An extracellular matrix signature in leukemia precursor cells and acute myeloid leukemia

Despite major therapeutic advancements, worldwide death rates of acute myeloid leukemia (AML) remain high, with approximately 20,830 people diagnosed in the USA in 2015, 10,460 (50.21%) of whom were estimated to die from the disease.[1] Several studies have shown that leukemia stem cells (LSCs), the founder cells from which AML arise,[2] are characterized by specific transcriptional[3] and epigenetic[4] profiles which can be applied to predict patient survival and prognosis.[3,4] The actual model for AML development[5] postulates that LSC arise within the same niches as the normal hematopoietic stem cells (HSCs), taking them over in time as the hematopoietic niche turns into a leukemic niche. While the altered expression of different extracellular matrix (ECM) elements within the leukemic niche has already been investigated,[6] the direct contribution of LSCs to the modification of the niche ECM has not been assessed systematically, and the prognostic relevance of alterations to the ECM homeostasis directly operated by LSCs, and AML cells remains untested. To this aim, we studied the transcriptional profile of ECM-related genes in LSCs, and applied the results to two AML cohorts to verify their prognostic potential.

The raw microarray profiles of normal HSCs, multipotent progenitors (MPPs), committed progenitors (megakaryocyte-erythrocyte progenitors, MEPs, common myeloid progenitors, CMPs, and granulocyte/monocyte progenitors, GMPs), LSCs, leukemia progenitor cells

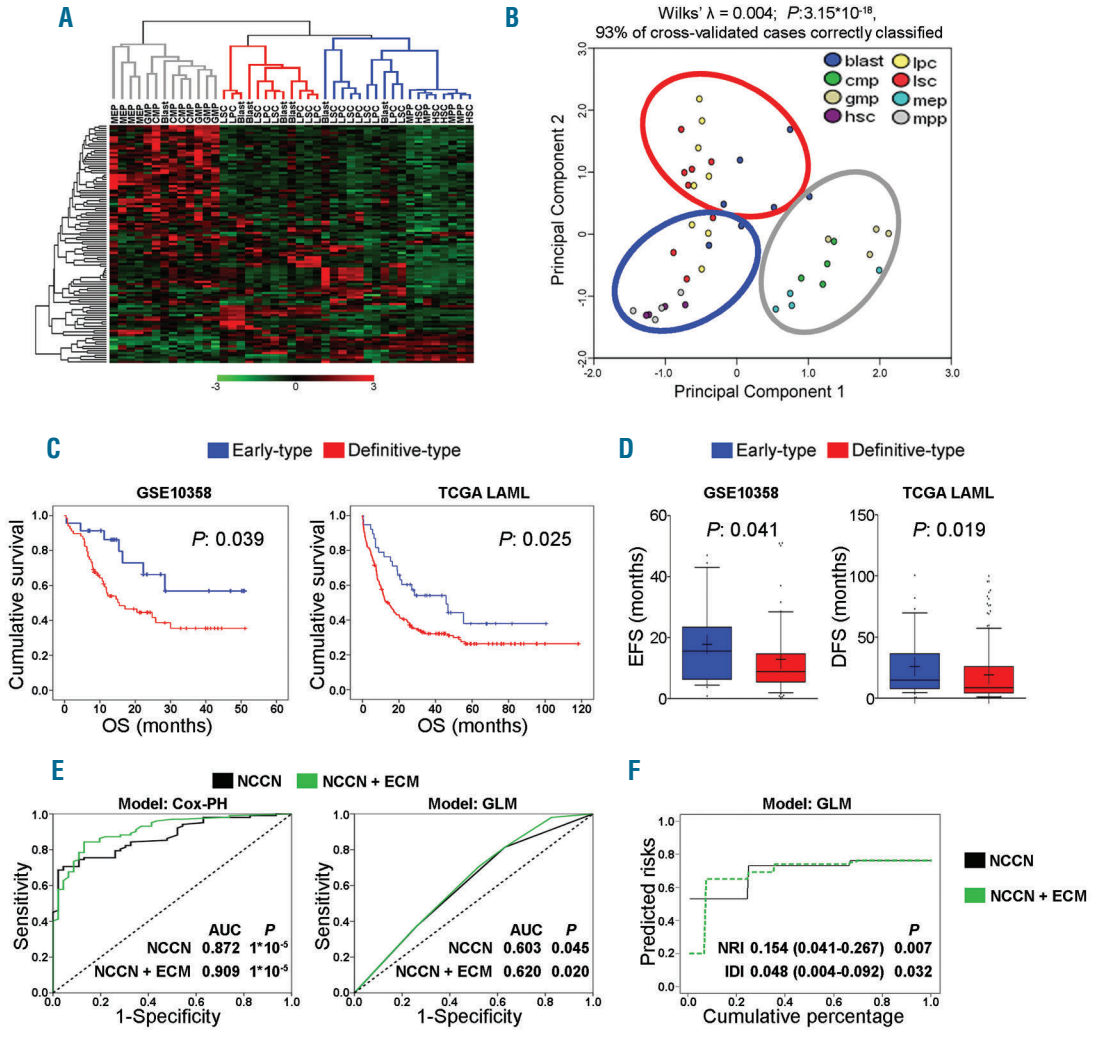

Figure 1. The prognostic value of the ECM signature. (a-b) ECM signature expression clusters leukemic and normal precursors into three groups, according to a) hierarchical clustering with Ward's method and b) PCA followed by LDA. (c, d) Analysis of c) the overall survival (OS) of the GSE10358 and TCGA LAML cohorts and d) mean event-free survival (EFS) for GSE10358 and mean disease-free survival (DFS) for TCGA LAML, using the early and definitive leukemic types identified by the ECM signature. *P*-values are from Log-Rank test in (c) and from Mann-Whitney U test in (d). In (d), data are reported as 10-90 percentile with outliers, median (thin internal line), mean (thin internal cross) and standard deviation. (e-f) A risk classification scheme adding the ECM signature to the NCCN classifier outperforms the classifier alone. (e) AUC-ROC analysis of Cox proportional hazard (Cox-PH) and generalized linear models (GLM). (f) Predictive curve, NRI and IDI from GLM.

(LPCs), and blasts were retrieved from the NCBI Gene Expression Omnibus (GEO) repository (*http://www.ncbi.nlm.nih.gov/geo/*) through the GEO Series accession number GSE24006. AML patient microarray and clinical data were retrieved from GEO for the accession number GSE10358 or from the GDC Legacy archive (*http://gdc-portal.nci.nih.gov/legacy-archive*) for The Cancer Genome Atlas AML cohort (TCGA LAML). Raw data from healthy donors (GSE11504 and GSE13159) were used as controls for AML. The raw intensity expression values were processed by Robust Multi-array Average procedure in Chipster software (*http://chipster.csc.fi/*) and significantly under- and over-expressed genes in leukemic cells (precursors and cohorts) tested with Empirical Bayes test with Benjamini-Hochberg correction. Only the genes which significantly varied in AML precursors *vs.* normal precursors and in AML cohorts *vs.* healthy donors were further studied. The list of ECM genes on which we focused was compiled by merging the genes in the Gene Ontology (GO) categories detailed in the *Online Supplementary Information*. Data standardization, hierarchical clustering (Ward's method), principal component analysis (PCA), Linear Discriminant Analysis (LDA), Fisher's Exact test (2-sided), Kaplan-Meier (Log-Rank), Cox multivariate models for survival analysis (Cox-PH) and generalized linear models (GLM) were performed in IBM SPSS Statistics 21. Net reclassification improvement (NRI), integrated discrimination improvement (IDI) and the area under the receiver operating characteristic curve (AUC-ROC) were calculated in R. A value of $P<0.05$ was considered significant. The support Vector Machine (SVM) algorithm used to select the 15 most important genes among the previously-identified ones was trained and tested as reported in the *Online Supplementary Information*. The retrospective Oulu AML cohort used to assess gene expression in patients was assembled with

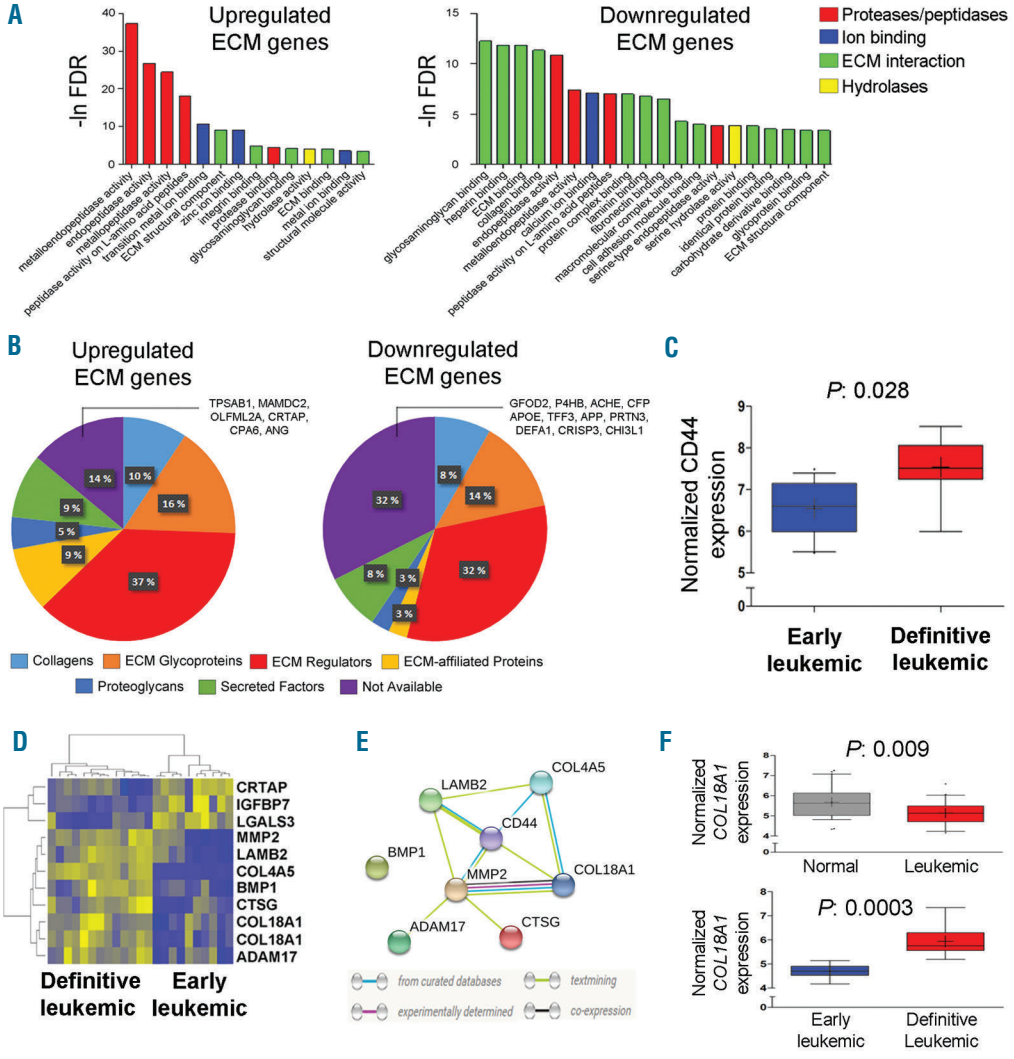

Figure 2. Components of the ECM signature and the CD44-ECM subnetwork. (a-b) Enrichment of the up- and downregulated genes of the ECM signature according to a) gene ontology (GO) categories and conceptual meta-categories, and b) categories from the Matrisome database. In a) data are presented as the natural antilogarithm of the false discovery rate (-ln FDR). (c) Normalized expression of *CD44* in early and definitive leukemic cells. (d-e) Differentially expressed genes in early and definitive leukemic cells (d) and network analysis of genes upregulated in definitive leukemic cells (e). (f) Normalized expression of *COL18A1* in leukemic cells *vs.* normal cells (upper panel) and in early *vs.* definitive leukemic cells (lower panel). (c, f) Data are reported as 10-90 percentile with outliers, median (thin internal line), mean (thin internal cross) and standard deviation. *P*-values are from Mann-Whitney U test.

approval of the Institutional Review Board and informed written consent of the patients, in accordance with the declaration of Helsinki. Patient samples were tested for the expression of each of the 15 genes by quantitative PCR, using the primers reported in the *Online Supplementary Information*. Gene network enrichment was performed in String (*http://string-db.org/*).

In total, 80 ECM genes were found differentially expressed in leukemic *vs.* normal precursor cells as well as in the two AML cohorts *vs.* healthy donors (*Online Supplementary Table S1*). Of them, the differential expression of the 15 most important genes, as resulted from the SVM algorithm, was validated via qPCR in the 65 AML patients of the Oulu retrospective cohort (*Online Supplementary Table S2* and *Online Supplementary Figure S1*). Grouping the precursors according to the 80 ECM genes by Ward's method or PCA followed by LDA resulted in three significant clusters; one containing only committed precursors (in grey), one containing only leukemic cells (in red), and the last containing leukemic cells admixed with normal HSCs and MPPs (in blue). Based on the proposed origin of LSCs from MPPs,[4] we called the leukemic cells within the blue cluster "early leukemic", and the ones within the red cluster "definitive leukemic" (Figure 1A–B). Notably, the amount of LSC and LPC in the two subgroups was equal (*Online Supplementary Figure S2*), discarding possible quantitative interferences as confounding factors for the clustering procedures. Next, we standardized the leukemic precursors and the patient data together and observed that the samples from the early and definitive groups continued to cluster independently (albeit fragmenting in smaller subgroups), never mixing and partitioning patient data into the same groups (*Online Supplemental Figure S3*). Patient grouping into "early-type" and "definitive-type" resulted in significant differences in overall survival (Figure 1C), even when data were adjusted for karyotypical (P=0.038 in GSE10358 and 0.016 in TCGA LAML) or molecular abnormalities (adjustment for FLT3, P=0.031 in GSE10358 and 0.027 in TCGA LAML; adjustment for IDH1, P=0.041 in GSE10358 and 0.033 in TCGA LAML; adjustment for NPM1, P=0.033 in GSE10358 and 0.024 in TCGA LAML). The two groups also significantly differed in event-free survival (EFS) and disease-free survival (DFS) (Figure 1D). Also, Fisher's Exact test returned no significant associations between clustering into the two groups and any risk factor (*Online Supplementary Table S3*), and multivariate survival models returned significant values for DFS in TCGA (P=0.015, HR: 4.45 [95% CI: 1.33-14.8]) and for EFS in GSE10358 (P=0.017, HR: 5.1 [95% CI: 1.34-20]), further confirming the independent prognostic value of these observations. Finally, we noticed that including the ECM signature into Cox-PH or GLM models based on the National Comprehensive Cancer Network (NCCN guidelines) significantly increased the discriminatory (Figure 1E) and the classificatory (Figure 1F) power of the NCCN classifier. Concerning the functions of the differentially expressed ECM genes, we found that leukemic cells upregulate proteases/peptidases (mostly metalloproteinases), but downregulate cell-ECM binding molecules (Figure 2A). Coherently, using the ontologies suggested by Naba *et al.*,[7] we observed that the group of genes with the highest variance was that of the "ECM regulators" (to which proteolytic enzymes are large contributors), followed by the "Not available" group (genes not included in the Matrisome DB) and by ECM glycoproteins (mostly proteins involved in cell adhesion). The remaining groups of collagens, secreted factors, proteoglycans and ECM-affiliated proteins were only marginally impacted (Figure 2B). The analysis of LSC markers[8] showed no major differences in the early and definitive groups (*Online Supplementary Figure S4*) except for the hyaluronic acid receptor CD44, which was significantly upregulated in definitive leukemic cells (Figure 2C). Notably, while CD44 has no effect on patient survival *per se* (P=0.513 for OS and P=0.982 for DFS in TCGA LAML), the same receptor has been reported to interact with collagen IV and laminin β2,[9,10] and we found both *COL4A5* and *LAMB2* among the seven genes significantly upregulated in the definitive leukemic cells (Figure 2D). Coinciding with overall ontological data, this small network is significantly enriched for ECM remodeling and peptidase/metallopeptidase (Figure 2E and *Online Supplementary Table S4*). Furthermore, this network also contains collagen XVIII (*COL18A1*), which has previously been associated with human and murine HSCs.[11] Notably, the absolute levels of *COL18A1* are lower in leukemic cells than in their normal counterparts (Figure 2F), but they are at their local highest in definitive leukemic cells, again remarking the peculiarities of this stage of leukemic development.

In this study, we show for the first time the existence of an "ECM signature" which is shared by leukemia precursor cells and circulating AML cells from patients. The most striking feature of the ECM signature was to partition leukemic precursors into two groups, which differed for a restricted set of ECM genes and for the expression of the CD44 receptor. CD44 belongs to a family of transmembrane glycoproteins whose primary function is to bind hyaluronic acid (HA), laminins, collagens, matrix metalloproteinases (MMPs), osteopontin, *etc.*,[12] and this receptor has previously been implicated in cell migration, proliferation, differentiation, survival, and bone marrow homing of hematopoietic stem/progenitor cells as well as in the homing of LSCs to intra-and extra-medullary niches and in resistance to chemotherapy.[12] We observed that the leukemic precursors with a higher expression of CD44 (the group of cells we called "definitive leukemic cells") also exhibit a parallel upregulation of genes whose products interact directly (*COL4A5, LAMB2*)[9,10] or indirectly (*MMP2, COL18A1*)[12] with CD44 and downregulation of MMP9 (which directly interacts with CD44 but whose levels correlate inversely with patient prognosis),[13] suggesting that the establishment of a "CD44-ECM network", rather than the expression of CD44 alone, is a crucial step in the progression of leukemic cells towards an aggressive phenotype. This also seems to be supported by the observation that, in two independent cohorts, patients with an ECM profile similar to that of the definitive leukemic cells showed significantly shorter survival (overall and endpoint), independently from well-known karyotypical or molecular drivers of AML. Notably, other genes upregulated in definitive leukemic cells include the matrix metalloproteinase 2 (*MMP2*), a disintegrin and metalloprotease domain 17 (*ADAM17*) and cathepsin G (*CTSG*), constituting a proteolytic subnetwork that sits well with the overall upregulation of proteases which we observe in the ECM signature and that others have already reported in AML.[14] Also, *MMP2* has been implicated in AML invasiveness,[15] and ADAM17 seems to play a central role in the survival of leukemic cells via the activation of the Lyn/Akt survival pathway.[16] In conclusion, the correlation of the ECM signature with AML outcome and leukemic precursor subtypes suggests a central role for ECM alteration in AML biology and encourages further studies to understand the regulatory mechanisms controlling it.

Valerio Izzi,[1] Juho Lakkala,[1] Raman Devarajan,[1] Heli Ruotsalainen,[1] Eeva-Riitta Savolainen,[2,3] Pirjo Koistinen,[3] Ritva Heljasvaara[1,4] and Taina Pihlajaniemi[1]

[1]Centre of Excellence in Cell-Extracellular Matrix Research and Biocenter Oulu, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland; [2]Nordlab Oulu and Institute of Diagnostics, Department of Clinical Chemistry, Oulu University Hospital, Finland; [3]Medical Research Center Oulu, Institute of Clinical Medicine, Oulu University Hospital, Finland and [4]Centre for Cancer Biomarkers (CCBIO), Department of Biomedicine, University of Bergen, Norway

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: 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.

Correspondence: valerio.izzi@oulu.fi
doi:10.3324/haematol.2017.167304

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Adult acute myeloid leukemia treatment - for health professionals (PDQ®), http://www.cancer.gov/types/leukemia/hp/adult-aml-treatment-pdq, 2015 (accessed: 6.12.2016).
2. Jordan CT. The leukemic stem cell. Best Pract Res Clin Haematol. 2007; 20(1):13-18.
3. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. JAMA. 2010;304(24):2706-2715.
4. Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. Nat Commun. 2015;6:8489.
5. Hope KJ, Jin L, Dick JE. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. Nat Immunol. 2004; 5(7): 738-743.
6. Boulais PE, Frenette PS. Making sense of hematopoietic stem cell niches Blood. 2015;125(17):2621-2629.
7. Naba A, Clauser KR, Hoersch S, Liu S, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics 2012;11(4):M111.014647.
8. Zhou J, Chng WJ. Identification and targeting leukemia stem cells: The path to the cure for acute myeloid leukemia. World J Stem Cells. 2014;6(4):473-484.
9. Ishii S, Ford R, Thomas P, Nachman A, Steele G Jr, Jessup JM. CD44 participates in the adhesion of human colorectal carcinoma cells to laminin and type IV collagen. Surg Oncol. 1993;2(4):255-264.
10. Klingbeil P, Marhaba R, Jung T, Kirmse R, Ludwig T, Zöller M. CD44 variant isoforms promote metastasis formation by a tumor cell-matrix cross-talk that supports adhesion and apoptosis resistance. Mol Cancer Res. 2009;7(2):168-179.
11. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. A stem cell molecular signature. Science. 2002;298(5593):601-604.
12. Ponta H, Sherman L, Herrlich PA. CD44: from adhesion molecules to signalling regulators. Nat Rev Mol Cell Biol. 2003;4(1):33-45.
13. Schmohl J, Santovito D, Guenther T, et al. Expression of surface-associated 82kDa-proMMP-9 in primary acute leukemia blast cells inversely correlates with patients' risk. Exp Hematol. 2016;44(5):358-362.e5.
14. Hatfield KJ, Reikvam H, Bruserud Ø. The crosstalk between the matrix metalloprotease system and the chemokine network in acute myeloid leukemia. Curr Med Chem. 2010;17(36):4448-4461.
15. Takahashi K., Eto H, Tanabe KK. Involvement of CD44 in matrix metalloproteinase-2 regulation in human melanoma cells. Int J Cancer. 1999;80(3):387-395.
16. Liu WH, Chang LS. Suppression of ADAM17- mediated Lyn/Akt pathways induces apoptosis of human leukemia U937 cells: Bungarus multicinctus protease inhibitor-like protein-1 uncovers the cytotoxic mechanism. J Biol Chem. 2010;285(40):30506-30515.