Leukemia Research 76 (2019) 1–10

Contents lists available at ScienceDirect

Leukemia Research

journal homepage: www.elsevier.com/locate/leukres

A precision medicine test predicts clinical response after idarubicin and cytarabine induction therapy in AML patients

David Martínez-Cuadrón[a,b], Cristina Gil[c], Josefina Serrano[d], Gabriela Rodríguez[e], Jaime Pérez-Oteyza[f], Raimundo García-Boyero[g], Santiago Jiménez-Bravo[h], Susana Vives[i], María Belén Vidriales[j], Esperanza Lavilla[k], José A. Pérez-Simón[l], Mar Tormo[m], Mercedes Colorado[n], Juan Bergua[o], Juan A. López[p], Pilar Herrera[q], Pilar Hernández-Campo[r], Julián Gorrochategui[r], Daniel Primo[r], Jose Luis Rojas[r], Jesús Villoria[s], Federico Moscardó[a], Iñaki Troconiz[t], María Linares Gómez[u], Joaquín Martínez-López[u], Joan Ballesteros[r], Miguel Sanz[a,b], Pau Montesinos[a,b,*,1,2], on behalf of the Spanish PETHEMA group

[a] Hospital Universitari i Politècnic La Fe, Valencia, Spain
[b] CIBERONC, Instituto Carlos III, Madrid, Spain
[c] Hospital General Universitario de Alicante, Alicante, Spain
[d] Hospital Universitario Reina Sofía, Córdoba, Spain
[e] Hospital General Universitario Gregorio Marañón, Madrid, Spain
[f] Hospital de Madrid Norte Sanchinarro, Madrid, Spain
[g] Hospital Universitario General de Castellón, Castellón, Spain
[h] Hospital Universitario de Gran Canaria Doctor Negrín, Gran Canaria, Spain
[i] ICO-Hospital Germans Trias i Pujol, Josep Carreras Leukemia Research Institute, Universitat Autònoma de Barcelona, Badalona, Spain
[j] Hospital General Universitario de Salamanca, Salamanca, Spain
[k] Hospital Universitario Lucus Augusti, Lugo, Spain
[l] Hospital Universitario Virgen del Rocío, Sevilla, Spain
[m] Hospital Clínico Universitario, Valencia, Spain
[n] Hospital Universitario Marqués de Valdecilla, Santander, Spain
[o] Hospital San Pedro de Alcántara, Cáceres, Spain
[p] Complejo Hospitalario de Jaén, Jaén, Spain
[q] Hospital Universitario Ramón y Cajal, Madrid, Spain
[r] Vivia Biotech, Tres Cantos, Madrid, Spain
[s] Medicxact, Madrid, Spain
[t] Universidad de Navarra, Pamplona, Spain
[u] Hospital Universitario 12 de Octubre, Madrid, Spain

ARTICLE INFO

Keywords:
Acute myeloid leukemia
Ex vivo assay
Pharmacological profile
Complete remission
Clinical correlation

ABSTRACT

Complete remission (CR) after induction therapy is the first treatment goal in acute myeloid leukemia (AML) patients and has prognostic impact. Our purpose is to determine the correlation between the observed CR/CRi rate after idarubicin (IDA) and cytarabine (CYT) 3 + 7 induction and the leukemic chemosensitivity measured by an *ex vivo* test of drug activity. Bone marrow samples from adult patients with newly diagnosed AML were included in this study. Whole bone marrow samples were incubated for 48 h in well plates containing IDA, CYT, or their combination. Pharmacological response parameters were estimated using population pharmacodynamic models. Patients attaining a CR/CRi with up to two induction cycles of 3 + 7 were classified as responders and the remaining as resistant. A total of 123 patients fulfilled the inclusion criteria and were evaluable for correlation analyses. The strongest clinical predictors were the area under the curve of the concentration response curves of CYT and IDA. The overall accuracy achieved using MaxSpSe criteria to define positivity was 81%, predicting better responder (93%) than non-responder patients (60%). The *ex vivo* test provides better yet similar

* Corresponding author.
E-mail address: montesinos_pau@gva.es (P. Montesinos).
[1] Hematology Department (Torre F, planta 7), Hospital Universitario La Fe, Avinguda Fernando Abril Martorell, 106, CP 46026 Valencia, Spain. Tel.1: +34 96 1244000. Ext 411966, Tel.2: +34 96 1245876, Fax: +34 96 1246201, Web: http://www.hematologialafe.es/.
[2] CIBERONC, Instituto Carlos III, Madrid, Spain.

https://doi.org/10.1016/j.leukres.2018.11.006
Received 6 June 2018; Received in revised form 29 October 2018; Accepted 12 November 2018
Available online 13 November 2018

information than cytogenetics, but can be provided before treatment representing a valuable in-time addition. After validation in an external cohort, this novel *ex vivo* test could be useful to select AML patients for 3 + 7 regimen vs. alternative schedules.

1. Introduction

Several clinical and biological features, like cytogenetic and molecular alterations, may predict the short- and long-term outcomes in patients with AML treated with intensive approaches [1,2]. However, the main prognostic factor after front-line induction treatment is the leukemic cell sensitivity to the chemotherapy itself (i.e, to achieve or not a first CR as well as the quality and duration of the CR) [3,4]. In order to predict response to chemotherapy, many individualized sensitivity and resistance assays have been deployed for detecting *ex vivo* drug-inducible cell death. Nevertheless, as far as we know, none of them have settled into the routine medical practice [5].

The combination of IDA and CYT (3 + 7 schedule) is widely used as front-line regimen in younger AML patients, resulting in CR or CR with incomplete recovery (CRi) rates of 70–80% after one or two cycles [6]. Among patients who do not achieve CR/CRi after induction therapy, categorized as primary refractory or resistant disease, the prognosis is dismal [4]. We hypothesize that a reliable and automated *ex vivo* drug sensitivity test accurately predicts the antileukemic efficacy of 3 + 7 induction would be valuable to preselect this regimen as the optimal upfront chemotherapy schedule.

We present here a non-interventional prospective study assessing a novel and actionable native environment precision medicine (PM) method [7] (PharmaFlow platform) that could overcome some of the previous shortcomings of *ex vivo* testing. The PharmaFlow platform is a cell-based multi-color screening flow cytometry platform that incorporates both automated sample preparation and automated flow cytometry, together with proprietary analytical software and a database that achieve rapid data acquisition, analysis, and reporting of results. This platform has the capacity to evaluate hundreds of drug combinations that form the basis of treatment protocols in AML, together with the fact that these drug combinations are evaluated in whole bone marrow retaining the biological system as intact as possible. The aim is to determine the ability of the PharmaFlow PM test to predict the response to first line induction chemotherapy with CYT and IDA (3 + 7) in young AML patients.

2. Patients and methods

2.1. Patients and study design

This multicenter, non-interventional prospective cohort study was carried out in 43 Spanish institutions of the PETHEMA group. The inclusion period lasted five years (2011–2016), enrolling patients aged 18 years and older with newly diagnosed AML (de novo or secondary to myelodysplastic syndromes, or therapy-related). Patients with acute promyelocytic leukemia and those with AML derived from a chronic myeloid leukemia were not screened. Patients were eligible for the current analysis if they received front-line induction therapy using IDA/CYT (3 + 7). All adult patients treated with other front-line regimens were excluded from the correlation analysis, but their AML samples were used to build pharmacodynamics (PD) population-based models. Diagnosis and classification of AML was made according to the WHO (World Health Classification) criteria [8]. The study was conducted according to the Spanish law 14/2007 of biomedical research, and was approved by the Research Ethics Board of each participating institution. All patients provided informed consent.

2.2. Chemotherapy regimen, drugs and evaluation

Induction therapy consisted of up to two cycles of the combination of intravenous (IV) IDA (12 mg/m^2/day), from days one to three, and IV continuous perfusion of CYT (200 mg/m^2/day), from days one to seven. A second 3 + 7 induction cycle was administered in patients showing a partial remission (PR) after the first cycle. Supportive measures were given according to local policies of each participating institution.

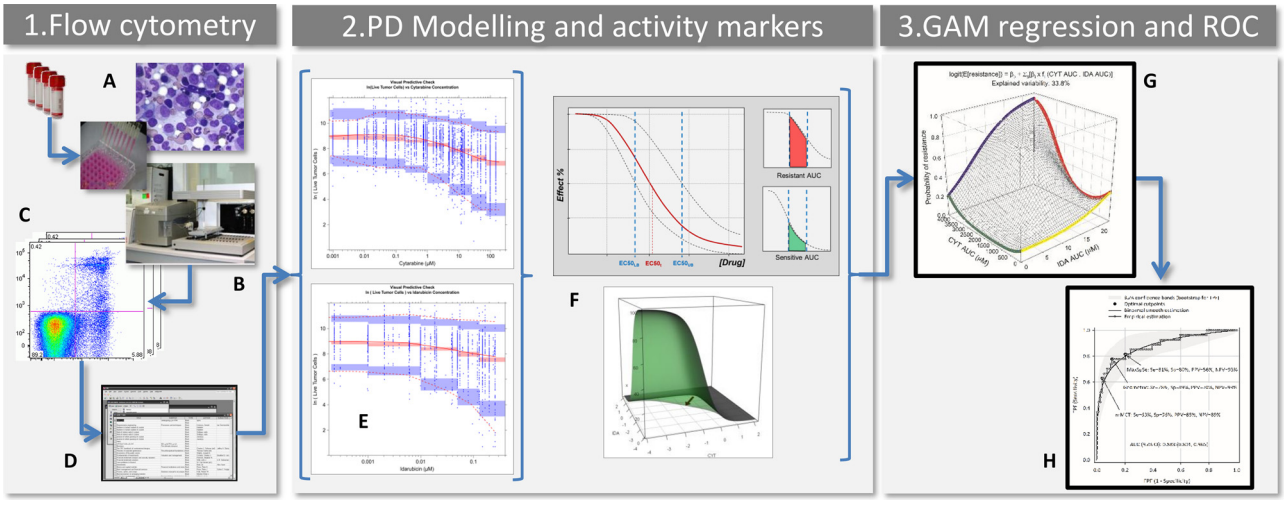

Fig. 1. Sequential workflow of experimental (1) and analytical (2 & 3) methods applied in the study. Whole bone marrow samples [A] were incubated preserving the native microenvironment with drugs and drugs mixtures. Automated flow cytometry [B] followed by dot-plots analysis [C] allowed the counting after incubation of Live Pathologic Cells (LPC) at control wells and wells with increasing drugs concentrations. Data was uploaded into the LIMS system. Response vs drug concentration relationships were analyzed through non-linear mixed effect population modelling [E]. Predicted pharmacodynamic profiles were integrated between the 80% confidence interval of the individual estimate of EC_{50} in order to calculate the area under the curve (AUC) used as a single activity marker. Similarly, a double integration of the two-variables interaction surface function allows the calculation of the volume under the surface (VUS) that is effected by the sign (synergy or antagonism) of the interaction [F]. Correlation of activity markers with clinical output was analyzed by Generalized Additive Models (GAM) [G] and ROC curves [H]. (See details on individual figures).

Table 1
Patients' characteristics and clinical response of the 123 patients incubated with CYT and IDA.

	Clinical response (ref. standard)		
	CR/CRi (N = 92)	PR/Res (N = 31)	p-value
Age (years)			
Median (range)	49 (19-71)	55 (22-67)	0.224 [a]
18–29 [n (%)]	8 (8.7)	2 (6.5)	0.758 [b]
30–39 [n (%)]	17 (18.5)	5 (16.1)	
40–49 [n (%)]	23 (25.0)	5 (16.1)	
50–59 [n (%)]	22 (23.9)	10 (32.3)	
> 60	22 (23.9)	9 (29.0)	
Gender			
Male [n (%)]	42 (45.7)	16 (51.6)	0.565 [b]
Female [n (%)]	50 (54.3)	15 (48.4)	
ECOG			
0–1 [n (%)]	69 (87.3)	27 (100.0)	0.062 [c]
2–4 [n (%)]	10 (12.7)	0 (0.0)	
FAB subtype			
M0 [n (%)]	3 (3.6)	5 (17.9)	0.166 [b]
M1 [n (%)]	19 (22.6)	4 (14.3)	
M2 [n (%)]	23 (27.4)	9 (32.1)	
M4 [n (%)]	22 (26.2)	5 (17.9)	
M5 [n (%)]	16 (19.0)	5 (17.9)	
M6 [n (%)]	1 (1.2)	0 (0.0)	
WBC (count × 10^9L^{-1})			
Median (range)	22.1 (0-288.4)	22.3 (1-157)	0.461 [a]
0–10 [n (%)]	31 (33.7)	12 (38.7)	0.528 [b]
10–50 [n (%)]	37 (40.2)	14 (45.2)	
> 50 [n (%)]	24 (26.1)	5 (16.1)	
Cytogenetic risk profile			
Favorable [n (%)]	11 (13.1)	0 (0.0)	< 0.001 [b]
Intermediate [n (%)]	68 (81.0)	11 (40.7)	
Adverse [n (%)]	5 (6.0)	16 (59.3)	
FLT3-ITD status			
Wild type [n (%)]	72 (79.1)	28 (90.3)	0.161 [b]
Mutant [n (%)]	19 (20.9)	3 (9.7)	
NPM1 status			
Wild type [n (%)]	47 (57.3)	22 (78.6)	0.045 [b]
Mutant [n (%)]	35 (42.7)	6 (21.4)	

Patients with missing data have not been included in the denominators of the relative frequencies.
[a] Mann-Whitney test.
[b] Pearson's chi-square test.
[c] Fisher's exact test.

Response to induction chemotherapy was assessed according to the revised Cheson criteria [9]. Patients were considered as responders if they achieved CR or CRi within the first two identical 3 + 7 induction cycles. Patients dying during induction before response assessment were considered as non-evaluable. The remaining patients were classified as resistant.

2.3. Vivia's PharmaFlow PM test

A representative workflow of the PharmaFlow PM test is shown in Fig. 1, collecting the experimental and analytical methods applied in this study.

2.3.1. Native environment whole bone marrow sample

Ex vivo drug sensitivity analysis was made using the PharmaFlow platform (previously termed ExviTech®) [7] maintaining the bone marrow (BM) microenvironment. A minimum BM sample volume between one and two ml was collected by aspiration at AML diagnosis, before starting induction chemotherapy, and was processed by an automated method in Vivia Biotech laboratories 24 h after extraction. Samples were incubated with CYT, IDA and/or CYT + IDA for 48 h. A more detailed description of the procedure has been published elsewhere [7].

2.3.2. Modeling of ex vivo activity of CYT, IDA, and their combination

Evaluation of drug response was done by counting the number of live pathological cells (LPC) remaining after incubation at increasing drug concentrations [10]. Dying cells (apoptosis) were excluded using Annexin V-FITC. Pharmacological responses were analyzed using PD population-based models [10] which essentially perform the fitting of the dependent variable (natural log of LPC) in a non-linear mixed-effects model to derive typical population values (fixed effects) and the magnitude of inter-patient and residual variability (random effects). Model development was performed with the first-order conditional estimation method using interaction option with the software NONMEM (v7.2) [11], according to the following equation:

Where LPC_0 parameter refers to the number of LPC after incubation in the absence of drug, E_{max} represents the maximum fractional decrease in LPC that the drug can elicit, EC_{50}, is the drug concentration exerting half of E_{max}, and γ is the parameter governing the steepness of the LPC vs drug concentration (C) curve. For interaction analysis a Surface Interaction model [12] was used to estimate the degree of synergy, referred as α parameter, between both drugs.

Interpatient variability (IPV) associated to all parameters was described by means of an exponential model of the components of variance. An additive error structure was used for the residual variability. Population PD models were built with BM samples from 473 patients that were incubated with CYT, 456 with IDA and 443 with CYT + IDA. Bayesian estimation methods were then used to retrieve individual patient parameters based on their available exposure-response measurements in conjunction with the PD population parameters.

Evaluation of the population PD models was done by the simulation-based procedure visual predictive check [13]. Five hundred experimental scenarios equal to the original ones were simulated using the selected models and the corresponding parameters. In each simulated set and for each concentration lever the 2.5, 50, and 97.5th percentiles of the LPC distribution were calculated, then the 95% confidence intervals for the above mentioned percentiles were computed and represented graphically together with the 2.5, 50, and 97.5th percentiles obtained from the raw data.

2.3.3. Probability of clinical outcome modeling clinical correlation

Individual response profiles normalized with respect to LPC_0, were integrated between the concentration points corresponding to the 20^{th} and 80^{th} percentiles of the distribution of estimated individual EC_{50} values, to obtain the values of the areas under the curves (AUCs) that were used as a descriptor of the *ex vivo* drug effect (i.e., the higher the AUC, the lower the cytotoxic effect (efficacy or potency) of the drug.

The individual AUC values were correlated to the actual patient´s response after induction therapy (non-responder [PR or resistant disease] vs. responder [CR or CRi]). The probability of being non-responder was modeled using binary logistic generalized additive models (GAM) based on the binomial distribution that included either one bi- or two univariate smooth functions of the AUCs of CYT and IDA. Additionally, univariate smooth functions of the LPC_0, α, and the pre-post incubation difference of the percentage of LPC in control wells (to detect for any possible effect of spontaneous cell death) were included as well, but discarded afterwards because they were not related to the clinical response. Also, the predictive ability of relevant patients' characteristics (age and sex, presenting leukocyte count, performance status, mutations in the NPM1 or FLT3 genes, and cytogenetic risk group) on top of pharmacodynamics data was explored by introducing them as parametric model terms in auxiliary GAMs. P-spline bases were used as smoothers for univariate smooth functions; tensor products of univariate P-spline smooths were used for constructing bi-variate smooth functions. All smoothing bases had dimension three.

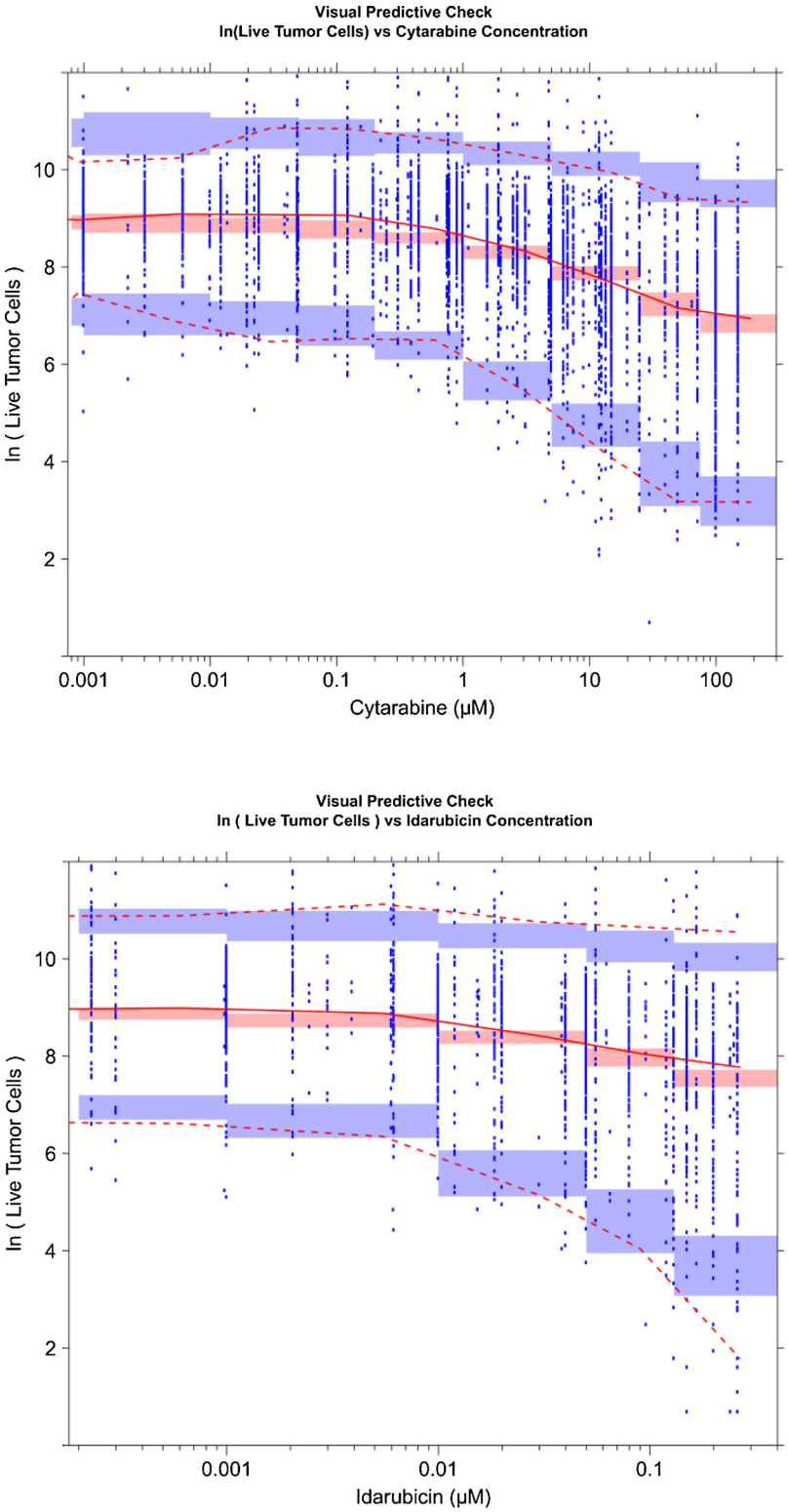

Fig. 2. Visual predictive check of the population pharmacokinetic models of cytarabine and idarubicin. Open circles are the observed data points, the solid and dashed red lines are, respectively, the median and the 5-95th percentiles of the observed distribution of ln(cells), and the semitransparent red and blue bands represent, respectively, the simulation-based 95% confidence intervals for the median and 5-95th percentiles of the estimated population distribution of ln(cells).

Coefficients of the smooth functions were estimated using penalized iteratively re-weighted least squares. Minima of the scaled Akaike information criteria were used to find the optimal values of the smoothing parameters.

2.4. Data collection and study endpoints

Demographic data (gender, age) were prospectively collected since diagnosis, as well as the following parameters: WBC in PB, ECOG performance status, type of AML (FAB classification, *de novo* vs. secondary AML), karyotype [14], FLT3 and NPM1 mutation status, hematological

Table 2
Estimation of the *ex vivo* population pharmacodynamic parameters. Parameters typical and random (variability and residual error percentage) are shown together with the corresponding relative standard error calculated as the ratio between the standard error provided by NONMEM and the estimate. Estimates of inter-patient variability (IPV) are expressed as coefficient of variation (%).

Single Drugs		
Parameter (units)	Cytarabine	Idarubicin
LPC_0 (cells)	7530 (4.2)	7270 (4.8)
E_{MAX} (unitless)	1 (-)	1 (-)
EC_{50} (μM)	6.94 (13.3)	0.087 (9·2)
γ(unitless)	0.684 (-)	1.14 (-)
Residual Error ($\log_{(\mu}$M))	0.231 (2.9)	0.237 (3.2)
Inter-patient variability (IPV)		
LPC_0	89.7 (2.4)	92.7 (2.6)
E_{MAX}	N/D	N/D
EC_{50}	229 (4.4)	159 (4.5)
γ	N/D	N/D
Residual Error	50.5 (3.8)	45.8 (4)

Drug Combination Parameter (units)	Cytarabine + Idarubicin
α (unitless)	1.1 (13)
IPV α [CV(%)]	176 (4·8)
Residual Error [$\log_{(\mu}$M)]	0.299 (4.2)
IPV Residual Error	58.3 (4.4)

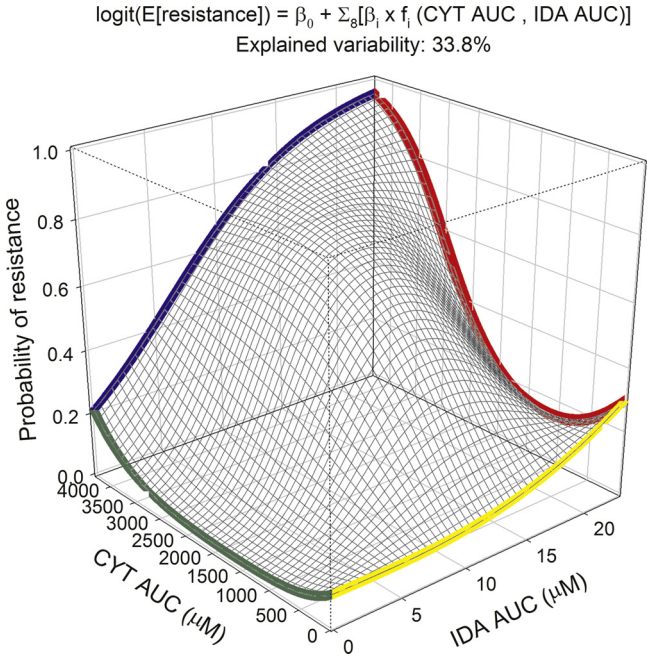

Fig. 3. Regression hyperplane of the predicted probability of resistance over the AUCs of cytarabine and idarubicin. The AUCs are a summary of pharmacodynamic parameters such that the higher the AUC the lower the cytotoxic effect (efficacy or potency) of the drug. The regression hyperplane has been obtained using bi-dimensional smooth functions in a binary logistic GAM. AUC: area under the curve, GAM: generalized additive model.

response, number of 3 + 7 induction cycles, date of response, date of last follow-up, and post-remission therapy. All data collection forms and clinical records were monitored.

The primary end-point was evaluate the predictive capacity of the *ex vivo* results. First, the CR/CRi rate observed in patients treated with up to two induction cycles of 3 + 7 was recorded and monitored. This was correlated with the *ex vivo* drug sensitivity analyses performed in the same cohort of patients. As a secondary end-point, the overall survival (OS) probability was also calculated according to the observed and predicted response after induction.

Since the prediction of outcome to front-line induction therapy is likely to be most beneficial to elderly patients, we performed sensitivity analyses of the clinical correlation by re-running the GAMs within the cohort aged 60 years or more (n = 31).

2.5. Statistical analyses

The probability of response modeling was performed with the mgcv package (v1.8–23) run in the R environment (v3.4.3) for statistical computing [15]. The empirical ROC curves were calculated for the probabilities of being non-responder from each GAM. The AUCs of ROC curves were computed using the trapezoidal rule. In addition, three cut-points to define positivity and derive classification probabilities (sensitivity and specificity) were established for each ROC curve. One used a geometric criterion, by selecting the closest point to the (1,0) coordinate (left upper corner of the [sensitivity,1 − specificity] plane), another was set by maximizing both sensitivity and specificity (MaxSpSe) and the other by minimizing a misclassification cost term (mMCT) [16], assigning a greater cost to false positives than to false negatives (prioritizing specificity over sensitivity).

The OS was described with the Kaplan-Meier method and compared between the patients predicted to be non-responder and responder as per the aforementioned three different cut-points using simple Cox regression.

3. Results

3.1. Patient characteristics

Overall, 954 BM samples from patients with AML suspicion were received at the laboratory. Of them, 316 (33%) were not evaluable because of the following laboratory technical issues: 1) low sample cellularity (187 patients), 2) low cell viability (below 60%) in control wells after incubation (67), 3) insufficient sample volume (< 500 μL) (38), and 4) other reasons such as clotted sample (24). Other 26 patients (3%) did not fulfill the diagnosis criteria. Among the 612 analyzed samples, 139 where used only for assay adjustment and did not contain necessary data for the final model. Overall, 473 patients samples (50%) were used to build the PD models, and a complete data set was monitored in 237 of them (50%). Among the monitored patients, 114 were not evaluable for the correlation analyses due to: 1) induction death (20 patients), 2) not first line of treatment (11), and 3) other induction schedule (83). Finally, 123 monitored patients (52%) fulfilled the inclusion criteria defined in the study and were evaluable for the correlation analyses (see CONSORT diagram in supplementary Fig. 1). The main patient and disease characteristics of these 123 patients regarding the clinical response are displayed in Table 1. In summary, median age was 50 years (range, 19 to 71), 109 patients (89%) were diagnosed with *de novo* AML, and 21 patients (17%) were categorized as having high-risk cytogenetics. Only the cytogenetic risk group and, marginally, the presence of mutations in the NPM1 gene were significantly associated with clinical response to induction and the result of the PM test. Post-remission therapy consisted of allogeneic stem cell transplant (SCT) in 33 patients (27%), and chemotherapy with or without autologous SCT in 66 patients (54%).

3.2. Ex vivo PharmaFlow test characterization of CYT-IDA combination

Visual predictive checks graphs were generated for the single drugs PD models (Fig. 2). Most of the observations were contained within the simulation-based 95% confidence intervals of the 2.5–97.5 th population percentiles proving good predictability of the selected models. Pharmacodynamic population parameters as well as variability and error values are shown in Table 2. The typical parameter values for

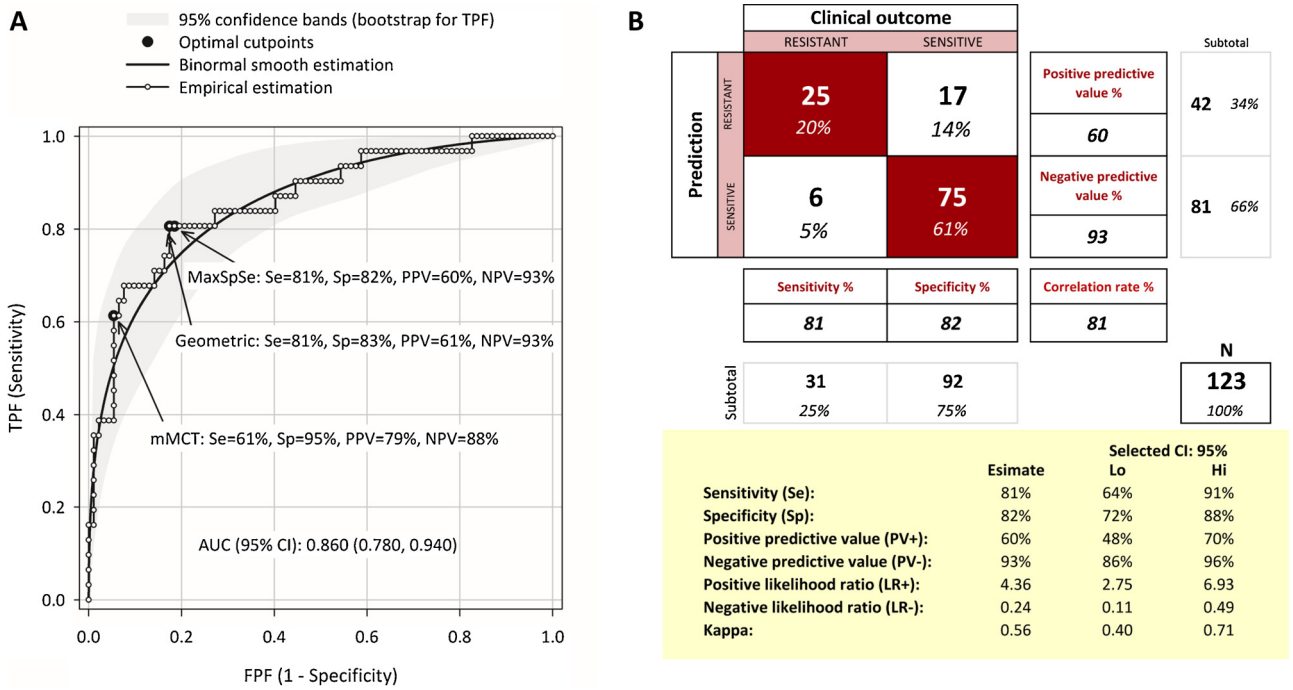

Fig. 4. Empirical and smoothed (binormal) ROC curves of the probability of resistance obtained in the binary logistic GAM. A. - Open circles are the pairs of sensitivity and 1 − specificity values at the estimated discrete individual values of the probability of resistance (used as a marker to classify the patients as responder or resistant), the solid large circles represent the pairs of sensitivity and 1 − specificity values at the selected cut-points that were obtained with each of the three criteria specified in the text: 'MaxSpSe' selects the point that maximizes both, the sensitivity and the specificity; 'Geometric' selects the closest point to the (1,0) coordinate (left upper corner of the [sensitivity,1 − specificity] plane); and 'mMCT' selects the point that minimizes a misclassification cost term that assigned a greater cost to false positives than to false negatives (prioritizes specificity over sensitivity).
AUC: area under the curve, CI: confidence interval, FPF: false positive fraction, NPV: negative predictive value, PPV: positive predictive value, Se: sensitivity, Sp: specificity, TPF: true positive fraction.
B.- Confusion matrix for MaxSpSe cutoff.

maximal fractional effect (E_{max}) were set to 1 for both drugs, and was limited to the range 0-3. The typical value for the alpha parameter of the interaction model was 1.1 (Table 2), indicating slight synergistic interaction between IDA and CYT in the *ex vivo* combination experiments.

3.3. Clinical responses among AML patients treated with CYT-IDA

CR/CRi was achieved after one (88, 96%) or two (4, 4%) identical induction cycles in 92 out of 123 patients (75%) included in the correlation study.

3.4. Correlation between ex vivo activity and clinical response to CYT-IDA

Fig. 3 depicts the predicted surface fitted by the GAM representing the probability of being non-responder for the observed range of the individual AUC values. The model presented used a bivariate smooth function of CYT and IDA; the models that used univariate smooths (see Supplementary Fig. 2) achieved worse fit. Higher CYT and IDA AUC values were associated with greater probability of being non-responder, albeit the relationship was non-monotonical. Sensitivity/specificity values ranged from 81% / 82%, to 61% / 95%, based on the cut-point selected (Fig. 4A). The positive/negative predictive values (PPV/NPV) ranged from 60% / 93% to 79% / 88%. Fig. 4B shows the confusion matrix obtained using the MaxSpSe cutoff point.

In the auxiliary models, only the cytogenetic risk group provided independent additional predictive information on top of CYT and IDA pharmacodynamics, the latter remaining significantly associated with response no matter of patients' clinical characteristics. A simple logistic model of the probability of being non-responder over the cytogenetic risk group (favorable/intermediate/adverse) explained less variability (29.4%) than the GAM over the AUC values (40.8% in the subset of 111 patients in whom the cytogenetic risk was informed) (Fig. 6). Furthermore, the model with the pharmacodynamics data explained significantly more variability than the cytogenetic information alone ($p = 0.001$, details available or request).

The sensitivity analyses showed that the predictive ability of the PM test remained intact within the cohort aged ≥60 years (Supplemental Fig. 3), although in this case most of the discriminative information was provided by CYT data; IDA AUC values were in general higher in older patients.

3.5. OS according to the ex vivo activity and observed clinical response

The OS was significantly shorter in patients predicted to be non-responders than in patients predicted to be responders regardless of the cut-point used to classify them. The median OS among patients predicted to be non-responders ranged from 344 to 589 days (Fig. 5). It was not reached in patients predicted to be responders. The hazard ratios (HR) of death (patients predicted to be non-responders vs. responders) ranged from 2.46 (1.38–4.36) to 3.44 (1.88–6.28). The values for the groups defined by actual clinical response were similar (median OS among non-responders: 279 days; HR [resistant vs. CR/CRi]: 3.17).

4. Discussion

This study shows good correlation between the hematological response to IDA/CYT 3 + 7 induction and the observed leukemic chemosensitivity measured by a novel approach to *ex vivo* testing of drug activity. The statistical model that was built using the PharmaFlow PM test showed a good correlation (81%) with hematological responses observed in patients. A 93% of patients predicted sensitive indeed

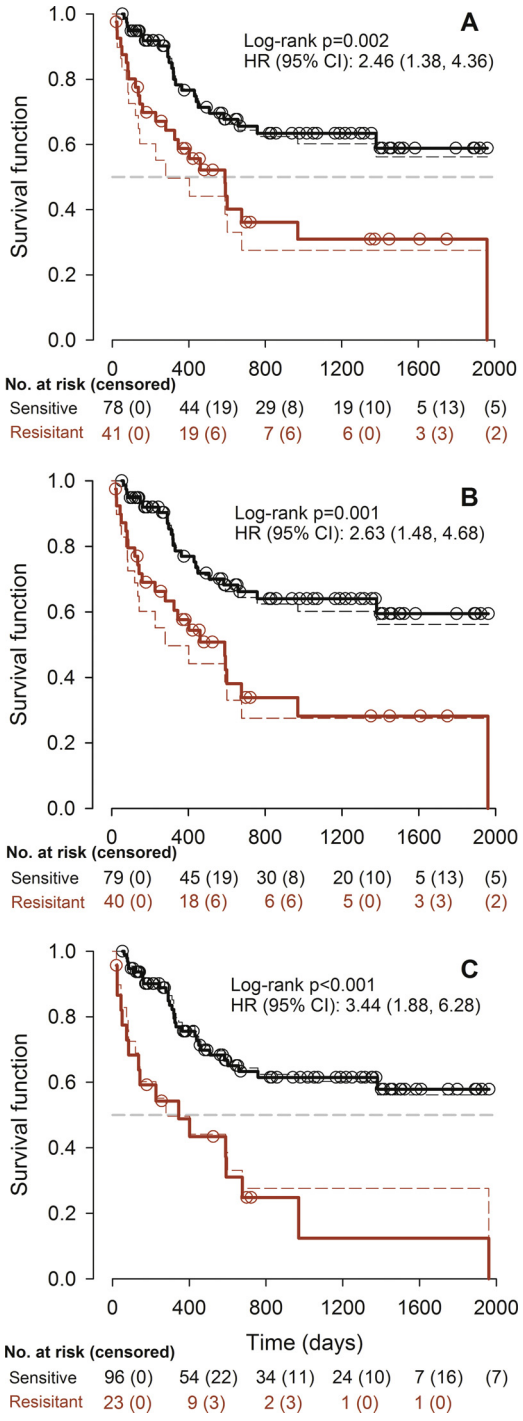

Fig. 5. Kaplan-Meier plots of overall survival. The three panels deptict the pairs of survival functions for patients classified as responder (solid black lines) and resistant (solid red lines) according to the cut-points of the estimated probability of resistance that were obtained with each of the three criteria specified in the text (panel A: 'MaxSpSe', panel B: 'Geometric', and panel C: 'mMCT'). The dashed lines represent the survival functions of clinical responders (black lines) and resistant patients (red lines). The hazard ratios of death were obtained from a Cox regression model that used the patients who were predicted to be responder as the reference category (patients predicted to be resistant over patients predicted to be responder).
CI: confidence interval, HR: hazard ratio

achieved CR/CRi, potentially allowing for better selection of the frontline chemotherapy schedule. However, a validation study in an external cohort must be performed before using the PharmaFlow PM test in the upfront clinical setting.

Several studies have analyzed the correlation between *ex vivo* drug testing and clinical outcomes in adult AML patients by different methods [17–23], but none was adopted in the routine clinical practice. Our novel approach was designed to overcome some of the *ex vivo* testing limitations of the last 30 years. Thus, the PharmaFlow PM test has several relevant aspects differing from previously reported assays: 1) contrarily to others using ficol isolated leukemic cells, we used whole BM samples diluted to keep, at least in part, the native environment. We have shown this prevents major artifacts on drug activity [7]. As the tumor microenvironment has been implicated in the drug-resistance mechanisms [24–27], we can hypothesize that using whole BM samples could lead to more reliable results; 2) we employed PD models to analyze the data albeit the correlations were performed in relatively small number of patients. This method increases accuracy by fitting dose-response population data in one single step [28]; and 3) we used a proprietary automated flow cytometry platform (PharmaFlow) that could lower experimental errors, providing accurate data from patient samples.

Our study methodology has several limitations that deserve some comments: 1) the study was designed to create a predictive model for the clinical response in terms of CR/CRi vs. PR/resistance, using the classical criteria, whilst no systematic/centralized minimal residual disease assessment was performed. We can speculate that a more precise response assessment by flow cytometry could improve the accuracy of our observations; 2) the PharmaFlow PM test analyzed the main baseline leukemic population, not facing up the presence of phenotypically different subclones before and after induction; 3) the test was not predicting the actual overall response (it was not aimed to predict induction death); 4) the test was not evaluable in a sizable proportion of patients mainly due to low cellularity or low viability samples; 5) although the incubation time was relatively short, additional transportation and processing time could lead, in some patients, to a non-affordable delay to start induction chemotherapy while receiving the test report; and 6) our results are not yet validated in an independent cohort. Finally, we should speculate about some potential causes about why the PharmaFlow could not predict some selected cases: 1) the test analyzed the main leukemic population, not facing up the presence or marginal resistant subclones, 2) the test was not reproducing the leukemic stem-cell niche who could have an *in vivo* protective effect against chemotherapy agents, 3) the test was not taking into account some pharmacodynamic and pharmacokinetic parameters which may influence the actual response to chemotherapy (e.g., drug-drug interactions during induction, pharmacogenetic polymorphisms); and 4) technical issues/limitations of the test.

To identify the *ex vivo* variables associated to hematological response, sophisticated regression analyses were performed. The AUC of the concentration-response curves yielded a useful summary of the pharmacodynamic parameters for the purposes of predicting the clinical response. Both drugs provided meaningful and independent information, as can be observed in predicted surface of being non-responder shown in Fig. 3; When IDA is inactive at maximal IDA AUC, the AUC of CYT leads the prediction with a classical dose-response curve (red). Conversely, When CYT is inactive IDA leads the prediction with a classical dose response curve (blue). When one of the drugs is very active (low AUC) the other shows a more limited effect, still consistent with higher AUC corresponding to higher probability of resistance. This behavior is coherent and consistent with the expectations; higher AUC of either drug, i.e. lower activity, implies higher probability of resistance.

Notably, the observed interaction between both drugs was not highly synergistic, and did not independently correlate with hematological response. Contrarily to other reports [29], our data suggest that the effectiveness of the CYT-IDA combination may rely in their complementary activities, and marginally on their synergism.

Diagnostic accuracy of a predictive test is usually assessed by the

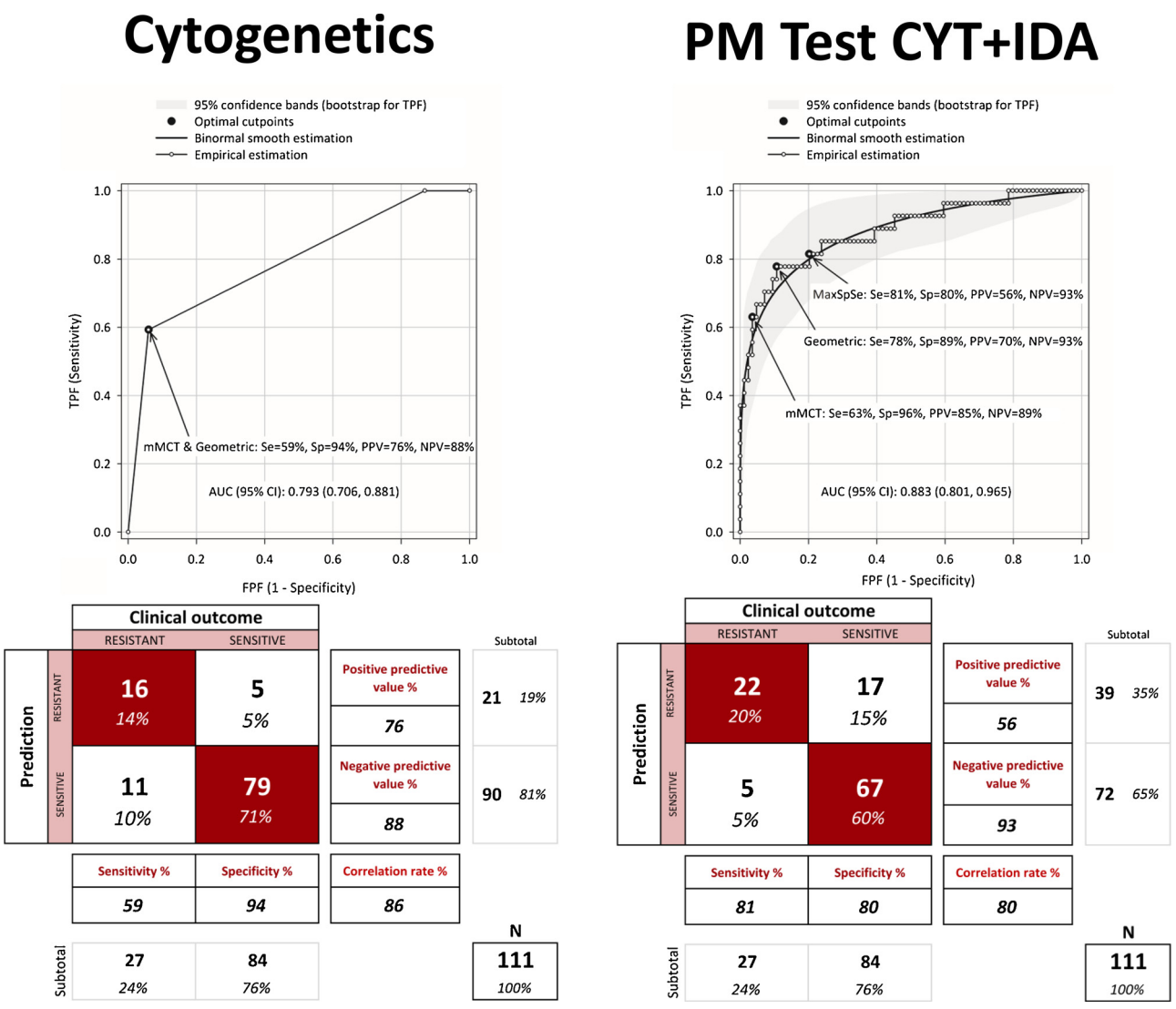

Fig. 6. Comparison between clinical correlation of cytogenetics (left) and PM Test (right), on a cohort of 111 patients sharing both results. ROC curves (top) and confusion matrices for MaxSpSe cutoff (down). Deviance explained is 29.4% for cytogenetics, and 40.9% for PM test.

maximum sensitivity and specificity values achieved in ROC curves (Fig. 4A, "maxSpSe"). However, we considered an alternative approach to maximize the ability to identify patients as non-responder (i.e, PPV) or responder (i.e, NPV), while prioritizing specificity over sensitivity ("mMCT") to account for the high clinical cost of false positives (denying potentially effective 3 + 7 schedule). Although the geometric cutoff point balanced both aspects, the final selected cutoff point in ROC curve was the MaxSpSe to construct the confusion matrix and derive classification probabilities, achieving high values for both specificity and sensitivity, as well as good PPV and NPV (Fig. 4B). The accuracy achieved by the PharmaFlow PM test using the aforementioned MaxSpSe criteria to define positivity was 81%, predicting better responder (93%, NPV) than non-responder patients (60%, PPV). We can speculate that, in order to improve first CR/CRi rates, patients predicted as non-responder to CYT-IDA could be proposed for alternative induction schedules or front-line clinical trials.

The estimated OS was significantly better in patients predicted to be responder. This is not surprising, because PharmaFlow PM test accurately predicted the response, and achieving a first CR leads to improved OS after 1st line in AML [30]. Interestingly, the OS probability was similarly discriminated by the test result (responder vs. non-responder) compared with the actual clinical response (CR/CRi vs. resistance).

The information about the cytogenetic risk was less predictive than the pharmacologic parameters; Deviance explained 29.4% for cytogenetics, and 40.8% for PM test. Nonetheless, cytogenetics still proved to be of prognostic relevance in these patients. Fig. 6 compares clinical correlation of cytogenetics vs PM Test, on a cohort of 111 patients sharing both results. In both approaches prediction of sensitive patients is better than resistant patients (NPV vs PPV). While PM Test has better prediction of sensitive patients (NPV = 93% vs 88%), cytogenetics shows a 20% improvement in the prediction of resistant patients. However, results from cytogenetic risk are available typically in 10–14 days and thus after patient treatment, while results from this novel *ex vivo* PM Test are available in 48–72 h prior to patient treatment. It should be noted that the model showed an acceptable performance among patients older than 60 years old (which are at higher risk of treatment related death), and retained its independent predictive value among the adverse cytogenetic subgroup of patients (which are frequently offered non-intensive options because of high resistance rate). We can hypothesize that the PM test could be valuable for selecting candidates to IDA and CYT 3 + 7 among these critical subsets.

In conclusion, our novel approach to *ex vivo* testing using the PharmaFlow PM platform provided drug sensitivity parameters that were integrated in a flexible generalized additive logistic regression model with an outstanding predictive accuracy for hematological

response after front-line induction with CYT-IDA 3 + 7. After validation in an external cohort, our diagnostic tool could be useful to select AML patients for 3 + 7 regimen vs. alternative schedules. The PETHEMA AML group is launching a clinical trial to confirm the predictive value of this new PM Test.

Acknowledgements

This study was supported by grants from the Spanish Ministry of Economy, Industry and Competitiveness, Enisa (Empresa Nacional de Innovación, S.A.) and Soprea (Sociedad para la Promoción y Reconversión Económica de Andalucía, S.A.U.).

DMC., JML, JB and PM designed the study, supervised research, and wrote the paper; CG, JS, GR, JPO, RGB, SJB, SV, MBV, EL, JAPS, MT, MC, JB, JAL, PH and MLG contributed samples/patient data and provided critical input; PHC and DP supervised research, analyzed the data and provided critical input; JG, JLR, JV and IT performed the statistical data analysis, supervised research and provided critical input; FM and MS design the study and supervised research. All authors contributed to the preparation of the manuscript and approved the submission in its current form.

We would like to thank the following investigators and institutions from the Spanish PETHEMA group for their support and collaboration: Dr. Carmen Burgaleta (Hospital Príncipe de Asturias, Alcalá de Henares, Madrid), Dr. Mª Ángeles Fernández (Hospital Xeral Cies, Vigo), Dr. Juan Antonio Vera (Hospital Virgen de la Macarena, Seville), Dr. Arancha Alonso (Hospital Universitario Quirón, Madrid), Dr. Jordi Sierra (Hospital de la Santa Creu y Sant Pau, Barcelona), Dr. Ataulfo González (Hospital Clínico San Carlos, Madrid), Dr. Begoña Navas (Hospital Universitario Moncloa, Madrid), Dr Esperanza Lavilla Rubira (Hospital Lucus Augusti, Lugo), Dr. Adriana Simiele (Hospital Povisa, Pontevedra), Dr. Concepción Bethancourt (Hospital Regional Universitario de Málaga Carlos Haya, Málaga), Dr. Guiomar Bautista (Hospital Universitario Puerta de Hierro, Majadahonda, Madrid), Dr. Jose Ángel Hernández Rivas (Hospital Infanta Leonor, Madrid), Dr. Bernando J. González (Hospital Universitario de Canarias, Santa Cruz de Tenerife, Canarias), Dr. Mª Asunción Mora Casado (Hospital Infanta Sofía, Madrid), Dr. Adolfo de la Fuente (MD Anderson Cancer Center, Madrid), Dr. Carlos Javier Cerveró (Hospital Virgen de la Luz, Cuenca), Dr. Consolación Rayón (Hospital Universitario Central de Asturias, Oviedo), Dr. Teresa Olavé (Hospital Clínico Universitario Lozano Blesa, Zaragoza), Dr. Aurelio López (Hospital Arnau de Vilanova, Valencia), Dr. Olga Salamero (Hospital Universitari Vall d'Hebron), Dr. Edelmira Martí (Hospital de Manises, Valencia), Dr. Juan Muñoz (Hospital Puerta del Mar, Cadiz), Dr. Guillermo Debén Ariznavarreta (Hospital Universitario de A Coruña, A Coruña), Dr. Fernando Ramos (Hospital Universitario de León, León), Dr. María Lourdes Amador (Hospital de Montecelo), Dr. Manuel Mateo Pérez (Complejo Hospitalario Universitario de Santiago, Santiago de Compostela, A Coruña) and Dr. Ana Kerguelen (Hospital Universitario la Paz, Madrid). We also thank the patients for providing their samples.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.leukres.2018.11.006.

References

[1] E.H. Estey, Acute myeloid leukemia: 2013 update on risk-stratification and management, Am. J. Hematol. 88 (2013) 318–327.
[2] M.A. Sanders, P.J. Valk, The evolving molecular genetic landscape in acute myeloid leukaemia, Curr. Opin. Hematol. 20 (2013) 79–85.
[3] J.P. Sasine, G.J. Schiller, Emerging strategies for high-risk and relapsed/refractory acute myeloid leukemia: novel agents and approaches currently in clinical trials, Blood Rev. 29 (2015) 1–9.
[4] F. Thol, R.F. Schlenk, M. Heuser, A. Ganser, How I treat refractory and early relapsed acute myeloid leukemia, Blood 126 (2015) 319–327.
[5] D. Schrag, H.S. Garewal, H.J. Burstein, D.J. Samson, D.D. Von Hoff, M.R. Somerfield, American Society of Clinical Oncology Technology Assessment: chemotherapy sensitivity and resistance assays, J. Clin. Oncol. 22 (2004) 3631–3638.
[6] M. Low, D. Lee, J. Coutsouvelis, S. Patil, S. Opat, P. Walker, A. Schwarer, H. Salem, S. Avery, A. Spencer, A. Wei, High-dose cytarabine (24 g/m^2) in combination with idarubicin (HiDAC-3) results in high first-cycle response with limited gastrointestinal toxicity in adult acute myeloid leukaemia, Intern. Med. J. 43 (2013) 294–297.
[7] T.A. Bennett, P. Montesinos, F. Moscardo, D. Martinez-Cuadron, J. Martinez, J. Sierra, R. Garcia, J.P. de Oteyza, P. Fernandez, J. Serrano, A. Fernandez, P. Herrera, A. Gonzalez, C. Bethancourt, G. Rodriguez-Macias, A. Alonso, J.A. Vera, B. Navas, E. Lavilla, J.A. Lopez, S. Jimenez, A. Simiele, B. Vidriales, B.J. Gonzalez, C. Burgaleta, J.A. Hernandez Rivas, R.C. Mascunano, G. Bautista, J.A. Perez Simon, L. Fuente Ade, C. Rayon, I.F. Troconiz, A. Janda, A.G. Bosanquet, P. Hernandez-Campo, D. Primo, R. Lopez, B. Liebana, J.L. Rojas, J. Gorrochategui, M.A. Sanz, J. Ballesteros, Pharmacological profiles of acute myeloid leukemia treatments in patient samples by automated flow cytometry: a bridge to individualized medicine, Clin. Lymphoma Myeloma Leuk. 14 (2014) 305–318.
[8] J.W. Vardiman, N.L. Harris, R.D. Brunning, The World Health Organization (WHO) classification of the myeloid neoplasms, Blood. 100 (2002) 2292–2302.
[9] B.D. Cheson, J.M. Bennett, K.J. Kopecky, T. Buchner, C.L. Willman, E.H. Estey, C.A. Schiffer, H. Doehner, M.S. Tallman, T.A. Lister, F. Lo-Coco, R. Willemze, A. Biondi, W. Hiddemann, R.A. Larson, B. Lowenberg, M.A. Sanz, D.R. Head, R. Ohno, C.D. Bloomfield, 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. 21 (2013) 4642–4649.
[10] R.N. Upton, D.R. Mould, Basic concepts in population modeling, simulation, and model-based drug development: part 3-introduction to pharmacodynamic modeling methods, CPT Pharmacometrics Syst. Pharmacol. 3 (2014) e88.
[11] S.L. Beal, L.B. Sheiner, A.J. Boeckmann, R.J. Bauer, NONMEM Users Guides. Icon Development Solutions, Ellicot City, Maryland, 1989–2001.
[12] W.R. Greco, G. Bravo, J.C. Parsons, The search for synergy: a critical review from a response surface perspective, Pharmacol. Rev. 47 (1995) 331–385.
[13] M. Bergstrand, A.C. Hooker, J.E. Wallin, M.O. Karlsson, Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models, AAPS J. 13 (2011) 143–151.
[14] D. Grimwade, R.K. Hills, A.V. Moorman, H. Walker, S. Chatters, A.H. Goldstone, K. Wheatley, C.J. Harrison, A.K. Burnett, Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials, Blood 116 (2010) 354–365.
[15] S.N. Wood, Generalized Additive Models. An Introduction with R. Chapman & Hall, CRC, Boca Raton, Florida, 2006.
[16] M. Greiner, Two-graph receiver operating characteristic (TG-ROC): update version supports optimisation of cut-off values that minimise overall misclassification costs, J. Immunol. Methods 191 (1996) 93–94.
[17] A.G. Bosanquet, P. Nygren, L.M. Weisenthal, Individualized tumor response testing in leukemia and lymphoma, in: G.J. Kaspers, B. Coiffier, M.C. Heinrich, E.H. Estey (Eds.), Innovative Leukemia and Lymphoma Therapy, Informa Healthcare, New York, 2008, pp. 23–44.
[18] K.R. Jun, S. Jang, H.S. Chi, K.H. Lee, J.H. Lee, S.J. Choi, J.J. Seo, H.N. Moon, H.J. Im, C.J. Park, Relationship between in vitro chemosensitivity assessed with MTT assay and clinical outcomes in 103 patients with acute leukemia, Korean J. Lab. Med. 27 (2007) 89–95.
[19] J.M. Norgaard, S.T. Langkjer, T. Palshof, B. Pedersen, P. Hokland, Pretreatment leukaemia cell drug resistance is correlated to clinical outcome in acute myeloid leukaemia, Eur. J. Haematol. 66 (2001) (2001) 160–167.
[20] T. Pemovska, M. Kontro, B. Yadav, H. Edgren, S. Eldfors, A. Szwajda, H. Almusa, M.M. Bespalov, P. Ellonen, E. Elonen, B.T. Gjertsen, R. Karjalainen, E. Kulesskiy, S. Lagstrom, A. Lehto, M. Lepisto, T. Lundan, M.M. Majumder, J.M. Marti, P. Mattila, A. Murumagi, S. Mustjoki, A. Palva, A. Parsons, T. Pirttinen, M.E. Ramet, M. Suvela, L. Turunen, I. Vastrik, M. Wolf, J. Knowles, T. Aittokallio, C.A. Heckman, K. Porkka, O. Kallioniemi, K. Wennerberg, Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia, Cancer Discov. 3 (2013) 1416–1429.
[21] W.E. Pierceall, S.M. Kornblau, N.E. Carlson, X. Huang, N. Blake, R. Lena, M. Elashoff, M. Konopleva, M.H. Cardone, M. Andreeff, BH3 profiling discriminates response to cytarabine-based treatment of acute myelogenous leukemia, Mol. Cancer Ther. 12 (2013) 2940–2949.
[22] P. Staib, E. Staltmeier, K. Neurohr, O. Cornely, M. Reiser, T. Schinkothe, Prediction of individual response to chemotherapy in patients with acute myeloid leukaemia using the chemosensitivity index Ci, Br. J. Haematol. 128 (2005) 783–791.
[23] S. Yamada, T. Hongo, S. Okada, C. Watanabe, Y. Fujii, T. Ohzeki, Clinical relevance of in vitro chemoresistance in childhood acute myeloid leukemia, Leukemia. 15 (2001) 1892–1897.
[24] Z.W. Li, W.S. Dalton, Tumor microenvironment and drug resistance in hematologic malignancies, Blood Rev. 20 (2006) 333–342.
[25] E.A. Sison, P. Brown, The bone marrow microenvironment and leukemia: biology and therapeutic targeting, Expert Rev. Hematol. 4 (2011) 271–283.
[26] Y. Tabe, M. Konopleva, Role of microenvironment in resistance to therapy in AML, Curr. Hematol. Malig. Rep. 10 (2015) 96–103.
[27] H. Zahreddine, K.L. Borden, Mechanisms and insights into drug resistance in cancer, Front. Pharmacol. 4 (2013) 28.

[28] A. Quartino, M.O. Karlsson, A. Freijs, N. Jonsson, P. Nygren, J. Kristensen, E. Lindhagen, R. Larsson, Modeling of in vitro drug activity and prediction of clinical outcome in acute myeloid leukemia, J. Clin. Pharmacol. 47 (2007) 1014–1021.

[29] J.E. Lancet, J.E. Cortes, D.E. Hogge, M.S. Tallman, T.J. Kovacsovics, L.E. Damon, R. Komrokji, S.R. Solomon, J.E. Kolitz, M. Cooper, A.M. Yeager, A.C. Louie, E.J. Feldman, Phase 2 trial of CPX-351, a fixed 5:1 molar ratio of cytarabine/daunorubicin, vs cytarabine/daunorubicin in older adults with untreated AML, Blood 123 (2014) 3239–3246.

[30] D.K. Webb, Management of relapsed acute myeloid leukaemia, Br. J. Haematol. 106 (1999) 851–859.