model development (training) phase of the study, AUC_{ROC} estimates for models (combinations of nodes) constructed in the training set were adjusted, using bootstrap methods, to better reflect the likely performance of the

model in an independent data set (Efron & Robert 1993). These adjusted AUC_{ROC} estimates of model performance obtained from the training data are referred to here as optimism-adjusted.

Validation Phase—The primary hypothesis required the AUC_{ROC} for the SCNP classifier to be significantly > 0.5 , where higher classifier scores indicate a greater probability of CR to anthracycline/cytarabine-based induction chemotherapy. The null hypothesis was tested using an exact permutation method based on the Wilcoxon rank sum statistic (Agresti 1992) at a two-sided significance level of 0.05. The Wilcoxon rank sum test is equivalent to testing the AUC_{ROC} against a null value of 0.5 (Pepe 2003).

Confidence intervals for AUC_{ROC} estimates of SCNP classifier performance in the Validation set were calculated using the bias corrected accelerated (BCa) estimator, which is similar to the bootstrap percentile confidence interval but is corrected for bias and the rate of change of the standard error of the estimate with respect to the true parameter value (Efron & Robert 1993). Multivariate logistic regression was used to evaluate which clinical variables were jointly associated with response, and to evaluate the association between the SCNP score and response, controlling for jointly significant clinical variables. Significance tests of independent variables in logistic regression models of response were based on the likelihood ratio chi-square test.

Results

Patient Characteristics

Clinical characteristics associated with the evaluable training and validation samples are summarized in Table I. The Training sample set had a high percentage of AML patients carrying *FLT3*-ITD mutations (36%, 20.8% with a *FLT3*-ITD allelic ratio > 0.4) and high ($> 100 \times 10^9$ cells/l) WBC count at diagnosis (34%), both negative prognostic factors. While most (63%) samples belonged to the intermediate risk cytogenetic group, as defined by National Comprehensive Cancer Network 2011 criteria (O'Donnell 2011), the favourable risk cytogenetic group was underrepresented (9.4%). In contrast, the Validation sample set had fewer samples with the *FLT3*-ITD mutation (20%, 14.7% with an allelic ratio > 0.4), high WBC count at diagnosis (26.5%), and intermediate risk cytogenetic characteristics (40%); but had more samples with favourable risk cytogenetic characteristics (29%). Both Training and Validation sets had similar representation of *NPM1* mutations (11.3% versus 10.3%). There was no statistically significant difference between the cohorts of patients with evaluable vs. non-evaluable samples with the exception of a different gender distribution (44% evaluable vs. 18% non-evaluable females in the Validation set) (Table SI). The only statistically significant differences were in the relative percent of cytogenetic risk groups and *FLT3*-ITD allelic ratio. These differences reflect sample bank limitations.

Pathways Evaluated

Based upon relevance to AML pathophysiology, cell surface receptors and signalling pathways involved in cell survival, proliferation, apoptosis and DNA damage response

(DDR) were investigated (Figure 2A). Apoptotic signalling and DDR were measured after *in vitro* exposure of AML samples to etoposide or anthracycline/cytarabine.

Classifier Development

Training Phase—Initially, 10 of 82 total nodes were selected based on their ability to classify donors according to their CR/NR status. The methodologies used for selecting the initial node subset were: 1) node-signalling differences by CR/NR, 2) Random Forest (Breiman 2001) node importance for CR/NR using all nodes, and 3) non-zero node coefficients by Penalized Logistic Regression (Firth 1993, Harrell 2001, Hollander & Wolfe 1999, Liaw, & Wiener 2002, Pepe 2003) for CR/NR using all nodes.

To find combinations of nodes that were better predictors of CR/NR than individual nodes, models based on 2- to 4-node combinations of the 10 nodes were developed (>300) using penalized logistic regression. The adjusted AUC_{ROC} for each of the models was then calculated (Mason & Graham 2002). Bootstrap re-sampling (n=500) was used to adjust the AUC_{ROC} for possible optimism. The models were then ranked by their adjusted AUC_{ROC} and several of the highest-ranking models were investigated further. The lead candidate was selected based on several criteria including signalling pathway, the range of node signalling, and experience from previous studies (Kornblau, *et al* 2010a).

All evaluable Training samples (n= 53: 36 CR, 17 NRs) were included for model building using classification algorithms such as penalized logistic regression and random forest. Candidate models were selected based on their ability to correctly classify CR/NR patients. The lead candidate was a penalized logistic regression model with 3 SCNP node-metrics representing apoptosis, PI3K signalling and proliferation pathways (Figure 2A,B). The optimism adjusted estimate of the AUC_{ROC} for this model was 0.85 with 95% CI = (0.66, 0.98). This SCNP classifier, with all parameters fixed (Figure 2B), was then applied to the validation sample to estimate its true accuracy.

Accuracy of the SCNP Classifier

The accuracy of the SCNP classifier in predicting response to induction therapy met the criteria for success in the validation study (Table II). The AUC_{ROC} of the classifier in the validation set was significantly greater than 0.5 (AUC_{ROC}=0.66, $p=0.042$, 95% CI = (0.52, 0.78)). The primary analysis was conducted with the intent to diagnose a population in which both induction failure (n=14) or induction death (n=6) were included in the NR group. As the biology underlying the two outcomes might be different and because only 2 samples from patients who died during the induction treatment were present in the training set, a pre-specified secondary analysis was also conducted after removal of samples from patients who died during the induction phase. The AUC_{ROC} for the SCNP classifier in this patient set (n=62) was 0.70 ($p=0.021$) (Table II).

Reproducibility—Reproducibility of the assay was assessed in both the training and validation studies by evaluating the linear relationship between SCNP classifier scores based on independent runs of the assay for a subset of samples. In the training set, the assay was repeated for 14 samples at different time points, specifically one year apart and in different

laboratories. The R^2 for the two sets of SCNP classifier scores was 0.96. Figure 3A shows the high concordance of SCNP scores for the two runs in the training set. In the validation set, the nodes required for the classifier were repeated twice for 43 subjects on 2 separate 96-well plates using different reagent batches. The R^2 was 0.942 for the SCNP classifier score (Figure 3B) and between 0.88–0.94 for each individual node-metric (Figures 3C–E).

Impact of Clinical Variables—Variables previously reported to be correlated to induction response in paediatric AML were categorized in (a) variables available before induction cycle 1 (i.e., age, gender and WBC) and (b) available after start of induction cycle 1 (cytogenetics and *FLT3*-ITD). The use of MRD in the analysis of banked samples was difficult because an evaluable baseline and post-induction 1 therapy sample are required for the analysis. Other genetic markers of interest in the adult AML setting, such as *NPM1* or *CEBPA* are of less value because a correlation to response in paediatric AML is difficult to demonstrate due to low prevalence of these molecular alterations in children. In this data set MRD, *NPM1*, and *CEBPA* were not associated with response to induction therapy ($p=0.51$, 0.99 and 0.52 respectively, data not shown).

In a multivariate logistic regression model of response in the validation set, WBC at diagnosis and cytogenetic risk groups were associated with induction response; but age, gender and *FLT3*-ITD status were not (Table III). After controlling for these significant clinical variables (WBC and cytogenetic risk), the SCNP classifier score remained significantly associated with CR ($p=0.017$), indicating that the SCNP classifier provides information independent of the other clinical variables that were jointly associated with response (Table III). Exploratory analyses excluding induction deaths suggested that the relationship between the SCNP score and induction response was strongest among patients with an intermediate risk cytogenetic classification ($n=23$) ($AUC_{ROC}=0.88$, $p=0.002$). Conversely, no relationship ($AUC_{ROC}=0.48$, $p=0.959$, $n=17$) was seen in those with an unfavourable risk cytogenetic classification (Figure 4B). *The p-value when comparing the AUC Curve in the ‘cytogenetics intermediate risk’ to that for the ‘cytogenetics poor risk’ group is $P=0.017$ when using a DeLong test.* Among the three SCNP signalling nodes contributing to the score, the node that measured drug-induced apoptosis and DNA damage response performed most consistently across the training and validation sets.

Discussion

Whereas current standard induction regimen CR rates approach 85% and 3-year event-free survival (EFS) 60% (Cooper, *et al* 2012, Pui, *et al* 2011, Rubnitz, *et al* 2010, Tsukimoto, *et al* 2009), no validated biomarkers to identify refractory disease at the time of diagnosis are available, thus preventing early adoption of alternative induction strategies. Using current state of the art therapy, induction failures in paediatric AML remain a significant proportion of overall treatment failures and warrant diagnostic biomarkers for the development of new therapies.

In this study, cryopreserved unsorted diagnostic BM samples, collected prospectively as part of three COG studies, were used for the development and validation of a SCNP assay-based classifier to predict response to anthracycline/cytarabine-based induction chemotherapy in

paediatric AML. SCNP is a multiparametric flow cytometry technique that allows functional pathway analysis at the level of the single cell without the need for cell sorting. Because it measures the ultimate functional consequences of multiple molecular (genetic and epigenetic) alterations at the downstream cellular level, SCNP assay readouts are likely to have higher predictive value than the evaluation of known molecular alterations from cytogenetic abnormalities or specific gene mutations. This study was limited to bulk leukaemia samples and did not include sufficient cell surface markers for detailed cell subset analysis amenable to statistical evaluation. However, this was a relative limitation given that the goal of the study was the prediction of induction CR, i.e., reduction of bulk leukaemia cells below 5%. This limitation would have a much greater and clinically relevant effect if trying to predict disease relapse where the responsible cells are in the 5% of residual blasts present after response or evolve from residual blasts by acquiring additional mutations.

To develop an SCNP assay-based test we used a standard approach for classifier development, with prospectively designed training and validation phases. After a broad evaluation of several signalling pathways selected for their relevance to the leukaemogenesis process (training study), a SCNP classifier with quantitative readouts for DNA damage response and apoptosis, PI3K–signalling, and proliferation pathways was developed. Consistent with results of SCNP studies in adult AML samples, *in vitro* readouts of induced DNA damage response and apoptosis, specifically etoposide-induced cPARP, was the key contributor to the model for prediction of complete response in paediatric samples. Although evaluated in the training set, neither surface markers themselves nor basal (resting) signalling readouts were top candidates for the response to induction classifier. In future studies we will evaluate the use of our training set surface markers and basal signalling readouts for a risk of relapse classifier.

In spite of differences in induction therapy regimens and schedules (samples came from 3 different clinical trials using two anthracycline/cytarabine-based induction therapy backbones), and varying proportions of AML subtypes in the training and validation sets from available banked samples, a robust and highly reproducible SCNP classifier was validated and showed an AUC_{ROC} of the classifier in the validation set of 0.70 ($p=0.02$) (AUC_{ROC} 0.66, $p=0.042$ for the intent to diagnose population). In view of the fact that the test and validation sets were studied one year apart, we could not merge all samples and split them into a new test and validation set with more homogeneous known biomarkers. Indeed, our intent was to assess the heterogeneity of paediatric AML based on signalling characteristics of the bulk population that integrates multiple genetic and epigenetic features into an overall population signalling phenotype.

The performance of this classifier was higher when early death cases were excluded (as per a pre-specified analysis) and NRs limited to resistant disease patients (AUC_{ROC} 0.70, $p=0.021$), supporting the hypothesis that the underlying biology is different between the latter two sets of patients with unfavourable outcome. Of note, 4 of the 6 samples associated with induction death in the validation set had been predicted as CR, and 2 predicted as NR by the SCNP classifier. Two out of 4 of these patients (predicted as CR) achieved a morphological CR after induction 1, but then died during induction 2 due to comorbidities. The latter patients were pre-specified to be removed in the study design.

In the validation set the only clinical variables significantly associated with induction response were WBC and cytogenetic risk group. Age, gender and *FLT3*-ITD status were not associated with induction response and, as noted before, other features used in adult AML (*NPM1* and *CEBPA*) are rare in children and probably not useful in predicting induction response. More importantly, after controlling for WBC and cytogenetic risk group, the SCNP classifier score remained significantly associated with CR ($p=0.017$), indicating that the SCNP classifier provides information independent of the other clinical variables that were jointly associated with response. Furthermore, within cytogenetic risk groups, the predictive accuracy of the SCNP model was the highest in the intermediate risk group with exclusion of early deaths (AUC_{ROC} 0.88, $p=0.002$). These patients represent 60% of newly diagnosed non-M3, non-DS paediatric AML and are a subgroup in which additional prognostic/predictive biomarkers to guide patient management are especially needed.

The remaining 40% of patients belong to both the favourable and unfavourable risk groups. Core binding factor AML has a favourable outcome with state of the art therapy, while the unfavourable risk group probably represents individuals with a more heterogeneous population of AML cells. The favourable risk group benefits from the robust predictive value of cytogenetics and can be identified early with real time quantitative polymerase chain reaction. In contrast, the unfavourable risk group will require separate in-depth analysis with a larger set of cell surface markers, as carried out in our training set, to better distinguish this complex population mix with multiple non-overlapping genetic and epigenetic mechanisms of drug resistance (Breems and Lowenberg 2011, Breems, *et al* 2008).

The SCNP assays described herein were performed on cryopreserved samples. Preliminary data in adult AML showed high correlation between SCNP readouts in paired fresh and cryopreserved aliquots of the same samples (Kornblau, *et al* 2010a). A separate study is being performed to confirm this observation for the paediatric AML setting. Once these data become available, the SCNP assay will be moved prospectively into the clinical setting to confirm its feasibility and value as provider of independent and actionable information for consideration prior to initiating therapy.

A parallel SCNP study, initially retrospective and latterly prospective, is planned to develop classifiers for prediction of relapse risk in bulk leukaemia cells from patients who achieve CR after induction therapy to guide individualized consolidation treatments. These studies await release of overall survival and EFS data (AAML0531) to model a classifier from the data sets used in this study. These analyses will be validated with a new cohort from the AAML03P1 and AAML0531 COG studies.

It will be important to correlate genetic and epigenetic studies (e.g., National Cancer Institute paediatric AML TARGET project) with SCNP analysis of single cells and leukaemic sub-populations to improve our understanding of the heterogeneity of childhood AML, especially in the unfavourable risk cytogenetic group. This knowledge will provide a roadmap that is urgently needed for efficient and targeted design of clinical trials and an individualized treatment approach for intermediate and unfavourable risk paediatric AML patients. Our challenge is to understand how the taxonomy of molecular lesions in paediatric

AML correlates with re-wiring of signalling networks in single cells and cell subsets at diagnosis, as well as in real-time while under therapeutic pressure.

If this approach is validated prospectively during therapy we envision the use of SCNP as a functional flow cytometry test to be integrated into each induction and consolidation course of paediatric AML therapy. These future studies may use newer technology with increased sensitivity for routine assessment of single cells (Bendall, *et al* 2011). Such an approach may allow assessment of response to induction, risk of relapse, emergence and evolution of drug resistance, and emergence of aberrant signalling leading to relapse. As more samples undergo SCNP studies a comprehensive taxonomy of signalling in *de novo* paediatric AML may provide insight into a rational addition of targeted therapies in the setting of a modified backbone of anthracycline/cytarabine induction schema. Moreover, these functional flow cytometry tests could facilitate rational targeted therapies for the limited number of relapse patients that stymie the field of developmental therapeutics in paediatric AML.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank all of the patients and their families. We also thank those who have contributed with research support and grant funding: Children's Oncology Group (COG), the Valvano Foundation: Novel Approaches in Myeloid Leukemia (N.J.L.; G.V.D.), Ariana Riccio Funds for Pediatric AML Research (N.J.L.; G.V.D.), and Nodality, Inc.

References

- Agresti A. A Survey of Exact Inference for Contingency Tables. *Statistical Science*. 1992; 7:46.
- Becton D, Dahl GV, Ravindranath Y, Chang MN, Behm FG, Raimondi SC, Head DR, Stine KC, Lacayo NJ, Sikic BI, Arceci RJ, Weinstein H. Randomized use of cyclosporin A (CsA) to modulate P-glycoprotein in children with AML in remission: Pediatric Oncology Group Study 9421. *Blood*. 2006; 107:1315–1324. [PubMed: 16254147]
- Bendall SC, Simonds EF, Qiu P, Amir el AD, Krutzik PO, Finck R, Bruggner RV, Melamed R, Trejo A, Ornatsky OI, Balderas RS, Plevritis SK, Sachs K, Pe'er D, Tanner SD, Nolan GP. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*. 2011; 332:687–696. [PubMed: 21551058]
- Berman JN, Gerbing RB, Alonzo TA, Ho PA, Miller K, Hurwitz C, Heerema NA, Hirsch B, Raimondi SC, Lange B, Franklin JL, Gamis A, Meshinchi S. Prevalence and clinical implications of NRAS mutations in childhood AML: a report from the Children's Oncology Group. *Leukemia*. 2011; 25:1039–1042. [PubMed: 21358716]
- Breems DA, Lowenberg B. Acute myeloid leukemia with monosomal karyotype at the far end of the unfavorable prognostic spectrum. *Haematologica*. 2011; 96:491–493. [PubMed: 21454879]
- Breems DA, Van Putten WL, De Greef GE, Van Zelderen-Bhola SL, Gerssen-Schoorl KB, Mellink CH, Nieuwint A, Jotterand M, Hagemeijer A, Beverloo HB, Lowenberg B. Monosomal karyotype in acute myeloid leukemia: a better indicator of poor prognosis than a complex karyotype. *J Clin Oncol*. 2008; 26:4791–4797. [PubMed: 18695255]
- Breiman L. Random Forests. *Machine Learning*. 2001; 45(1):5–32.
- Brown P, Meshinchi S, Levis M, Alonzo TA, Gerbing R, Lange B, Arceci R, Small D. Pediatric AML primary samples with FLT3/ITD mutations are preferentially killed by FLT3 inhibition. *Blood*. 2004; 104:1841–1849. [PubMed: 15166029]

- Cesano A, Rosen DB, O'Meara P, Putta S, Gayko U, Spellmeyer DC, Cripe LD, Sun Z, Uno H, Litzow MR, Tallman MS, Paietta E. Functional pathway analysis in acute myeloid leukemia using single cell network profiling assay: Effect of specimen source (bone marrow or peripheral blood) on assay readouts. *Cytometry B Clin Cytom.* 2012; 82:158–172. [PubMed: 22334473]
- Cheson BD, Bennett JM, Kopecky KJ, Buchner T, Willman CL, Estey EH, Schiffer CA, Doehner H, Tallman MS, Lister TA, Lo-Coco F, Willemze R, Biondi A, Hiddemann W, Larson RA, Lowenberg B, Sanz MA, Head DR, Ohno R, Bloomfield CD. Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J Clin Oncol.* 2003; 21:4642–4649. [PubMed: 14673054]
- Cooper TM, Franklin J, Gerbing RB, Alonzo TA, Hurwitz C, Raimondi SC, Hirsch B, Smith FO, Mathew P, Arceci RJ, Feusner J, Iannone R, Lavey RS, Meshinchi S, Gamis A. AAML03P1, a pilot study of the safety of gemtuzumab ozogamicin in combination with chemotherapy for newly diagnosed childhood acute myeloid leukemia: A report from the Children's Oncology Group. *Cancer.* 2012; 118:761–769. [PubMed: 21766293]
- Efron, BT.; Robert, J. An Introduction to Bootstrap. Boca Raton, FL: Chapman and Hall; 1993.
- Firth D. Bias Reduction of Maximum Likelihood Estimates. *Biometrika.* 1993; 80:41.
- Gentles AJ, Plevritis ScK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA.* 2010; 304:2706–2715. [PubMed: 21177505]
- Gibbs KD Jr, Gilbert PM, Sachs K, Zhao F, Blau HM, Weissman IL, Nolan GP, Majeti R. Single-cell phospho-specific flow cytometric analysis demonstrates biochemical and functional heterogeneity in human hematopoietic stem and progenitor compartments. *Blood.* 2011; 117:4226–4233. [PubMed: 21357764]
- Harrell, FE, Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer-Verlag; 2001.
- Ho PA, Alonzo TA, Gerbing RB, Pollard J, Stirewalt DL, Hurwitz C, Heerema NA, Hirsch B, Raimondi SC, Lange B, Franklin JL, Radich JP, Meshinchi S. Prevalence and prognostic implications of CEBPA mutations in pediatric acute myeloid leukemia (AML): a report from the Children's Oncology Group. *Blood.* 2009; 113:6558–6566. [PubMed: 19304957]
- Ho PA, Alonzo TA, Kopecky KJ, Miller KL, Kuhn J, Zeng R, Gerbing RB, Raimondi SC, Hirsch BA, Oehler V, Hurwitz CA, Franklin JL, Gamis AS, Petersdorf SH, Anderson JE, Reaman GH, Baker LH, Willman CL, Bernstein ID, Radich JP, Appelbaum FR, Stirewalt DL, Meshinchi S. Molecular alterations of the IDH1 gene in AML: a Children's Oncology Group and Southwest Oncology Group study. *Leukemia.* 2010; 24:909–913. [PubMed: 20376086]
- Ho PA, Kutny MA, Alonzo TA, Gerbing RB, Joaquin J, Raimondi SC, Gamis AS, Meshinchi S. Leukemic mutations in the methylation-associated genes DNMT3A and IDH2 are rare events in pediatric AML: a report from the Children's Oncology Group. *Pediatr Blood Cancer.* 2011; 57:204–209. [PubMed: 21504050]
- Hollander, M.; Wolfe, DA. Nonparametric Statistical Methods. John Wiley and Sons, Inc.; 1999. xiv +787 pages
- Hotson AN, Hardy JW, Hale MB, Contag CH, Nolan GP. The T cell STAT signaling network is reprogrammed within hours of bacteremia via secondary signals. *J Immunol.* 2009; 182:7558–7568. [PubMed: 19494279]
- Irish JM, Hovland R, Krutzik PO, Perez OD, Bruserud O, Gjertsen BT, Nolan GP. Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell.* 2004; 118:217–228. [PubMed: 15260991]
- Irish JM, Czerwinski DK, Nolan GP, Levy R. Altered B-cell receptor signaling kinetics distinguish human follicular lymphoma B cells from tumor-infiltrating nonmalignant B cells. *Blood.* 2006a; 108:3135–3142. [PubMed: 16835385]
- Irish JM, Kotecha N, Nolan GP. Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat Rev Cancer.* 2006b; 6:146–155. [PubMed: 16491074]
- Irish JM, Myklebust JH, Alizadeh AA, Houot R, Sharman JP, Czerwinski DK, Nolan GP, Levy R. B-cell signaling networks reveal a negative prognostic human lymphoma cell subset that emerges

- during tumor progression. *Proc Natl Acad Sci U S A*. 2010; 107:12747–12754. [PubMed: 20543139]
- Jan M, Chao MP, Cha AC, Alizadeh AA, Gentles AJ, Weissman IL, Majeti R. Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. *Proc Natl Acad Sci U S A*. 2011; 108:5009–5014. [PubMed: 21383193]
- Kornblau SM, Tibes R, Qiu YH, Chen W, Kantarjian HM, Andreeff M, Coombes KR, Mills GB. Functional proteomic profiling of AML predicts response and survival. *Blood*. 2009; 113:154–164. [PubMed: 18840713]
- Kornblau SM, Minden MD, Rosen DB, Putta S, Cohen A, Covey T, Spellmeyer DC, Fantl WJ, Gayko U, Cesano A. Dynamic single-cell network profiles in acute myelogenous leukemia are associated with patient response to standard induction therapy. *Clin Cancer Res*. 2010a; 16:3721–3733. [PubMed: 20525753]
- Kornblau SM, McCue D, Singh N, Chen W, Estrov Z, Coombes KR. Recurrent expression signatures of cytokines and chemokines are present and are independently prognostic in acute myelogenous leukemia and myelodysplasia. *Blood*. 2010b; 116:4251–4261. [PubMed: 20679526]
- Kotecha N, Flores NJ, Irish JM, Simonds EF, Sakai DS, Archambeault S, Diaz-Flores E, Coram M, Shannon KM, Nolan GP, Loh ML. Single-cell profiling identifies aberrant STAT5 activation in myeloid malignancies with specific clinical and biologic correlates. *Cancer Cell*. 2008; 14:335–343. [PubMed: 18835035]
- Krutzik PO, Nolan GP. Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. *Nat Methods*. 2006; 3:361–368. [PubMed: 16628206]
- Krutzik PO, Irish JM, Nolan GP, Perez OD. Analysis of protein phosphorylation and cellular signaling events by flow cytometry: techniques and clinical applications. *Clin Immunol*. 2004; 110:206–221. [PubMed: 15047199]
- Lacayo NJ, Meshinchi S, Kinnunen P, Yu R, Wang Y, Stuber CM, Douglas L, Wahab R, Becton DL, Weinstein H, Chang MN, Willman CL, Radich JP, Tibshirani R, Ravindranath Y, Sikic BI, Dahl GV. Gene expression profiles at diagnosis in de novo childhood AML patients identify FLT3 mutations with good clinical outcomes. *Blood*. 2004; 104:2646–2654. [PubMed: 15251987]
- Liaw A, Wiener M. Classifications and Regression by random Forest. *R News*. 2002; 2/3:18–22.
- Loken MR, Alonzo TA, Pardo L, Gerbing RB, Raimondi SC, Hirsch BA, Ho PA, Franklin J, Cooper TM, Gamis AS, Meshinchi S. Residual disease detected by multidimensional flow cytometry signifies high relapse risk in patients with de novo acute myeloid leukemia: a report from Children's Oncology Group. *Blood*. 2012; 120:1581–1588. [PubMed: 22649108]
- Majeti R, Weissman IL. Human acute myelogenous leukemia stem cells revisited: there's more than meets the eye. *Cancer Cell*. 2011; 19:9–10. [PubMed: 21251611]
- Majeti R, Becker MW, Tian Q, Lee TL, Yan X, Liu R, Chiang JH, Hood L, Clarke MF, Weissman IL. Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. *Proc Natl Acad Sci U S A*. 2009; 106:3396–3401. [PubMed: 19218430]
- Mason SJ, Graham NE. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q. J. R. Meteorological Society*. 2002; 128:21.
- Meshinchi S, Woods WG, Stirewalt DL, Sweetser DA, Buckley JD, Tjoa TK, Bernstein ID, Radich JP. Prevalence and prognostic significance of Flt3 internal tandem duplication in pediatric acute myeloid leukemia. *Blood*. 2001; 97:89–94. [PubMed: 11133746]
- O'Donnell MRACN, Altman J. NCCN clinical practice guidelines in oncology: Acute Myeloid Leukemia. *J Natl Compr Canc Netw*. 2011; 9:37.
- Pepe, MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press; 2003.
- Perez OD, Nolan GP. Phospho-proteomic immune analysis by flow cytometry: from mechanism to translational medicine at the single-cell level. *Immunol Rev*. 2006; 210:208–228. [PubMed: 16623773]
- Pollard JA, Alonzo TA, Gerbing RB, Ho PA, Zeng R, Ravindranath Y, Dahl G, Lacayo NJ, Becton D, Chang M, Weinstein HJ, Hirsch B, Raimondi SC, Heerema NA, Woods WG, Lange BJ, Hurwitz

C, Arceci RJ, Radich JP, Bernstein ID, Heinrich MC, Meshinchi S. Prevalence and prognostic significance of KIT mutations in pediatric patients with core binding factor AML enrolled on serial pediatric cooperative trials for de novo AML. *Blood*. 2010; 115:2372–2379. [PubMed: 20056794]

Pui CH, Carroll WL, Meshinchi S, Arceci RJ. Biology, risk stratification, and therapy of pediatric acute leukemias: an update. *J Clin Oncol*. 2011; 29:551–565. [PubMed: 21220611]

Razzouk BI, Estey E, Pounds S, Lensing S, Pierce S, Brandt M, Rubnitz JE, Ribeiro RC, Rytting M, Pui CH, Kantarjian H, Jeha S. Impact of age on outcome of pediatric acute myeloid leukemia: a report from 2 institutions. *Cancer*. 2006; 106:2495–2502. [PubMed: 16639734]

Rosen DB, Minden MD, Kornblau SM, Cohen A, Gayko U, Putta S, Woronicz J, Evensen E, Fantl WJ, Cesano A. Functional characterization of FLT3 receptor signaling deregulation in acute myeloid leukemia by single cell network profiling (SCNP). *PLoS One*. 2010a; 5:e13543. [PubMed: 21048955]

Rosen DB, Putta S, Covey T, Huang YW, Nolan GP, Cesano A, Minden MD, Fantl WJ. Distinct patterns of DNA damage response and apoptosis correlate with Jak/Stat and PI3kinase response profiles in human acute myelogenous leukemia. *PLoS One*. 2010b; 5:e12405. [PubMed: 20811632]

Rubnitz JE, Inaba H, Dahl G, Ribeiro RC, Bowman WP, Taub J, Pounds S, Razzouk BI, Lacayo NJ, Cao X, Meshinchi S, Degar B, Airewele G, Raimondi SC, Onciu M, Coustan-Smith E, Downing JR, Leung W, Pui CH, Campana D. Minimal residual disease-directed therapy for childhood acute myeloid leukaemia: results of the AML02 multicentre trial. *Lancet Oncol*. 2010; 11:543–552. [PubMed: 20451454]

Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*. 2005; 308:523–529. [PubMed: 15845847]

Stelzer, G.; Goodpasture, L. *Use of Multiparameter Flow Cytometry and Immunophenotyping for the Diagnosis and Classification of Acute Myeloid Leukemia*. Wilmington, DE: Wiley-Liss; 2000.

Stirewalt DL, Kopecky KJ, Meshinchi S, Appelbaum FR, Slovak ML, Willman CL, Radich JP. FLT3, RAS, and TP53 mutations in elderly patients with acute myeloid leukemia. *Blood*. 2001; 97:3589–3595. [PubMed: 11369655]

Tsukimoto I, Tawa A, Horibe K, Tabuchi K, Kigasawa H, Tsuchida M, Yabe H, Nakayama H, Kudo K, Kobayashi R, Hamamoto K, Imaizumi M, Morimoto A, Tsuchiya S, Hanada R. Risk-stratified therapy and the intensive use of cytarabine improves the outcome in childhood acute myeloid leukemia: the AML99 trial from the Japanese Childhood AML Cooperative Study Group. *J Clin Oncol*. 2009; 27:4007–4013. [PubMed: 19620491]

Zwaan CM, Meshinchi S, Radich JP, Veerman AJ, Huismans DR, Munske L, Podleschny M, Hahlen K, Pieters R, Zimmermann M, Reinhardt D, Harbott J, Creutzig U, Kaspers GJ, Griesinger F. FLT3 internal tandem duplication in 234 children with acute myeloid leukemia: prognostic significance and relation to cellular drug resistance. *Blood*. 2003; 102:2387–2394. [PubMed: 12816873]

Figure 1. Study Design and Sample Disposition

Flowchart of the study design with descriptive schematics of the samples analysed in the Training and Test (Validation) sets and sample disposition.

AML, acute myeloid leukaemia; NR, non-responder; CR, complete remission; BM, bone marrow; Rx, therapy

A: Pathways Analysed in the Training Study

B: Locked SCNP Classifier

SCNP Node		Metric	Coefficient
Modulator	Readout		
			-4.4032 (intercept)
Etoposide	cPARP	AdjU _u	7.0139 (<i>b</i> ₁)
Thapsigargin	p-Erk	log ₂ (AdjFold)	-1.3572 (<i>b</i> ₂)
FLT3L	p-S6	log ₂ (AdjFold)	0.7843 (<i>b</i> ₃)

The SCNP classifier score is defined as the probability of a patient to achieve a complete response, as calculated by the model: $\text{logit}[p(CR)] = \text{intercept} + b_1 * \text{NodeMetric}_1 + b_2 * \text{NodeMetric}_2 + b_3 * \text{NodeMetric}_3$

Figure 2.
A: Pathways Analysed in Training Study Schematic of the cell signalling pathways probed in the Training study. A SCNP node consists of the combination of a modulator and the corresponding intracellular readout. Modulators are shown initiating signalling pathways that produce an intracellular proteomic response.

B: Locked Classifiers The final SCNP assay classifier locked and tested in the independent Validation sample set. Classifier components are 3 SCNP node-metrics and the affiliated coefficients. Metrics are further described in Figure S1.

A: SCNP Assay Reproducibility in Training Study

B: SCNP Assay Reproducibility in Validation: Classifier Score

C: SCNP Assay Reproducibility in Validation: Node 1

Thapsigargin_15_pErk_Fold Node Reproducibility (n=43)

D: SCNP Assay Reproducibility in Validation: Node 2

FLT3L_15_pS6_Fold Node Reproducibility (n=43)

E: SCNP Assay Reproducibility in Validation: Node 3

Figure 3.

A: SCNP Assay Reproducibility in Training Study Duplicate samples from 14 patient samples were processed at two time-points (one year apart) and two different laboratories to evaluate the reproducibility of the SCNP classifier score. The classifier scores from Run 1 were regressed against the scores from Run 2 to calculate the Pearson correlation coefficient.

B-E: SCNP Assay Reproducibility in Validation Study Duplicate samples from 43 patients (with sufficient cells) were processed in tandem using 2 different reagent batches/cocktails on different 96-well plates to assess the reproducibility of the SCNP classifier

score. The score from Run 1 were regressed against the score from Run 2 to calculate the Pearson correlation coefficient for each of the 3 node-metrics in the classifier (nodes 1–3) as well as the final score.

Figure 4. Accuracy of the SCNP Model in Cytogenetic Risk Groups (Validation)

Relationship between the SCNP score and CR was strongest among patients with intermediate risk cytogenetics, while no relationship was observed in patients with unfavourable risk cytogenetics using a DeLong test. Induction deaths were excluded from this analysis.

Table I

Patient Clinical Characteristics for Training and Test (Validation) Set

Category		Training Study	Validation Study	<i>p</i> -Value
		n= 53	n=68	
Age (years)	Median (range)	12.3 (0.9 to 18.4)	11.65 (0.22 to 20.37)	0.402 ¹
Age Group	<2 years	3 (5.6%)	7 (10.3%)	0.666 ²
	2–10 years	19 (35.9%)	22 (32.4%)	
	>10 years	31 (58.5%)	39 (57.3%)	
Gender	Male	24 (45.3%)	38 (55.9%)	0.275 ²
	Female	29 (54.7%)	30 (44.1%)	
Cytogenetic Risk	Favourable	5 (9.4%)	20 (29.5%)	0.001 ²
	Intermediate	33 (62.3%)	27 (39.7%)	
	Unfavourable	8 (15.1%)	19 (27.9%)	
	Unknown	7 (13.2%)	2 (2.9%)	
<i>FLT3</i>-ITD Mutation	Yes	19 (35.9%)	14 (20.6%)	0.050 ²
	No	33 (62.3%)	54 (79.4%)	
	ND	1 (1.8%)	0	
<i>FLT3</i>-ITD Allelic Ratio	> 0.4	11 (20.8%)	10 (14.7%)	0.031 ²
	0.4	38 (71.7%)	58 (85.3%)	
	ND	4 (7.5%)	0 (0.0%)	
<i>NPM1</i> Mutation	Yes	6 (11.3%)	7 (10.3%)	0.462 ²
	No	45 (84.9%)	54 (79.4%)	
	ND	2 (3.8%)	7 (10.3%)	
<i>CEBPA</i> Mutation	Yes		2 (2.9%)	NA
	No		60 (88.2%)	
	ND	53 (100%)	6 (8.8%)	
Minimal Residual Disease	Yes		2 (2.9%)	NA
	No		60 (88.2%)	
	Unknown	53 (100%)	6 (8.8%)	
WBC Counts (×10⁹/l)	<100	35 (66%)	50 (73.5%)	0.425 ²
	100	18 (34%)	18 (26.5%)	
	Median	79.3	53.1	
Induction Therapy		DAT: 32 (60.4%) High-dose DAT: 21 (39.6%)	ADE+GO: 37 (54.4%) ADE: 31 (45.6%)	NA

Category		Training Study	Validation Study	p-Value
		n= 53	n=68	
Induction Therapy Response	CR	36 (67.9%)	48 (70.6%)	0.843 ²
	NR	17 (32.1%)	20 (29.4%)	
Overall Survival	at 3 years	54.5%	65.3%	0.274 ³

¹ Wilcox Test

² Fisher Test

³ Logrank Test

FLT3-ITD, fms-like tyrosine kinase 3 internal tandem duplication; ND, not done; *NPM1*, nucleophosmin 1; WBC, white blood cell; DAT, daunorubicin, cytarabine, and thioguanine; ADE, ara-C, daunorubicin, etoposide; GO, gemtuzumab ozogamicin; NA, not applicable; CR, complete response; NR, non-responder

Table II

Classifier Performance in Validation Set

Pre-specified Analysis	AUC _{ROC} (95% CI ¹)	<i>p</i> -value ² (Sample Size)
Primary (Intended to Diagnose)	0.66 (0.52, 0.78)	0.042 (n=68)
Sensitivity analysis: Induction Deaths Removed from NR	0.70 (0.55, 0.83)	0.021 (n=62)

¹ Bias corrected accelerated bootstrap method

² Wilcoxon exact test

AUC_{ROC}, area under the curve of a receiver operating characteristic curve; CI, confidence interval; NR, non-responder

Table III
Association of Clinical Variables and the SCNP Score with Induction Therapy Response
(Validation Study)

WBC and cytogenetics were associated with induction response in the Validation Set while age, gender and *FLT3*-ITD status were not. When WBC and cytogenetics are combined with SCNP classifier score, only cytogenetics and SCNP remain significant, indicating that the SCNP classifier adds independent value.

Clinical Variable* / SCNP	Associated with CR/NR	Model Including Only Clinical Variable (p-value ¹)	Model Including Significant Variables & SCNP Score (p-value ¹)
Variable available before starting Induction Rx			
Age	No	0.98	N/A
Sex	No	0.965	N/A
WBC Pre-induction	Yes	0.023	0.056
SCNP Classifier Score	N/A	N/A	0.017
Variable Available after starting Induction therapy			
Cytogenetics Risk Group ²	Yes	0.009	0.001
<i>FLT3</i> -ITD	No	0.976	N/A

* Variables were pre-specified based on prior evidence that variable can be associated with induction response in this patient population,

¹ LR chi-square test, Regression coefficient not shown,

² per NCCN guideline 2011 (O'Donnell *et al* 2011),

SCNP = Single-Cell Network Profiling, CR = complete remission, NR = non-responder, WBC = white blood cell, N/A = not applicable, *FLT3*-ITD = *FLT3* internal tandem duplication