

Expression of a specific extracellular matrix signature is a favorable prognostic factor in acute myeloid leukemia

Valerio Izzi^{a,*}, Juho Lakkala^a, Raman Devarajan^a, Eeva-Riitta Savolainen^b, Pirjo Koistinen^c, Ritva Heljasvaara^{a,d}, Taina Pihlajaniemi^a

^a Centre of Excellence in Cell-Extracellular Matrix Research and Biocenter Oulu, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland

^b Nordlab Oulu and Institute of Diagnostics, Department of Clinical Chemistry, Oulu University Hospital, University of Oulu, Oulu, Finland

^c Medical Research Center Oulu, Institute of Clinical Medicine, Oulu University Hospital, University of Oulu, Oulu, Finland

^d Centre for Cancer Biomarkers (CCBIO), Department of Biomedicine, University of Bergen, N-5009 Bergen, Norway

ARTICLE INFO

Keywords:

Extracellular matrix (ECM)
Acute myeloid leukemia (AML)
Hematopoiesis
Clinical study
PCR

ABSTRACT

Relapse of acute myeloid leukemia (AML) is still dramatically frequent, imposing the need for early markers to quantify such risk. Recent evidence point to a prominent role for extracellular matrix (ECM) in AML, but its prognostic value has not yet been investigated. Here we have investigated whether the expression of a 15-ECM gene signature could be applied to clinical AML research evaluating a retrospective cohort of 61 AML patients and 12 healthy donors. Results show that patients whose ECM signature expression is at least twice as that of healthy donors have considerably longer relapse-free survival, with further stage-specific therapy outcomes.

1. Introduction

Although up to 80% of acute myeloid leukemia (AML) patients can expect to enter a first complete remission period (CR1) after appropriate induction regimen, many of them will subsequently relapse and face a dismal prognosis [1]. This adverse outcome is at the root of AML's still dramatically high death rate (approximately 21380 new AML cases will be diagnosed in USA in 2017, 50% of whom estimated to die within the same year- the highest death rate among hematological malignancies) [2], and the identification of new prognostic factors predisposing to either a better or a worse outcome is imperative to increase patients' chance to survive AML. Gene expression signatures have long since proven their potential usefulness in AML [3–6], but their translation to the clinics has been largely unsuccessful mainly because of the sophisticated methods they are based on (such as microarrays or specific chip platforms and RNA-seq) which are not readily available in clinical labs [7].

The extracellular matrix (ECM), the non-cellular microenvironment in which cells are embedded, plays crucial roles in both tissue homeostasis and disease [8]. In particular, in the hematopoietic niches, the ECM has key roles in anchoring hematopoietic stem cells (HSC) to the endosteal or the vascular structures, instructing the balance between proliferative and anti-proliferative signals and ultimately allowing for fine-tuning of the hematopoietic process throughout the life of the organism [8,9]. On the other hand, the ability of leukemia stem cells

(LSC) and AML cells to interact with the ECM is a detrimental feature which generally fosters resistance to therapy and survival of minimal leukemic clones, which relapse in time and re-install the disease [10]. It is the case, in example, of CD44, the prototypical hyaluronic acid receptor with the further ability to bind to other ECM components (such as osteopontin, fibronectin and selectins) [11]. It has been reported, in fact, that CD44 expression on LSC and AML cells associates with resistance to chemotherapy and increased aggressiveness of the disease [11]. Much alike, integrin-mediated sensing of fibronectin determines post-therapy survival of AML clones, thus ultimately facilitating its relapse [12].

While many evidence can be found in the literature about the ability of both normal and neoplastic hematopoietic clones to sense, and to bind to, ECM, there is conversely a dramatic scarcity of knowledge on the production of ECM by AML cells themselves, which also implies an almost complete lack of knowledge on what roles direct ECM regulation by AML cells play in the context of biological and clinical features of AML.

Recently, we and others have reported on common and widespread mechanisms controlling the expression of extracellular matrix genes in AML and leukemic precursors [5,13], and shown the prognostic value of what we called the “extracellular matrix signature of AML” [5]. Also, we showed that machine-learning algorithms such as support vector machine (SVM) can reduce the 80-genes signature to a more practicable 15-genes signature (which can be assessed by real-time quantitative

* Correspondence to: Center for Cell-Matrix Research, Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, P.O. Box 5000, FIN-90014 University of Oulu, Finland.
E-mail address: valerio.izzi@oulu.fi (V. Izzi).

PCR – RT-qPCR) without losing sensitivity [5], but did not test whether this reduced signature could be applied to define patients' prognosis.

Combining the need for a better understanding of ECM roles in AML with the necessity of having tests that can be performed in clinical laboratories without the need for sophisticated methods and high-end mathematics, we have here addressed the question whether the restricted set of ECM genes which we previously identified [5] could provide relevant clinical information on AML patients, and found that the expression of this ECM signature at levels twice as that of healthy donors marked patients with a better response to therapy, reduced minimal residual disease (MRD) and overall longer relapse-free survival. We also observe that these findings, obtained using the simplest techniques currently in use in hematological laboratories worldwide, can be largely recapitulated in previously-published AML cohorts investigated via microarrays, further suggesting the importance of this signature in the biology and clinical features of AML.

2. Material and methods

2.1. Analysis of the Oulu AML retrospective cohort

The Oulu AML retrospective cohort was assembled with approval of the Institutional Review Board and informed written consent of the patients, in accordance with the declaration of Helsinki. Details about the 73 patients studied (61 AML + 12 healthy controls), as well as about the composition of the reduced ECM signature and the primers used for RT-qPCR are reported in the Appendix. The expression values of the 15 genes constituting the ECM signature (normalized to *GAPDH*) were collapsed to a single value per AML patient or healthy donor by calculating their geometric mean, using the formula:

$$\left(\prod_{i=1}^n a_i \right)^{\frac{1}{n}} = \sqrt[n]{a_1 a_2 \dots a_n}$$

in which the geometrical mean is defined as the n^{th} root of the product of n elements a (n being the number of elements, in this case the genes - a). The arithmetic mean of all geometric mean values from the healthy donors was then calculated and the standard deviation value multiplied by 2 and then added to the average to obtain the upper and lower cutoff thresholds. All AML patients whose gene expression (geometric mean) fell within the thresholds were allocated to the ECM^{norm} group, while those whose expression was higher than the upper 2-SD threshold were allocated into the ECM^{high} group. In the Oulu cohort there were also 3 AML patients whose expression was lower than the bottom 2-SD threshold. Upon analysis, we found that these patients had no difference with the ECM^{norm} group, while showed exactly the same differences that the ECM^{norm} exhibited in respect to the ECM^{high} group. Hence, these patients were allocated back into the ECM^{norm} group.

For the analysis of outcome (post therapy)-specific results, patients were assessed at the following time-points: end of the induction protocol (Ind1), end of the first consolidation protocol (Cons1), and last available follow-up (Last).

2.2. Analysis of ECM signature expression in hematopoietic precursors

Raw microarray data (Affymetrix Human Genome U133 Plus 2.0 Array) were downloaded for the samples reported by Gentles et al. (GSE24006) [3] and by Novershtern et al., (GSE24759) [14], imported into Chipster (<http://chipster.csc.fi/>), normalized using robust multi-array average (RMA) protocol and the expression of the ECM signature studied. To facilitate cross-comparison with GSE24006, data from the GSE24759 were subset (post-normalization) to remove more mature cells, finally including only hematopoietic precursors (CD133⁺ and CD34⁺ HSC), committed progenitors (CMP, GMP and MEP), single-colony forming unit (CFU) progenitors (monocytic, granulocytic and megakaryocytic), and naïve B and T lymphocytes

2.3. Statistics

Fisher's Exact test (2-sided), Mann-Whitney U test, Analysis of Variance (ANOVA) followed by Tukey's HSD or Dunnett's T3 post-hoc tests, Kaplan-Meier (Log-Rank method, KM) and Cox proportional hazards (Cox-PH) survival analyses were performed in IBM SPSS Statistics 21, and all tests were bootstrapped 1000 times unless otherwise specified. Gene network enrichment analysis was performed in *String-DB* (<http://string-db.org/>) and the results imported into Cytoscape for easier visualization. The Linear Support Vector Machine (LSVM) algorithm used to analyze the contribution of the ECM gene expression to prognosis was trained and tested as reported in the Appendix, using IBM SPSS Modeler 18. In all analyses, a value of $P < 0.05$ was considered significant.

3. Results

3.1. Features of the ECM signature

The ECM signature which we tested in this work was previously reported [5] and comprises the following genes: *ADAM17*, *COL24A1*, *EMILIN2*, *CHI3L1*, *COL17A1*, *COL18A1*, *CRISP3*, *CRISPLD2*, *DEFA1*, *ELANE*, *LGALS3*, *MMP8*, *MMP9*, *PRTN3* and *SLPI*. This specific ECM signature is significantly enriched for protein-protein interactions (PPI) and includes ECM regulators (proteinases, 45% of the total gene-set), collagens (27%), glycoproteins (18%) and ECM-affiliated proteins (9%) (Fig. 1A) and overlaps with human AML signatures and mouse models of immunological and hematological phenotypes, which is an indication of the specific involvement of its constituents in the development (either normal or neoplastic) and functions of white blood cells (Fig. 1B,C and Appendix Table 1). Further ontological analyses of the signature are reported in Appendix Table 1.

Notably, signature expression is overall low in early hematopoietic stem and progenitor cells (CD133⁺ and CD34⁺ hematopoietic stem cells -HSC- and multipotent precursors -MPP), while it significantly increased with differentiation along the erythro-myeloid branch (myelo-erythroid progenitors -MEP-, common myeloid progenitors -CMP-, and granulocyte-monocyte progenitors -GMP-) and reached its maximum at the monocytic stage (CFU-mono) (Appendix Fig. 1A,B). In a similar way, the expression of the ECM signature in neoplastic clones was at its lowest in leukemia stem cells (LSC), while it increased constantly with more-differentiated cell states (leukemia precursor cells -LPC- and AML blasts) (Appendix Fig. 1B). Altogether, these results indicate that acquisition of this signature is globally associated with a more mature phenotype and, accordingly, we observed a significant negative association between signature expression and mRNA levels for CD34, a typical HSC and LSC marker [15], and a positive association with CD14, the phenotyping marker of monocytes [16].

3.2. Clinical significance of the ECM signature

Since this signature includes genes both up-and down-regulated in respect to healthy donors (Appendix Fig. 2) [5], and since relative expression values could not be collapsed into a single "global" value without using complicate approaches (such as principal component analysis) [3,6] unsuitable for direct clinical use, we undertook a

different approach, which separated AML patients into those who expressed the signature more than 2-times standard deviation (2-SD) that of healthy donors' expression and those whose expression was less than 2-SD that of healthy donors (see Supplemental Material for further details). All AML patients within the 2-SD limit were considered as "normal-like ECM" (ECM^{norm}), while patients outside these borders were considered significant outliers. Interestingly, we could not detect AML patients below the lower 2-SD threshold, but we could identify patients above the highest 2-SD thresholds, which we termed ECM^{high}. We found that ECM^{high} patients (in total 24 out of the 61 patients) had

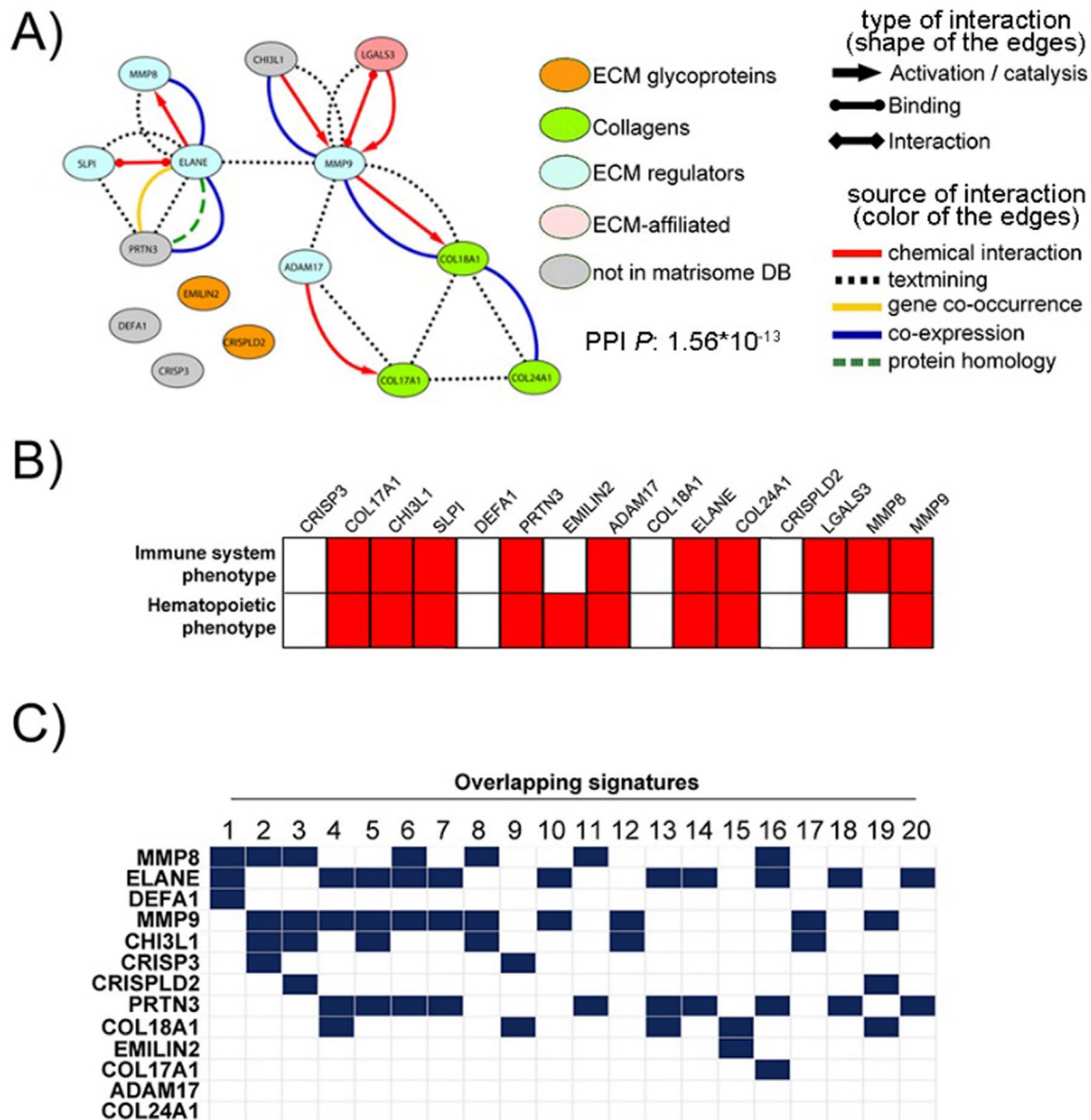


Fig. 1. Features of the ECM signature. The 15-ECM gene signature (A) is significantly enriched for interacting proteins, spanning different categories in the Matrisome Database and types of interaction. This signature also (B) hosts a significant amount of genes previously reported to produce altered hematopoietic or immune system phenotypes when altered in mice, and (C) overlaps significantly with signatures of neoplastic hematopoiesis (reported in Appendix Table 1). PPI: protein-protein interaction value. Data in (A) are from *String DB* (<https://string-db.org/>), in (B) from *MGI* (<http://www.informatics.jax.org/>) and in (C) from *MSigDB* (<http://software.broadinstitute.org/gsea/msigdb/index.jsp>).

significantly longer relapse-free survival (RFS) in respect to ECM^{norm} patients, both in KM and Cox-PH models (Fig. 2A). Particularly, in Cox-PH, ECM^{high} patients' hazard was 0.381 (95% confidence interval: 0.15–0.97, Table 1), indicating an approximate 69% reduction in the risk of an unfavorable event.

Table 1. The ECM patients' groups (ECM^{high} or ECM^{norm}), in red, were inputted together with gender, age, and molecular and cytogenetic abnormalities into a multivariable (Cox proportional hazards) model of relapse-free survival. Df: degrees of freedom; HR: hazard rate; 95% CI: 95% confidence interval.

Notably, the ECM^{high} and ECM^{norm} groups did not differ in overall survival (OS, Appendix Table 2), nor did they show association with gender, age, cytogenetic or molecular abnormalities (Appendix Table 3), suggesting a specific involvement of the ECM signature in the mechanisms underlying patients' chemosensitivity. Further analyses evidenced that the ECM^{high} group had lesser relapse event overall (41% vs. 80% in the ECM^{norm} group) (Fig. 2B) and significantly different

outcomes at different steps of the therapy. We observed, in fact, similar response to therapy (% of patient attaining CR) after the first induction cycle, followed by a steady increase at later stages in the ECM^{high} group. Conversely, in the ECM^{norm} group, the good response at consolidation was followed by a sharp decrease at the last follow-up (Fig. 2C), a clear indication of the rise of relapses during the post-first consolidation stages (coinciding with discharge from hospital and follow-up periods) in the ECM^{norm} group. Notably, the increase in CR in the ECM^{high} group over time was linear (P:0.0124), indicating a trend towards gradual amelioration over time in this group.

These data are also in agreement with the % of patients having minimal residual disease, MRD (< 5% detectable AML blasts in the blood) [1], at the same cycles: while, in fact, both the groups showed similar levels after the first induction and consolidation, the % of MRD patients decreased significantly at last follow-up in the ECM^{high} group only, indicating a favorable resolution of the disease (Fig. 2D). Notably, as already observed for microarray data, *CD34* and *CD14* mRNA levels

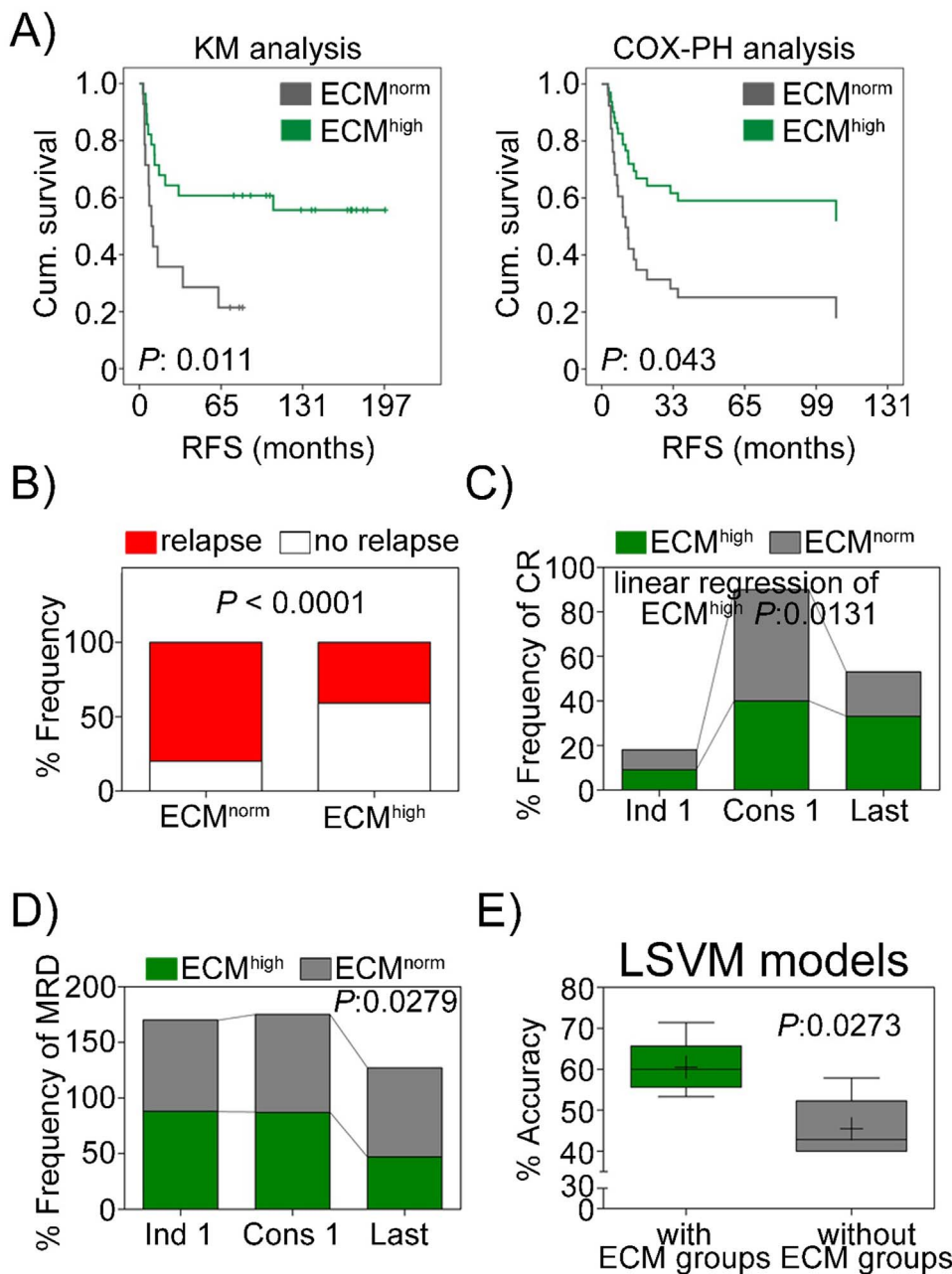


Fig. 2. High ECM gene expression marks favorable outcome in AML. Patients with high expression of the ECM gene-set in respect to healthy donors (ECM^{high} , > 2 times the standard deviation of the healthy donors) had significantly longer relapse-free survival (RFS) than patients with ECM gene-set expression comparable to the healthy donors (ECM^{norm}) in both univariable and multivariable analyses (A). ECM^{high} patients had also quantitatively less relapses overall (B), and exhibited higher complete remission (CR, C) and lower minimal residual disease (MRD, D) frequencies at last follow-up. (E) Incorporating the ECM gene-set information into a linear support vector machine (LSVM) classifier increases the accuracy of a model based on age, gender, molecular and cytogenetic abnormalities. *P* values are from (A) Log-rank and Cox proportional hazards, (B,D) Fisher's Exact, (C) linear regression, and (E) Mann-Whitney *U* test.

Table 1
Multivariable (Cox-PH) relapse-free survival (RFS) analysis of the Oulu retrospective cohort.

Omnibus test of model coefficients								
–2 Log Likelihood	Chi-square	df	<i>P</i> value					
126.2604801	11.91467	5	0.035976					
Variables in the equation								
	B	SE	Wald	df	<i>P</i> value	HR	95% CI for HR	
							Lower	Upper
ECM gene-set	–0.96469	0.476416	4.100165	1	0.042879	0.381102	0.149801	0.969543
Gender	–0.11388	0.472162	0.058167	1	0.809417	0.892369	0.353704	2.251386
Age	–0.00774	0.015038	0.265134	1	0.606615	0.992287	0.963466	1.021969
Molecular abn.	1.133062	0.762361	2.208953	1	0.137212	3.10515	0.696883	13.83583
Cytogenetic abn.	0.768643	0.458744	2.807422	1	0.093829	2.156836	0.877674	5.300306

in the Oulu cohort regressed negatively and positively, respectively, with that of the ECM signature (Appendix Figure 3), further suggesting that acquisition of the ECM signature is a sign of cell maturation. Furthermore, data show that the overall accuracy of different

automated algorithms (including linear support vector machine - LSVM, k-nearest neighbors - KNN, and naïve Bayes network -NBN) in predicting patients' relapse was largely improved if the ECM signature status was added to the age, gender, cytogenetic and molecular

information about the patients (Fig. 2E), further supporting the potential relevance of the ECM signature expression in driving clinical decisions.

Finally, since the signature wraps all expression data into a single value, we further investigated on the different expression of each gene in the groups as described in the Supplemental Material, and found that only 3 genes were differentially expressed (up-regulated) in ECM^{high} patients vs. both ECM^{norm} and healthy donors (*COL24A1*, *ELANE* and *MMP9*) (Appendix Figure 4 and Appendix Table 4), suggestive of their central role in establishing the ECM^{high} phenotype.

4. Discussion

Our study shows, for the first time, the direct prognostic value of a specific set of ECM genes' expression in predicting relapse-free survival in adult AML. Furthermore, our results come from a context (the 2-SD cutoff in respect to healthy donors) and a methodology (the RT-qPCR) that suits clinical hematology laboratories, thus directly translating our previous biomarker discovery work [5] into practice.

Owing to the scarcity of data about ECM and AML, it is difficult to discuss the individual roles of the ECM genes in the specific signature. It seems, nevertheless, notable that two of the three up-regulated genes characterizing the ECM^{high} phenotype have been already implicated in AML: *MMP9*, in fact, has been already recently described as generally down-regulated in AML [13], and its higher expression postulated to be a positive factor in patients' prognosis [17], thus completely matching our observation. Intriguingly, we also observe significant up-regulation of *ELANE*, the neutrophils' elastase, in these patients. *ELANE* has not only been reported to directly interact with *MMP9* [18], but also to be linked to a higher risk of neutropenic patients to develop AML when its content is lower than normal [19]. Notably, *COL24A1* has also been recently found to have a prognostic value in cancer [20], though this is the first time it is associated to AML.

It is, furthermore, important to notice the inverse correlation between the ECM signature and *CD34*, which is a *bona fide* marker of LSC [15]. It has been already reported, in fact, that ECM gene expression is generally down-regulated in AML [5,13], and so it is conceivable that higher ECM associates with a more differentiated phenotype. Further sustain to this hypothesis comes from the observed down-regulation of *COL18A1*, which has been conversely associated with normal hematopoietic precursors in both mice and humans [21], and the fact that its down-regulation might trigger proliferation of myeloid clones [22].

In conclusion, the correlation of the ECM signature with AML outcome and survival suggests once more a crucial role for specific ECM regulation in AML biology and encourages further studies into the translation of these knowledge into the clinical practice.

Acknowledgments

The authors would like to Jaana Peters and Kirsi Kvist-Mäkelä for technical assistance and the CSC - IT Center for Science Ltd. for software development and technical support.

Funding details

This work was supported by the Centre of Excellence Grant 2012–2017 of the Academy of Finland (284605), the Sigrid Jusélius Foundation, the Finnish Cancer Foundation; and the University of Oulu Graduate School.

Disclosure of interest

The authors report no conflicts of interest

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.lrr.2017.12.001>.

References

- [1] H. Dohner, E. Estey, D. Grimwade, S. Amadori, F.R. Appelbaum, T. Buchner, et al., Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel, *Blood* 129 (2017) 424–447.
- [2] Adult Acute Myeloid Leukemia Treatment (PDQ®)–Health Professional Version, 2017.
- [3] A.J. Gentles, S.K. Plevritis, R. Majeti, A.A. Alizadeh, Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia, *JAMA* 304 (2010) 2706–2715.
- [4] H. Hackl, K. Steinleitner, K. Lind, S. Hofer, N. Tomic, S. Pavlovic, et al., A gene expression profile associated with relapse of cytogenetically normal acute myeloid leukemia is enriched for leukemia stem cell genes, *Leuk. Lymphoma* 56 (2015) 1126–1128.
- [5] V. Izzi, J. Lakkala, R. Devarajan, H. Ruotsalainen, E.R. Savolainen, P. Koistinen, et al., An extracellular matrix signature in leukemia precursor cells and acute myeloid leukemia, *Haematologica* (2017).
- [6] S. Wilop, W.C. Chou, E. Jost, M. Crysandt, J. Panse, M.K. Chuang, et al., A three-gene expression-based risk score can refine the European LeukemiaNet AML classification, *J. Hematol. Oncol.* 9 (78) (2016) 016–0308–8.
- [7] F. Chibon, Cancer gene expression signatures—the rise and fall? *Eur. J. Cancer* 49 (2013) 2000–2009.
- [8] F. Gattazzo, A. Urciuolo, P. Bonaldo, Extracellular matrix: a dynamic micro-environment for stem cell niche, *Biochim Biophys. Acta* 1840 (2014) 2506–2519.
- [9] A. Nakamura-Ishizu, Y. Okuno, Y. Omatsu, K. Okabe, J. Morimoto, T. Uede, et al., Extracellular matrix protein tenascin-C is required in the bone marrow micro-environment primed for hematopoietic regeneration, *Blood* 119 (2012) 5429–5437.
- [10] O. Bruserud, E. Aasebo, M. Hernandez-Valladares, G. Tsykunova, H. Reikvam, Therapeutic targeting of leukemic stem cells in acute myeloid leukemia—the biological background for possible strategies, *Expert Opin. Drug Discov.* 12 (2017) 1053–1065.
- [11] M. Zoller, CD44, hyaluronan, the hematopoietic stem cell, and leukemia-initiating cells, *Front Immunol.* 6 (2015) 235.
- [12] T. Matsunaga, N. Takemoto, T. Sato, R. Takimoto, I. Tanaka, A. Fujimi, et al., Interaction between leukemic-cell VLA-4 and stromal fibronectin is a decisive factor for minimal residual disease of acute myelogenous leukemia, *Nat. Med* 9 (2003) 1158–1165.
- [13] A. Foroushani, R. Aghahari, R. Docking, L. Chang, G. Duns, M. Hudoba, et al., Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications, *BMC Med Genom.* 10 (16) (2017) 017–0253–6.
- [14] N. Novershtern, A. Subramanian, L.N. Lawton, R.H. Mak, W.N. Haining, M.E. McConkey, et al., Densely interconnected transcriptional circuits control cell states in human hematopoiesis, *Cell* 144 (2011) 296–309.
- [15] G.J. Schuurhuis, M.H. Meel, F. Wouters, L.A. Min, M. Terwijn, N.A. de Jonge, et al., Normal hematopoietic stem cells within the AML bone marrow have a distinct and higher ALDH activity level than co-existing leukemic stem cells, *PLoS One* 8 (2013) e78897.
- [16] P. Antal-Szalmas, J.A. Strijp, A.J. Weersink, J. Verhoef, K.P. Van Kessel, Quantitation of surface CD14 on human monocytes and neutrophils, *J. Leukoc. Biol.* 61 (1997) 721–728.
- [17] J. Schmohl, D. Santovito, T. Guenther, W. Sutanto, T. Kroell, H. Salih, et al., Expression of surface-associated 82kDa-proMMP-9 in primary acute leukemia blast cells inversely correlates with patients' risk, *Exp. Hematol.* 44 (358) (2016) (362.e5).
- [18] Y. Guo, L. Ma, F. Zhang, R. Sun, T. Li, Neutrophil elastase ameliorates matrix metalloproteinase-9 to promote lipopolysaccharide-induced acute lung injury in mice 1, *Acta Cir. Bras.* 31 (2016) 382–388.
- [19] V. Makaryan, C. Zeidler, A.A. Bolyard, J. Skokowa, E. Rodger, M.L. Kelley, et al., The diversity of mutations and clinical outcomes for *ELANE*-associated neutropenia, *Curr. Opin. Hematol.* 22 (2015) 3–11.
- [20] K. Misawa, T. Kanazawa, A. Imai, S. Endo, D. Mochizuki, H. Fukushima, et al., Prognostic value of type XXII and XXIV collagen mRNA expression in head and neck cancer patients, *Mol. Clin. Oncol.* 2 (2014) 285–291.
- [21] N.B. Ivanova, J.T. Dimos, C. Schaniel, J.A. Hackney, K.A. Moore, I.R. Lemischka, A stem cell molecular signature, *Science* 298 (2002) 601–604.
- [22] V. Izzi, D.G. Vicente, R. Devarajan, H. Ruotsalainen, R. Heljasvaara, T. Pihlajaniemi, The Frizzled-like domain from collagen XVIII (FZC18) controls hematopoiesis and myeloidogenesis in vivo, *Immunology* 137 (Supplement 1) (2012) (233-233).