Pediatric RESEARCH

www.nature.com/pr

BASIC SCIENCE ARTICLE

Deciphering molecular heterogeneity in pediatric AML using a cancer vs. normal transcriptomic approach

Barbara Depreter[1], Barbara De Moerloose[1,2,3], Karl Vandepoele[2,4], Anne Uyttebroeck[5], An Van Damme[6], Eva Terras[3], Barbara Denys[4], Laurence Dedeken[7], Marie-Françoise Dresse[8], Jutte Van der Werff Ten Bosch[9], Mattias Hofmans[1,2], Jan Philippé[2,4,10] and Tim Lammens[1,2,3]

BACKGROUND: Still 30–40% of pediatric acute myeloid leukemia (pedAML) patients relapse. Delineation of the transcriptomic profile of leukemic subpopulations could aid in a better understanding of molecular biology and provide novel biomarkers.
METHODS: Using microarray profiling and quantitative PCR validation, transcript expression was measured in leukemic stem cells (LSC, $n = 24$) and leukemic blasts (L-blast, $n = 25$) from pedAML patients in comparison to hematopoietic stem cells (HSCs, $n = 19$) and control myeloblasts (C-blast, $n = 20$) sorted from healthy subjects. Gene set enrichment analysis was performed to identify relevant gene set enrichment signatures, and functional protein associations were identified by STRING analysis.
RESULTS: Highly significantly overexpressed genes in LSC and L-blast were identified with a vast majority not studied in AML. *CDKN1A*, *CFP*, and *CFD* (LSC) and *HOMER3*, *CTSA*, and *GADD45B* (L-blast) represent potentially interesting biomarkers and therapeutic targets. Eleven LSC downregulated targets were identified that potentially qualify as tumor suppressor genes, with *MYCT1*, *PBX1*, and *PTPRD* of highest interest. Inflammatory and immune dysregulation appeared to be perturbed biological networks in LSC, whereas dysregulated metabolic profiles were observed in L-blast.
CONCLUSION: Our study illustrates the power of taking into account cell population heterogeneity and reveals novel targets eligible for functional evaluation and therapy in pedAML.

Pediatric Research (2021) 89:1695–1705; https://doi.org/10.1038/s41390-020-01199-3

IMPACT:

- Novel transcriptional targets were discovered showing a significant differential expression in LSCs and blasts from pedAML patients compared to their normal counterparts from healthy controls.
- Deregulated pathways, including immune and metabolic dysregulation, were addressed for the first time in children, offering a deeper understanding of the molecular pathogenesis.
- These novel targets have the potential of acting as biomarkers for risk stratification, follow-up, and targeted therapy.
- Multiple LSC-downregulated targets endow tumor suppressor roles in other cancer entities, and further investigation whether hypomethylating therapy could result into LSC eradication in pedAML is warranted.

INTRODUCTION

Pediatric acute myeloid leukemia (pedAML) is a rare hematological disease that accounts for 20% of all pediatric leukemias.[1] Cytogenetic risk stratification combined with response-guided therapeutic decisions considerably improved prognostication.[2,3] Unfortunately, still 30–40% of the good responders experience relapse.[2] During the past decade, ample evidence showed that relapse is associated with a high leukemic stem cell (LSC) load at diagnosis as well as LSC persistence during an apparent state of remission.[4–8] However, relapse still occurs in a considerable part of LSC^{low} pedAML patients identified by flow cytometry,[7] emphasizing a high need for a more profound molecular and phenotypic characterization of LSC.

Hitherto, most pedAML gene expression profiles (GEPs) were established in bulk leukemic samples,[9–15] not taking into account cellular heterogeneity, and thus fail to identify critical LSC-specific genes and pathways. In adult AML, by contrast, several LSC gene signatures have been developed these past years.[16–23] Interestingly, the LSC17 signature by Ng et al.[16] also held a prognostic value in pedAML.[24,25] Moreover, it was used to develop a pediatric-specific LSC6 score able to identify high-risk (HR) pedAML patients.[26] Unfortunately, this LSC signature also contains

[1]Department of Internal Medicine and Pediatrics, Ghent University, Ghent, Belgium; [2]Cancer Research Institute Ghent, Ghent, Belgium; [3]Department of Pediatric Hematology-Oncology and Stem Cell Transplantation, Ghent University Hospital, Ghent, Belgium; [4]Department of Laboratory Medicine, Ghent University Hospital, Ghent, Belgium; [5]Department of Pediatrics, University Hospital Gasthuisberg, Leuven, Belgium; [6]Department of Pediatric Hematology Oncology, University Hospital Saint-Luc, Brussels, Belgium; [7]Department of Pediatric Hematology Oncology, Queen Fabiola Children's University Hospital, Brussels, Belgium; [8]Department of Pediatric Hematology Oncology, University Hospital Liège, Liège, Belgium; [9]Department of Pediatric Hematology Oncology, University Hospital Brussel, Brussels, Belgium and [10]Department of Diagnostic Sciences, Ghent University, Ghent, Belgium
Correspondence: Barbara Depreter (barbara.depreter@uzbrussel.be)
These authors jointly supervised this work: Jan Philippé, Tim Lammens.

Received: 13 April 2020 Revised: 30 July 2020 Accepted: 25 September 2020
Published online: 17 October 2020

Table 1. Demographics of the pedAML patient cohort evaluated by qPCR.

	LSC		L-blast	
No. of patients (fractions)	24 (42)		25 (35)	
	Median (range)		Median (range)	
Age (years)	8 (1–17)		9 (1–17)	
WBC count ($\times 10^9$/L)	24.3 (3.1–336)*		25.5 (3.4-336)*	
Morphological blast count				
BM (%)	65 (27–96)†		70 (31–96)†	
PB (%)	48.5 (1–95)		51 (1–95)	
	N	%	*N*	%
Time point				
Diagnosis (Dx)	22	92%	24	96%
Relapse (R)	2	8%	1	4%
Treatment protocol				
DB-AML01	5	21%	9	36%
NOPHO-DBH AML2012	14	58%	13	52%
Other	5	21%	3	12%
Gender				
Female	11	46%	12	48%
Male	13	54%	13	52%
Sample				
BM and PB	11	46%	10	40%
Only BM	7	29%	9	36%
Only PB	6	25%	6	24%
LSC phenotype couples	7	29%	NA	NA
WT1 overexpression	15	63%	17	68%
Fusion transcript	14	58%	14	56%
CBF leukemia	9	38%	10	40%
AML1-ETO + *C-KIT*^{WT}	1	4%	1	4%
AML1-ETO + *C-KIT*^{MUT}	2	8%	2	8%
AML1-ETO + *C-KIT*ND	2	8%	3	12%
CBFB-MYH11	4	17%	4	16%
NPM1				
MUT	0	0%	0	0%
WT	23	96%	24	96%
Unknown	1	4%	1	4%
FLT3				
ITD	8	33%	10	40%
WT	16	67%	15	60%
CEBPA				
Single MUT	0	0%	0	0%
Double MUT	1	4%	1	4%
WT	22	92%	23	92%
Unknown	1	4%	1	4%
Karyotype				
Normal	6	25%	6	24%
Abnormal	12	50%	15	60%
Unknown	6	25%	4	16%
CNS involvement				
Yes	3	13%	4	16%
No	16	67%	18	72%
Unknown	5	21%	3	12%

Table 1. continued

	N	%	*N*	%
Risk classification				
SR	14	58%	16	64%
HR	4	17%	5	20%
Unknown	6	25%	4	16%
FAB classification				
M0	0	0%	1	4%
M1	3	13%	3	12%
M2	8	33%	8	32%
M3	1	4%	1	4%
M4	5	21%	6	24%
M5	6	25%	5	20%
M7	1	4%	1	4%

Characteristics of pedAML patients used for sorting and qPCR analysis of the most significant DEGs identified between LSC and HSC, and between L-blast and C-blast. Superscripts indicate one (*) or two (†) missing data.
PedAML pediatric acute myeloid leukemia, *LSC* leukemic stem cell, *L-blast* leukemic blast, *F* female, *M* male, *WBC* white blood cell, *BM* bone marrow, *PB* peripheral blood, *FLT3* fms-like tyrosine kinase receptor-3, *NPM1* nucleophosmin, *CEBPA* CCAAT/enhancer-binding protein alpha, *ITD* internal tandem duplication, *WT* wild type, *MUT* mutated, *FAB* French–British–American, *CBF* core-binding factor, *WT1* Wilms' tumor 1, *CNS* central nervous system, *NA* not available.

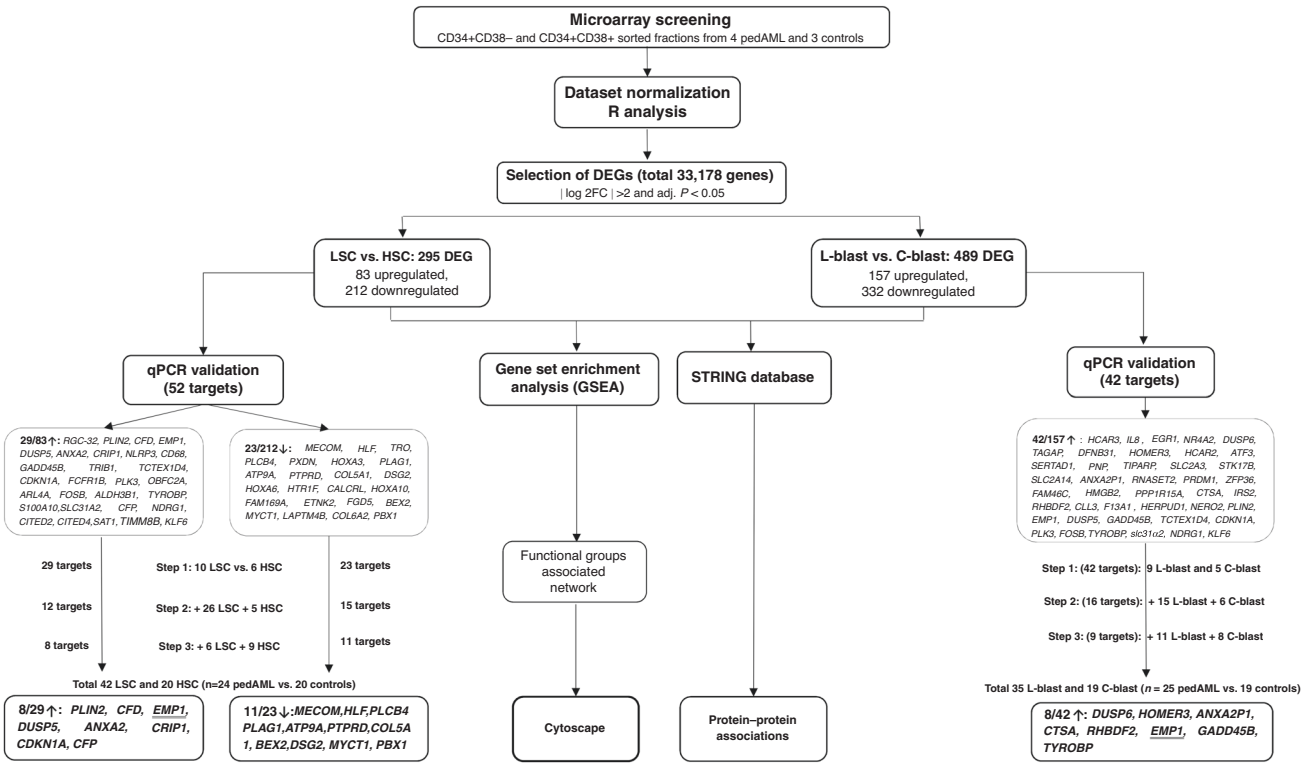
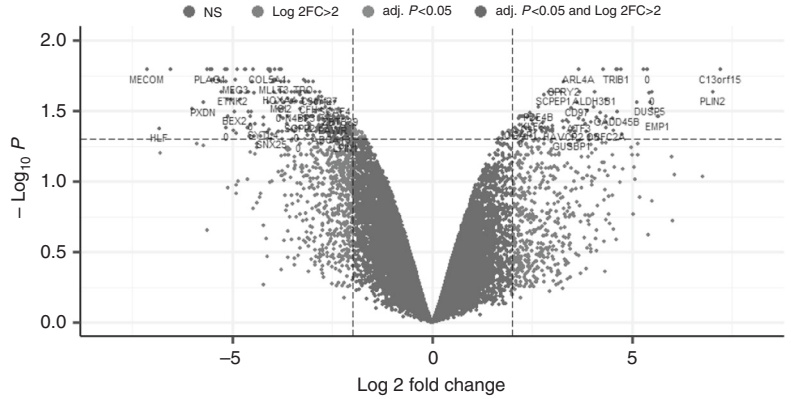


Fig. 1 Experimental setup and data processing steps. Step-by-step workflow illustrating the experimental and data processing steps pursued to filter out highly significant DEGs in leukemic subpopulations starting from a microarray profiling dataset. PedAML pediatric acute myeloid leukemia, CB cord blood, FC fold change, *P P* value, DEGs differentially expressed genes, LSC leukemic stem cell, HSC hematopoietic stem cell, L-blast leukemic blast, C-blast control blast.

genes that are expressed in hematopoietic stem cells (HSCs) and lack the inclusion of downregulated targets.[27] It was however shown that adding *PCD17*, an LSC-specific downregulated tumor suppressor gene (TSG), to the LSC17 score improved risk stratification in adult AML.[28] Hence, the identification of novel differentially expressed genes (DEGs) in pedAML leukemic subpopulations could aid in providing novel biomarkers for risk stratification, follow-up, and targeted therapy.

Here, we describe LSC and leukemic blast (L-blast) targets in pedAML discovered by microarray profiling followed by quantitative PCR (qPCR) validation following a cancer vs. normal (CvN) approach. We reveal deregulated pathways that have not yet



a

Top 50 LSC upregulated targets		
Gene symbol	log2FC	adj.P.Val
RGC-32	7.20	0.016
PLIN2	7.02	0.023
CFD	5.65	0.035
EMP1	5.64	0.035
ANXA2	5.50	0.031
DUSP5	5.44	0.027
CRIP1	5.28	0.016
NLRP3	5.13	0.028
CD68	4.75	0.046
42795	4.64	0.016
GADD45B	4.63	0.032
TRIB1	4.62	0.016
TCTEX1D4	4.54	0.043
CDKN1A	4.51	0.036
FCGR1B	4.36	0.038
ANXA2P1	4.35	0.026
PLK3	4.34	0.035
OBFC2A	4.34	0.042
ARL4A	4.27	0.016
SLED1	4.20	0.033
FOSB	4.18	0.037
ALDH3B1	4.08	0.023
ANXA1	4.05	0.030
TYROBP	3.96	0.038
S100A10	3.93	0.048
SLC31A2	3.87	0.032
CFP	3.85	0.044
ANXA2P3	3.83	0.039
NDRG1	3.69	0.023
ATF3	3.66	0.036
CD97	3.63	0.028
CRNDE	3.57	0.044
NAMPT	3.52	0.042
GUSBP1	3.50	0.048
CITED4	3.47	0.030
TOB1	3.45	0.044
SAT1	3.41	0.038
DOK2	3.37	0.037
HAVCR2	3.30	0.042
SPRY2	3.29	0.020
DUSP6	3.28	0.038
STK17B	3.22	0.023
SCPEP1	3.03	0.023
AEN	3.01	0.042
LOC439911	3.01	0.043
EMILIN2	3.00	0.041
SLC2A14	3.00	0.044
OGFRL1	2.94	0.043
TUBB4B	2.90	0.042
TUBB6	2.86	0.043

Top 50 LSC downregulated targets		
Gene symbol	log2FC	adj.P.Val
MECOM	–7.14	0.016
HLF	–6.82	0.042
TRO	–6.53	0.016
PLCB4	–6.01	0.031
PXDN	–5.72	0.028
HOXA3	–5.62	0.016
PLAG1	–5.56	0.016
ATP9A	–5.51	0.019
PTPRD	–5.50	0.016
COL5A1	–5.36	0.019
DSG2	–5.29	0.023
HOXA6	–5.28	0.020
HTR1F	–5.16	0.020
CALCRL	–5.16	0.020
NKAIN2	–5.13	0.016
HOXA10	–5.12	0.026
FAM169A	–5.02	0.026
ETNK2	–5.00	0.023
FGD5	–4.98	0.043
BEX2	–4.94	0.032
MEG3	–4.92	0.019
MYCT1	–4.90	0.046
LAPTM4B	–4.87	0.016
LOC100132526	–4.82	0.026
ZNF204P	–4.68	0.016
COL6A2	–4.68	0.016
WHAMMP3	–4.62	0.023
PBX1	–4.61	0.042
GPR56	–4.60	0.032
TMEM163	–4.58	0.023
LRRC16A	–4.58	0.048
CALN1	–4.55	0.040
BACH2	–4.52	0.032
SLC22A15	–4.50	0.020
RAPGEF2	–4.47	0.016
OPTN	–4.45	0.023
TSPYL5	–4.45	0.048
HBG1	–4.43	0.039
AIG1	–4.28	0.048
SYTL4	–4.22	0.040
BTBD11	–4.22	0.035
ANKS6	–4.13	0.045
HOPX	–4.12	0.050
SCML4	–4.11	0.027
LIN28B	–4.09	0.043
FAM3C	–4.05	0.017
SNX25	–4.02	0.047
GATA3	–3.98	0.020
MLLT3	–3.97	0.019
HOXA9	–3.92	0.019

b

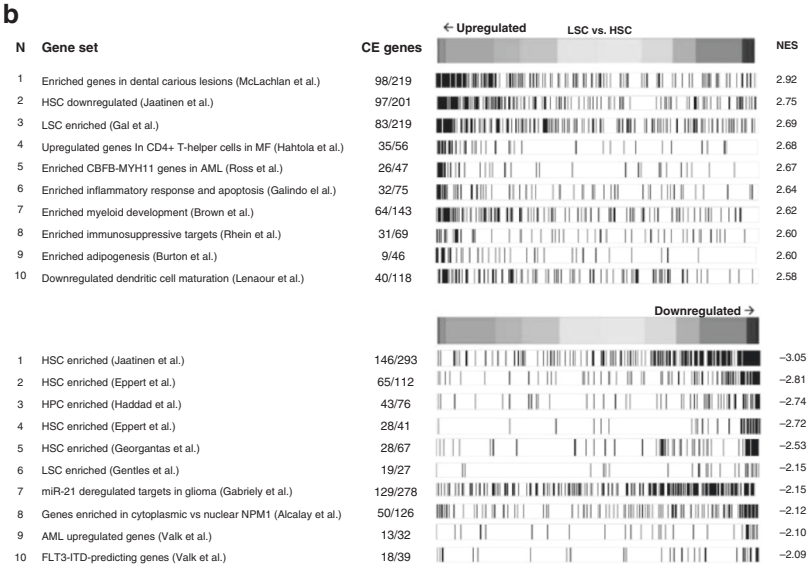
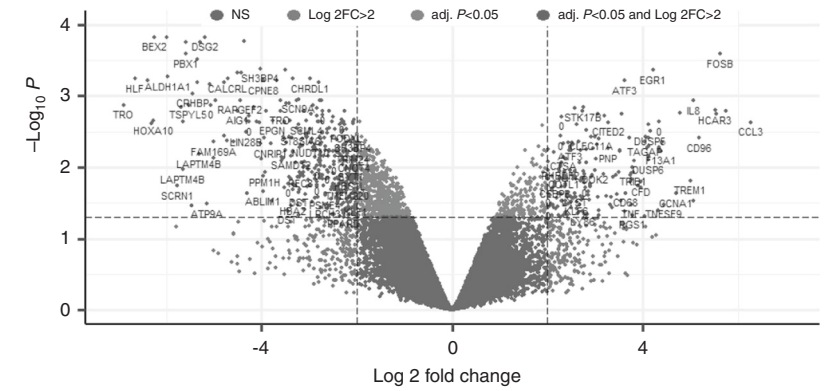


Fig. 2 Transcriptional differences and (anti-)correlated gene sets between LSC and HSC. a Visualization of DEGs identified between LSC ($n = 3$) and HSC ($n = 2$). Genes are plotted in a volcano plot as log 2 FC values against $-\log_{10}$ adj. P values. Thresholds $|\log 2\ FC| > 2$ and $-\log 10$ adj. $P < 0.05$ are shown as dashed lines. Genes selected as significantly different are highlighted in red. The top 50-ranked downregulated genes (left) and upregulated genes (right), annotated with gene symbols, are sorted by log 2 FC values. **b** Top 10 most correlated (top) and anti-correlated (bottom) gene sets identified through GSEA. The number of concordantly expressed (CE) genes/total genes and normalized enrichment scores (NESs) is shown for each gene set individually. FC fold change, DEGs differentially expressed genes, LSC leukemic stem cell, HSC hematopoietic stem cell.

Table 2. Enriched and suppressed pathways identified by functional protein association.

DEGs between LSC and HSC		DEGs between L-blast and C-blast
Enriched pathways (DEG = 83)	Suppressed pathways (DEG = 212)	Enriched pathways (DEG = 157)
Pathways in cancer	Tight junctions	FoxO signaling pathway
Cancer transcriptional misregulation	Cancer transcriptional misregulation	Cancer transcriptional misregulation
MAPK signaling pathway	Rap1 signaling pathway	Influenza A
Colorectal cancer	MAPK signaling pathway	HTLV-I infection
Osteoclast differentiation	Th17 cell differentiation	Osteoclast differentiation
NOD-like receptor signaling	Calcium signaling pathway	Epstein–Barr virus infection
Apoptosis		Cytokine receptor interaction
Breast cancer		
FoxO signaling pathway		

KEGG annotated pathways were derived by STRING analysis based on significant DEGs (number shown between brackets) identified for each comparison. LSC pathways that overlap with the top 10 ranked dysregulated pathways in adult AML (GSE 17054) are indicated in italic.
DEGs differentially expressed genes, *LSC* leukemic stem cell, *HSC* hematopoietic stem cell, *L-blast* leukemic blast, *C-blast* control blast.



a

Top 50 L-blast upregulated targets		
Gene symbol	**log2FC**	**adj.Pvalue**
CCL3	6.26	0.0023
C13orf15	5.74	0.0016
FOSB	5.62	0.0003
PLIN2	5.54	0.0018
HCAR3	5.53	0.0016
CD96	5.18	0.0039
SGK1	5.07	0.0294
IL8	5.06	0.0012
TREM1	5.00	0.0152
GADD45B	4.77	0.0017
CCNA1	4.71	0.0232
TNFSF9	4.43	0.0328
TCTEX1D4	4.40	0.0044
F13A1	4.37	0.0058
CDKN1A	4.34	0.0052
HSPA1B	4.32	0.0404
EGR1	4.22	0.0004
DUSP5	4.15	0.0031
C11orf96	4.12	0.0051
NR4A2	4.12	0.0038
EMP1	4.12	0.0025
DUSP6	4.11	0.0079
HOMER3	4.11	0.0069
PLK3	4.09	0.0052
HILPDA	4.04	0.0488
TAGAP	4.04	0.0044
SNAI1	3.96	0.0193
CFD	3,.96	0.0160
P2RY2	3.90	0.0179
DFNB31	3.81	0.0080
TRIB1	3.79	0.0115
RGS1	3.77	0.0466
IL13RA1	3.76	0.0149
SLED1	3.75	0.0146
TNF	3.75	0.0332
S100P	3.75	0.0121
HCAR2	3.72	0.0039
CD68	3.63	0.0228
HSPA8	3.61	0.0410
ATF3	3.61	0.0006
SERTAD1	3.55	0.0018
DNAJB1	3.53	0.0291
TNFAIP3	3.49	0.0116
NLRP3	3.45	0.0238
CD83	3.44	0.0128
CDKN2C	3.38	0.0135
CACNA2D4	3.36	0.0367
CLEC12A	3.33	0.0262
PLAU	3.31	0.0332
PNP	3.28	0.0054

b

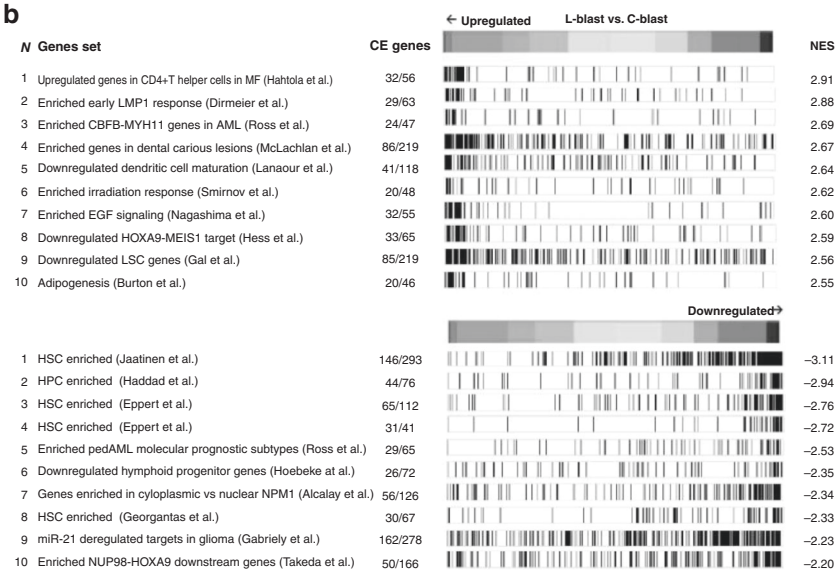


Fig. 3 Transcriptional differences and (anti-)correlated gene sets between L-blast and C-blast. a Visualization of genes identified to be upregulated in L-blast ($n = 4$) compared to C-blast ($n = 3$). Genes are plotted in a volcano plot as log 2 FC values against $-\log_{10}$ adj. P values. Thresholds $|\log 2\ FC| > 2$ and $-\log 10$ adj. $P < 0.05$ are shown as dashed lines, DEGs are highlighted in red. The top 50-ranked upregulated genes, sorted by log 2 FC values, are annotated with gene symbols on the right. **b** Top 10 most correlated (top) and anti-correlated (bottom) gene sets identified through GSEA, based on the DEGs between L-blast and C-blast. The number of concordantly expressed (CE) genes/total genes and normalized enrichment scores (NESs) is shown for each gene set individually. FC fold change, DEGs differentially expressed genes, L-blast leukemic blast, C-blast control blast.

been addressed in children, and aid in a further understanding of the pedAML molecular biology.

METHODS

Patients and controls

Bone marrow (BM) and/or peripheral blood (PB) from a total of 28 pedAML patients were selected based on cell availability (>50×10^6 after routine work-up) and CD34 positivity (≥1%). For 21/28 patients, both LSC and L-blast fractions were available, whereas for 3/28 and 4/28 patients only LSC or L-blast, respectively, could be evaluated. Demographics of patients used for LSC ($n = 24$) and L-blast ($n = 25$) expression evaluation are shown in Table 1. Details on treatment protocols and outcome definitions are described in Supplementary information. In addition, HSC and control myeloblasts (C-blast) were sorted from 19 and 20 healthy controls, respectively. Pediatric normal BM (NBM, $n = 9$, 12–18 years) was collected from posterior iliac crest during scoliosis surgery. Cord blood (CB, $n = 11$) was obtained after full-term delivery. All subjects gave their informed consent for inclusion before participation. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the University Hospital of Ghent (EC2015-1443 and EC2019-0294).

Cell sorting

Mononuclear cells were isolated by Ficoll density gradient (Axis-Shield), complemented by CD34 isolation if expression was <50% (CD34 MicroBead Kit, Miltenyi). Cell sorting was performed to isolate CD34+/CD38− and CD34+/CD38+ cells from patients and controls, defined as LSC and HSC, and L-blast and C-blast, respectively. Availability of both PB and BM yielded a total of 35 cell fractions in the LSC cohort (24 patients, 11 PB–BM couples) and in the L-blast cohort (25 patients, 10 PB–BM couples). The LSC compartment of seven patients could additionally be phenotypical distinguished based on differential expression of CD45RA ($n = 3$), CLL-1 ($n = 2$), CD123 ($n = 1$), or GPR56 ($n = 1$), yielding a total of 42 LSC fractions. Staining and sorting strategy are described in Supplementary information. Monoclonal antibodies are described in Supplementary Table S1. Sorted cells were collected in RPMI supplemented with 50% fetal calf serum, with a minimum post-sort purity >90%, spun down (10 min, 3000 r.p.m., 4 °C), and resuspended in 700 µL TRIzol for RNA extraction and cDNA synthesis as described in Supplementary information.

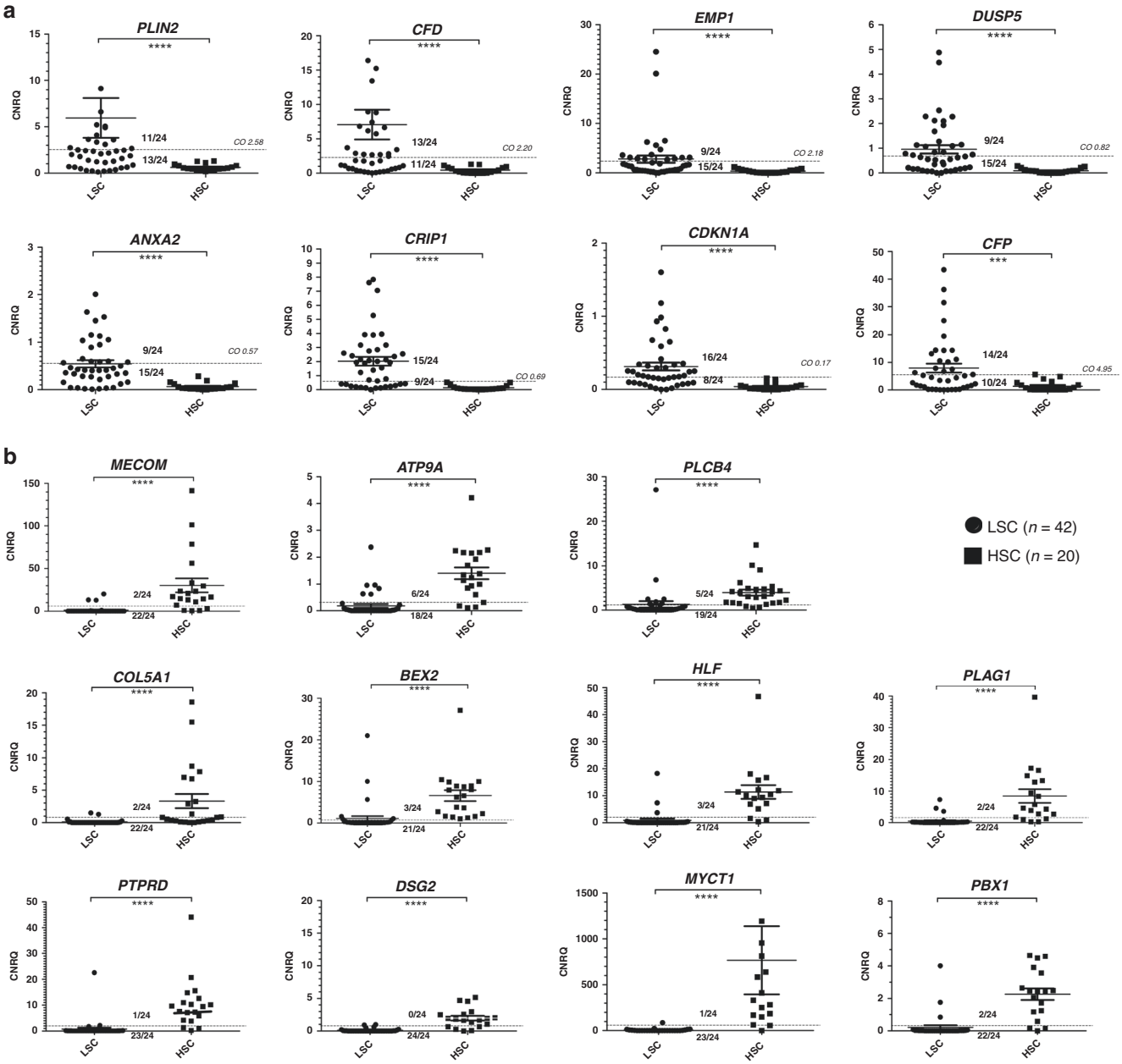


Fig. 4 qPCR validation of significantly DEGs in LSC vs. HSC. Targets with significant differential expression in LSC compared to HSC, shown by microarray analysis, were evaluated by qPCR analysis. Eight upregulated targets (**a**) and eleven downregulated targets (**b**) were evaluated in all available LSC ($n = 42$) and HSC ($n = 20$) fractions. Mean values are shown by horizontal lines, error bars indicate the 95% confidence interval of the mean, and the dotted line indicates the cut-off used to define overexpression, with the respective numbers of patients classified as high or low indicated above or below the line, respectively. *P* values (one-tailed) were calculated by the Mann–Whitney *U* test, and *, **, ***, or **** indicate the level of significance ($P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$, respectively). LSC leukemic stem cell, HSC hematopoietic stem cell, DEG differentially expressed gene, CNRQ calibrated normalized relative quantities.

Microarray downstream analyses

Microarray profiling was performed on LSC and L-blast sorted from 3/24 and 4/25 pedAML patients (Supplementary Table S2), respectively, next to two HSC and three C-blast fractions. Technical details are described in Supplementary information. DEGs were identified based on $|\log 2\ FC| > 2$ and adjusted *P* values (adj. *P*) < 0.05. Functional networks between protein–protein associations encoded by DEGs were identified by STRING at a high evidence level.[29] Only KEGG (Kyoto Encyclopedia of Genes and Genomes) annotated pathways were derived from significant pathway analysis. Gene set enrichment analysis (GSEA) was performed by combining independent omics datasets through pathway enrichment meta-analysis in order to obtain gene set enrichment signatures.[30] Unsupervised organization and visualization of enriched gene sets was performed in Cytoscape.[31] At least two clustered gene sets based on $P < 0.05$, false discovery rate (FDR) < 0.25 and Jaccard overlap combined index 0.375 were required for node visualization.

In addition, we re-analyzed the publicly available GSE 17054 microarray dataset,[17] containing GEPs of nine LSC from adult AML patients and four HSC from healthy adults.

Table 3. Frequency and magnitude of overexpressed targets in LSC and L-blast.

Expression	LSC upregulated targets							
	PLIN2	*CFD*	*EMP1*	*DUSP5*	*ANXA2*	*CRIP1*	*CDKN1A*	*CFP*
HSC ($n = 20$)								
Mean	0.88	0.59	0.53	0.16	0.11	0.18	0.08	1.35
SD	0.85	0.80	0.83	0.33	0.23	0.25	0.05	1.80
LSC ($n = 24$)								
Mean	5.95	7.06	2.80	0.96	0.55	2.03	0.32	7.93
SD	13.90	14.00	4.76	1.10	0.48	2.02	0.35	10.30
Cut-off	2.58	2.20	2.18	0.82	0.57	0.69	0.17	4.95
Overexpression	46%	54%	38%	38%	38%	63%	67%	58%
Fold change	6.79	11.89	5.28	5.99	4.83	11.34	3.78	5.87
	L-blast upregulated targets							
	DUSP6	*HOMER3*	*PNP*	*ANXA2P1*	*CTSA*	*RHBDF2*	*EMP1*	*GADD45B*	*TYROBP*
C-blast ($n = 19$)									
Mean	0.66	0.03	0.64	0.19	0.43	0.72	0.45	0.19	9.33
SD	0.65	0.04	0.34	0.13	0.27	0.32	0.47	0.14	12.80
L-blast ($n = 25$)									
Mean	3.22	0.45	1.31	0.67	1.42	3.30	3.57	0.78	63.80
SD	2.87	0.52	0.81	0.54	0.81	4.04	6.71	0.65	39.70
Cut-off	1.97	0.11	1.32	0.44	0.97	1.37	1.39	0.47	34.93
Overexpression	36%	84%	47%	72%	44%	88%	68%	72%	72%
Fold change	4.85	13.62	2.04	3.52	3.29	4.56	8.02	4.16	6.84

Mean expression and standard deviations (SDs) calculated for the most significantly upregulated targets in LSC vs. HSC and L-blast vs. C-blast. Cut-offs were calculated based on the mean expression measured in the healthy counterparts plus two times the SD. Percentage of patients classified as having overexpression are shown, together with the magnitude of overexpression expressed as the fold change of the mean expression in the leukemic vs. normal compartment.

Real-time quantitative PCR

Ninety-four targets were selected for qPCR validation based on the magnitude and significance of differential expression and feasibility of primer development. Available cell fractions were subdivided into three cohorts based on sample availability, while respecting a balanced distribution of the genetic variations, and subsequently measured by qPCR in three steps. Cell fractions with only limited material available were reserved for the most significant DEGs and evaluated in the third step, hereby avoiding sample paucity to present as an issue. Analytical details of qPCR experiments are described in Supplementary information, primers are shown in Supplementary Table S3. Data analysis were performed according to state-of-the-art methods[30] as follows: Ct values were corrected for primer efficiency and expressed as relative quantities. Normalized relative quantities (NRQs) were calculated against the expression of housekeeping genes *GAPD*, *HPRT1*, and *TBP*. To allow inter-run comparison, calibrated NRQ values (CNRQ) were generated by taking into account the expression of an inter-run calibrator. Target-specific cut-offs for overexpression were calculated based on the average expression plus two standard deviations measured in the respective normal counterparts.

RESULTS

In order to identify novel targets in LSC and L-blast, a CvN approach integrating different experimental and analysis techniques was used (Fig. 1). Each set of DEGs was explored in the context of intertwining pathways and databases to define the top deregulated pathways, and GSEA to gain insight into (anti-) correlated biological pathways. Full lists of significant DEGs and (anti-)correlated gene sets identified between LSC vs. HSC and L-blasts vs. C-blasts are shown in Supplementary Tables S4–S13. Below, we discuss the top-ranked significant DEGs and (anti-) correlated gene sets for each comparison separately.

Immune dysregulation separates LSC in pedAML from HSC

The expression of 295 targets significantly differed between LSC and HSC, with 83 targets up- and 212 downregulated in LSC. The top 50 ranked DEGs is shown in Fig. 2a. Well-known oncogenes were present among the highest LSC-overexpressed targets (e.g., *CFD*, *ANXA2*, *NLRP3*) next to genes with yet undefined roles in AML (e.g., *PLIN2*, *CRIP1*). The top 10 most downregulated genes also contained targets for which no role was yet reported in pedAML (e.g., *ATP9A*, *PLCB4*, *COL5A1*).

Analysis of functional protein associations (STRING) showed upregulated pathways in LSC related to (breast) cancer, osteoclast differentiation, and apoptosis, whereas transcriptional misregulation, and Rap1/MAPK signaling were downregulated (Table 2). Myeloid cell activation networks involved in immune response were enriched in LSC (FDR 4.6e − 6), whereas networks related to stimuli responses, signaling, and cell communication were suppressed (FDR 5.9e − 3). From a total of 3650 signatures available through GSEA, 240 and 18 gene sets were significantly enriched or suppressed in LSC, respectively. The top 10 LSC-enriched gene sets involved LSC signatures, inflammatory response, apoptosis, immune suppression, and adipogenesis, whereas HSC signatures were anti-correlated (Fig. 2b). Unsupervised visualization (Cytoscape) identified LSC-upregulated pathways related to abnormal cell division,

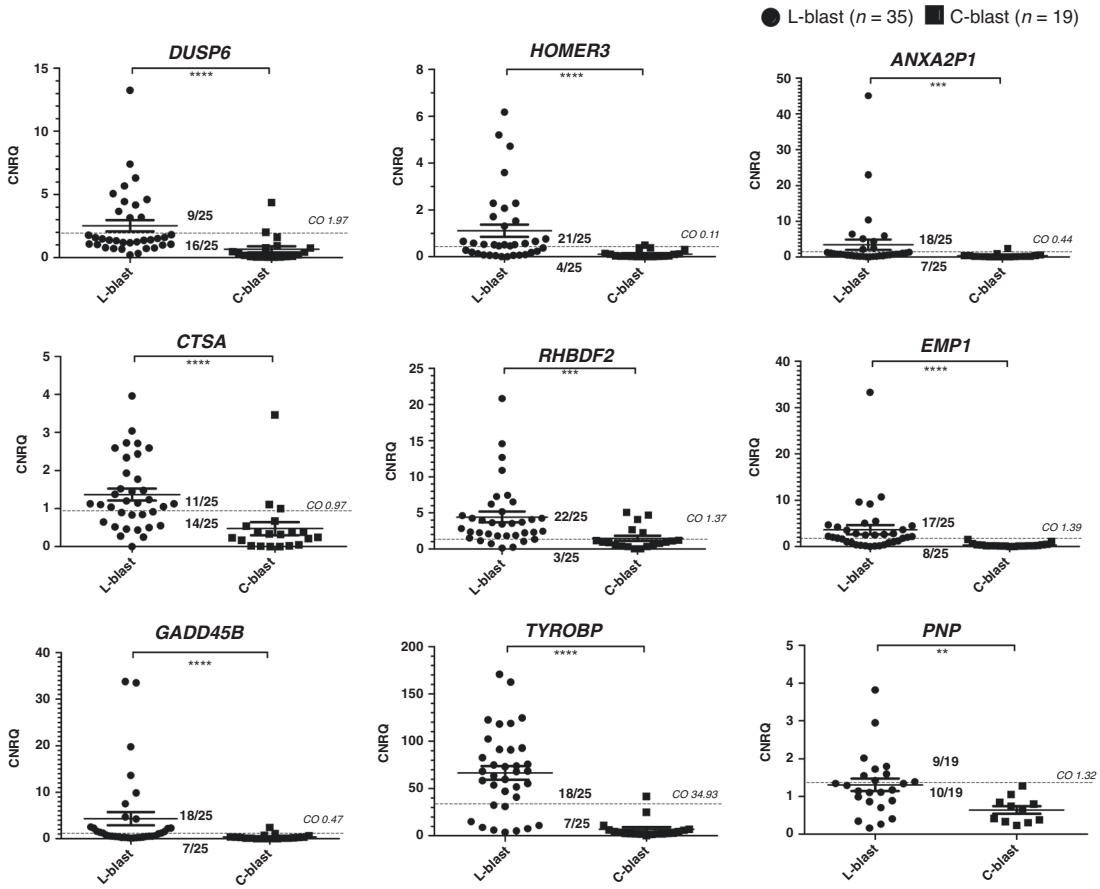


Fig. 5 qPCR validation of significantly DEGs in L-blast vs. C-blast. Targets with significant differential expression in L-blast compared to C-blast, shown by microarray analysis, were evaluated by qPCR analysis. Nine upregulated targets were evaluated in all available L-blast ($n = 35$) and C-blast ($n = 19$) fractions. Mean values are shown by horizontal lines, error bars indicate the 95% confidence interval of the mean, and the dotted line indicates the cut-off used to define overexpression, with the respective numbers of patients classified as high or low indicated above or below the line, respectively. *P* values (one-tailed) were calculated by the Mann–Whitney *U* test, and *, **, ***, or **** indicate the level of significance ($P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$, respectively). L-blast leukemic blast, C-blast control blast, DEG differentially expressed gene, CNRQ calibrated normalized relative quantities.

quiescence, autoimmune regulation, and environmental stress, while gene sets involved in normal quiescence and cell death signaling were downregulated (Supplementary Fig. S1A). Altogether, these data suggest that dysregulation of the immune system contributes to the leukemic transformation of stem cells in pedAML.

As pediatric and adult AML represent two genetically distinct diseases,[32] we wondered if this heterogeneity is also reflected in the stem cell transcriptome. To this end, we re-analyzed the GSE 17054 microarray dataset from Majeti et al.[17] and identified 486 significant DEGs between adult LSC and HSC (Supplementary Fig. S2A). Comparing the set of LSC-HSC DEGs identified in pediatric vs. adult AML revealed 71 common downregulated targets (Supplementary Fig. S2B), which was translated into mutual repressed pathways, for example, tight junction and MAPK signaling[17] (Table 2). In sharp contrast, only three common LSC-upregulated transcripts were identified (*TYROBP*, *CFP*, and *PTH2R*).

Metabolic changes in pedAML L-blasts enhance proliferation compared to normal counterparts

We identified 157 and 332 significantly up- and downregulated targets in L-blast vs. C-blast. The top 50 upregulated transcripts is shown in Fig. 3a. Pathways enriched in L-blasts involved cancer transcriptional misregulation, FoxO signaling, and cytokine–cytokine receptor interaction (Table 2).

Functional protein network analysis (STRING) illustrated a significant enrichment in L-blasts of stimuli responses (FDR 6.8e − 11) and metabolic processes (FDR 1.4e − 05). GSEA identified 163 enriched and 23 suppressed gene sets in L-blast vs. C-blast. The top-ranked adipogenesis gene set correlates with metabolic dysregulation, whereas increased EGF signaling and decreased stemness signatures relate to high proliferation (Fig. 3b). Among others, upregulated cancer and EGFR signaling, and downregulated death signaling, were confirmed by unsupervised clustering (Supplementary Fig. S1B).

Interestingly, the top-ranked (anti-)correlated gene sets identified in LSC (Fig. 2b) and L-blast (Fig. 3b) partially overlapped, and also enriched and suppressed pathways were recurrent (Table 2). Therefore, we sought similarities in the DEGs identified in LSC and L-blast. From the 83 and 157 significantly upregulated genes in LSC and L-blasts, respectively, 49 genes appeared to be common (Supplementary Fig. S3A). On the other hand, 134 targets were mutually downregulated from a total set of 212 and 332 transcripts, respectively (Supplementary Fig. S3B). Taken together, we conclude that LSC and L-blast share pan-leukemic molecular aberrancies compared to their normal counterparts.

Novel candidate targets in pedAML leukemic subpopulations validated by qPCR

A selection of the top-ranked up- and downregulated targets in LSC vs. HSC and upregulated targets in L-blast vs. C-blast, as identified by microarray analysis, were validated by qPCR (29/83,

23/212, and 42/157, respectively). A total of 94 targets were evaluated according to a three-step exclusion strategy, allowing the most significant DEGs to be evaluated in the highest number of cell fractions. Patients were dichotomized as high and low for the highest differentially expressed targets. Per target, overexpression was correlated to cytogenetic and molecular markers, and if patients were included in the NOPHO-DBH AML2012 protocol, correlated with event-free survival (EFS).

First, differential expression was confirmed at a significant level ($P < 0.05$) for 24/29 LSC upregulated targets (Supplementary Fig. S4A). Moreover, differential expression was significant at $P < 0.01$ with concomitant low HSC expression for 12/24 targets. Expression of these 12 targets was further evaluated using additional LSC and HSC fractions (Supplementary Fig. S4B). Too low LSC expression, or too high HSC expression, led to the exclusion of 4/12 targets. Finally, expression of *PLIN2*, *CFD*, *EMP1*, *DUSP5*, *ANXA2*, *CRIP1*, *CDKN1A*, and *CFP* was evaluated in all fractions and shown to be highly significantly overexpressed in LSC ($n = 42$) compared to HSC ($n = 20$) (Fig. 4a). Overexpression of these eight targets was observed in 38–67% of the patients, with *CDKN1A*, *CRIP1*, *CFP*, and *CFD* overexpressed in more than half of the patients (Table 3). Expression was averagely 4- to 12-fold higher in LSC compared to HSC, with *CFD* and *CRIP1* the most upregulated targets. *CFP* overexpression significantly correlated to *FLT3*-ITD mutations, for example, 46% in *CFP*-high ($n = 14$) vs. 10% in *CFP*-low ($n = 10$) patients ($P = 0.043$). High *ANXA2* levels beneficially impacted EFS at a borderline significant level ($P = 0.061$, 4 *ANXA2*-high vs. 10 *ANXA2*-low patients, Supplementary Fig. S5A).

Second, significant LSC downregulation was confirmed for 21/23 targets (Supplementary Fig. S4C, $P < 0.05$). Among these, 15/21 targets were even downregulated at $P < 0.01$, with virtual absent expression in LSC. Including more sample fractions led to the exclusion of 4/15 targets due to too low differential expression levels (Supplementary Fig. S4D). Evaluation of all fractions illustrated that *MECOM*, *HLF*, *PLCB4*, *PLAG1*, *ATP9A*, *PTPRD*, *COL5A1*, *BEX2*, *DSG2*, *MYCT1*, and *PBX1* transcripts are highly significantly repressed in LSC compared to HSC ($P < 0.0001$, Fig. 4b). These targets appeared to be suppressed in 75–100% of all patients. *BEX2* downregulation was significantly anticorrelated to *KMT2A*-rearrangements ($P < 0.01$), as previously demonstrated using cell lines.[33] On the other hand, the previously reported association between *PLAG1* and inv[16] (p13q22), or between *MYCT1* and FAB classifications M1/M5/M6, was not confirmed ($P > 0.05$).[34,35] Interestingly, 9/11 pedAML LSC-downregulated targets were also significantly suppressed in adult LSC (Supplementary Fig. S1B). Furthermore, *PBX1*, *MYCT1*, *HLF*, *ATP9A*, and *PLCB4* appeared to be also suppressed in other pediatric malignancies, suggesting a possible tumor suppressor role (Supplementary Fig. S6).

Third, qPCR confirmed a significant upregulation in L-blast vs. C-blast for 16/42 targets (Supplementary Fig. S4E, $P < 0.05$). Further analysis using more samples showed that expression was too low in L-blasts, or too high in C-blasts, for 7/16 targets (Supplementary Fig. S4F). Evaluation of the remaining nine targets in all samples illustrated that *DUSP6*, *HOMER3*, *ANXA2P1*, *CTSA*, *RHBDF2*, *EMP1*, *GADD45B*, *TYROBP*, and *PNP* transcripts were highly significantly overexpressed in L-blasts ($n = 36$) vs. C-blasts ($n = 19$) (Fig. 5). Five out of nine targets (*RHBDF2*, *HOMER3*, *ANXA2P1*, *GADD45B*, and *TYROBP*) were overexpressed in more than two-thirds of the patients, with *HOMER3* showing the highest differential expression (Table 3). Interestingly, *HOMER3*-high cases ($n = 21$) showed significantly less inv[16] (p13q22) ($P = 0.014$) and FAB M4 ($P = 0.031$), compared to *HOMER3*-low pedAML ($n = 4$). *DUSP6*, overexpressed in one-third of the patients, was previously shown to be significantly associated with *FLT3*-ITD in adult AML,[36] which we could not confirm in a pediatric setting ($P = 0.49$). *PNP*-high patients showed a significant lower EFS (Supplementary Fig. S5B), which was confirmed by Cox log-rank univariate analysis (hazard ratio 9.24, $P = 0.04$), but did not remain significant in multivariate analysis.

DISCUSSION

We here describe a novel set of differentially expressed targets in LSC and L-blast of pedAML patients identified based on a CvN approach. Moreover, we reveal previously unexplored deregulated pathways in these leukemic subpopulations of children with AML.

Eight targets were found highly significantly overexpressed in LSC compared to HSC. *CDKN1A*, *ANXA2*, *EMP1*, and *CFD* were previously linked to leukemogenesis, whereas the role of *PLIN2*, *DUSP5*, *CRIP1*, and *CFP* in AML remains elusive. *CDKN1A*, *CRIP1*, *CFP*, and *CFD* were found to be most frequently overexpressed (>50% of the patients), with *CRIP1* and *CFD* showing the highest differential expression. *CDKN1A* might represent an interesting target for LSC eradication, since elevated expression was reported to maintain LSC activity,[37,38] and *CDKN1A* knockdown indirectly reversed stem cell quiescence.[39] Overexpression of *CFP* and *CFD* in LSC of pedAML patients suggest a disturbed complement pathway regulation. *CFP* overexpression was significantly associated with *FLT3*-ITD mutations in pedAML ($P = 0.043$) and concomitantly overexpressed in adult AML, suggesting a role as age-independent LSC marker in (high-risk) AML. *CFD* expression was previously linked to poor outcome in adult AML,[40] and its prognostic value in children awaits validation. Last, we found that patients with high *ANXA2* LSC expression (38%, 4.8-fold higher expression than HSC) show a trend towards prolonged EFS ($P = 0.061$). This finding is in agreement with a previously reported favorable prognostic effect of *ANXA2* in bulk pedAML cells.[41] Hence, *ANXA2* could hold promise as a prognostic biomarker in pedAML. Although only the top-ranked DEGs were validated by qPCR in our study, microarray analysis additionally revealed interesting targets for flow cytometric validation. Our data confirm upregulated CD96 and CLECL12 expression in pedAML LSC compared to HSC.[42–44] A potential role for targeting CD180 and CD68 in LSC, or their qualification as follow-up marker, deserves further attention.

Suppression of LSC-specific downregulated targets was highly consistent across the different genetic subgroups (75–100% of all patients). Since several of these targets endow tumor suppressor roles in other cancer entities and are inactivated upon promoter hypermethylation, further investigation whether hypomethylating therapy could result into LSC eradication in pedAML is warranted. *MYCT1* was already identified as a TSG in AML,[35] and deserves the highest attention since expression levels are 100-fold lower in LSC compared to HSC. *PBX1* was previously reported to act as both an oncogene and TSG.[45] Suppressed *PTPRD* levels in LSC is in agreement with a previous report on *PTPRD* suppression in pedAML bulk leukemic cells.[46]

Transcriptional misregulation in cancer, osteoclastogenesis, and tight junction pathways were dysregulated in LSC. The "cancer transcriptional misregulation" pathway is associated with myeloid leukemogenesis and held responsible for tumorigenic epigenetic abnormalities.[47,48] Distortion of osteoclastogenesis and tight junction pathways might provide LSC an advantage over HSC during homing towards the endosteal–vascular niche. The observed immune dysregulation, separating LSC from HSC, strokes with a previous statement that multiple inflammatory signaling pathways are involved in the generation of pre-LSCs.[49]

L-blast upregulated targets were overexpressed in a larger portion of patients compared to LSC upregulated targets (36–88%, median 72% vs. 38–67%, median 50%, respectively). Among these, only *DUSP6* and *HOMER3* were previously addressed in AML. *DUSP6* is an important cellular signaling regulator overexpressed in AML.[36] *HOMER3* relates to the occurrence and development of AML,[50] and increased levels were significantly associated with favorable cytogenetics in adult AML.[51] Since *HOMER3* also showed the highest differential expression compared to C-blasts, targeting

could be of therapeutic value. *CTSA* also represent an interesting target, since several other cathepsins were shown to have a diagnostic, prognostic, or therapeutic significance in AML.[52–55] *GADD45B* is known to be involved in negative growth control during myeloid differentiation,[56] and suggested to play a role in the tumorigenesis of colorectal carcinoma.[57] Finally, high *PNP* expression was significantly associated with a worse EFS. Studies with PNP inhibitors in relapsed and refractory leukemias are ongoing,[58] and investigation of their applicability in pedAML might be worthwhile.

Deregulated pathways identified in L-blast compared to their normal counterparts suggest that disturbed regulation of cell cycling, apoptosis, glycolysis/gluconeogenesis, and oxidative stress resistance promote the maintenance and proliferation of blasts in a leukemic setting. Indeed, adapting to hypoxic conditions and switching from oxidative phosphorylation towards glycolysis was shown to correlate with an aggressive disease course in solid cancers.[59] Gluconeogenesis blocking, previously proposed as general antitumor therapeutic strategy, should be further explored in pedAML.[60]

EMP1 was the only target highly significantly upregulated in LSC and L-blast (38% and 68% of patients, respectively, $P < 0.0001$). Although Ng et al.[16] did not elaborate on its role, *EMP1* was included in the LSC17 score,[25] but not retained in the pedAML-specific LSC6-score.[26] However, based on the here observed CD38-independent overexpression, and the previously reported in vitro targetability of *EMP1* in B-ALL,[61] its role as a therapeutic target in pedAML should be further explored.

We detected a high molecular heterogeneity between pediatric and adult AML at the stem cell level. LSC populations from both entities shared 71 suppressed transcripts, but only three mutual upregulated targets (*TYROBP*, *CFP*, and *PTH2R*) were identified. *TYROBP* and *CFP* have not been functionally associated with AML, and their role in LSC transformation should be further explored. *PTH2R*, on the other hand, is known to be upregulated in AML and MDS,[62] including adult AML LSC.[63] Further research is warranted to evaluate whether these three targets could serve as pan-LSC targets, irrespective of the age of onset. Although these findings further underline the distinct biology between pediatric and adult AML,[32] it should be taken into account that, due to the small number of patients evaluated, the genetic subtypes might also play a role besides the age of the patients.

Our investigation has some limitations, including its design as a single-center study recruiting patients from different clinical trials, and the lack of protein expression data. We acknowledge that extensive in vitro and in vivo validation of several of the identified targets is needed to fully unlock the potential of the presented dataset. Nevertheless, we here provide a repository to the pediatric acute leukemia community allowing to fulfill the high need for alternative therapeutic strategies in pediatric AML. The limited number of patient samples asks for data to be interpreted with caution within the framework of generalizability. Although promising, these data need confirmation in larger, preferentially multicenter trials, as survival analyses were performed in a limited number patients. It is important to acknowledge that the pedAML cohort included one secondary AML evolved from juvenile myelomonocytic leukemia, one acute promyelocytic leukemia, and two relapsed patients, all excluded from outcome analysis.

In conclusion, we here report a unique set of LSC and L-blast-specific overexpressed genes in pedAML. Most targets have not been studied in AML, and are involved in immune regulation, apoptosis, adhesion, or intracellular signaling, making them attractive candidates for functional studies, refining signatures, and targeted therapy. Inflammatory pathways and immune regulation are critical biological networks perturbed in pedAML LSC, and L-blast presents a high proliferative cell cycle activity combined with metabolic dysregulation. In addition, we identified novel LSC-specific downregulated targets, often described as TSGs in solid tumors, of which some are relevant in adult AML LSC or other pediatric hematological diseases.

ACKNOWLEDGEMENTS

Our gratitude goes to Dr. F. Plasschaert, C. Matthys of the Cord Blood Bank, and the Department of Pediatric Hematology and Oncology of the Ghent University Hospital (Ghent, Belgium) for providing samples. We thank all patients and their parents for their participation in the study, as well as the data managers involved in the clinical trials. This research was funded by the Belgian Foundation against Cancer (grant 2014-265), FOD-KankerPlan (Actie29, grant to J.P.), vzw Kinderkankerfonds (grant to T.L.), and the Research Foundation—Flanders (Fonds voor Wetenschappelijk Onderzoek Vlaanderen, FWO, grant 1113117 to B. Depreter).

AUTHOR CONTRIBUTIONS

Conception and design, acquisition of data, or analysis and interpretation of data: B. Depreter, T.L., J.P., and M.H.; drafting the article or revising it critically for important intellectual content: B.D.M., E.T., K.V., L.D.; final approval of the version to be published: B.D.P., T.L., J.P., B.D.M., M.H., E.T., K.V., L.D., B.Denys, A.U., A.V.D., M.-F.D., J.V. d.W.T.B.

ADDITIONAL INFORMATION

The online version of this article (https://doi.org/10.1038/s41390-020-01199-3) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

Patient consent: All patients gave consent for inclusion in this study.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Rasche, M. et al. Successes and challenges in the treatment of pediatric acute myeloid leukemia: a retrospective analysis of the AML-BFM trials from 1987 to 2012. *Leukemia* **32**, 2167–2177 (2018).
2. De Moerloose, B. et al. Response-guided chemotherapy for pediatric acute myeloid leukemia without hematopoietic stem cell transplantation in first complete remission: results from protocol DB AML-01. *Pediatr. Blood Cancer* **66**, e27605 (2019).
3. Abrahamsson, J. et al. Response-guided induction therapy in pediatric acute myeloid leukemia with excellent remission rate. *J. Clin. Oncol.* **29**, 310–315 (2011).
4. De Kouchkovsky, I. & Abdul-Hay, M. Acute myeloid leukemia: a comprehensive review and 2016 update. *Blood Cancer J.* **6**, e441 (2016).
5. Hope, K. J., Jin, L. & Dick, J. E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat. Immunol.* **5**, 738–743 (2004).
6. Shlush, L. I. et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* **547**, 104–108 (2017).
7. Hanekamp, D. et al. Leukaemic stem cell load at diagnosis predicts the development of relapse in young acute myeloid leukaemia patients. *Br. J. Hematol.* **183**, 512–516 (2018).
8. Witte, K. E. et al. High proportion of leukemic stem cells at diagnosis is correlated with unfavorable prognosis in childhood acute myeloid leukemia. *Pediatr. Hematol. Oncol.* **28**, 91–99 (2011).
9. Bachas, C. et al. Gene expression profiles associated with pediatric relapsed AML. *PLoS ONE* **10**, e0121730 (2015).
10. Pigazzi, M. et al. MLL partner genes drive distinct gene expression profiles and genomic alterations in pediatric acute myeloid leukemia: an AIEOP study. *Leukemia* **25**, 560–563 (2011).
11. Jo, A. et al. High expression of EVI1 and MEL1 is a compelling poor prognostic marker of pediatric AML. *Leukemia* **29**, 1076–1083 (2015).
12. Andersson, A. et al. Microarray-based classification of a consecutive series of 121 childhood acute leukemias: prediction of leukemic and genetic subtype as well as of minimal residual disease status. *Leukemia* **21**, 1198–1203 (2007).
13. Balgobind, B. V. et al. Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. *Haematologica* **96**, 221–230 (2011).
14. Zangrando, A., Dell'orto, M. C., Te Kronnie, G. & Basso, G. MLL rearrangements in pediatric acute lymphoblastic and myeloblastic leukemias: MLL specific and lineage specific signatures. *BMC Med. Genomics* **2**, 36 (2009).

15. Ross, M. E. et al. Gene expression profiling of pediatric acute myelogenous leukemia. *Blood* **104**, 3679–3687 (2004).
16. Ng, S. W. et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).
17. Majeti, R. et al. Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. *Proc. Natl Acad. Sci. USA* **106**, 3396–3401 (2009).
18. Eppert, K. et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17**, 1086–1093 (2011).
19. Forsberg, E. C. et al. Molecular signatures of quiescent, mobilized and leukemia-initiating hematopoietic stem cells. *PLoS ONE* **5**, e8785 (2010).
20. Gal, H. et al. Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells. *Leukemia* **20**, 2147–2154 (2006).
21. Gentles, A. J., Plevritis, S. K., Majeti, R. & Alizadeh, A. A. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA* **304**, 2706–2715 (2010).
22. Saito, Y. et al. Identification of therapeutic targets for quiescent, chemotherapy-resistant human leukemia stem cells. *Sci. Transl. Med.* **2**, 17ra9 (2010).
23. de Leeuw, D. C. et al. Attenuation of microRNA-126 expression that drives CD34 +38- stem/progenitor cells in acute myeloid leukemia leads to tumor eradication. *Cancer Res.* **74**, 2094–2105 (2014).
24. Duployez, N., Preudhomme, C. & Cheok, M. A 17-gene-expression profile to improve prognosis prediction in childhood acute myeloid leukemia. *Oncotarget* **9**, 33869–33870 (2018).
25. Duployez, N. et al. The stem cell-associated gene expression signature allows risk stratification in pediatric acute myeloid leukemia. *Leukemia* **33**, 348–357 (2018).
26. Elsayed, A. H. et al. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. *Leukemia* https://doi.org/10.1038/s41375-019-0604-8 (2019).
27. Richard-Carpentier, G. & Sauvageau, G. Bringing a leukemic stem cell gene signature into clinics: are we there yet? *Cell Stem Cell* **20**, 300–301 (2017).
28. Xu, Z. J. et al. Reduced protocadherin17 expression in leukemia stem cells: the clinical and biological effect in acute myeloid leukemia. *J. Transl. Med.* **17**, 102 (2019).
29. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, 607–613 (2019).
30. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
31. Reimand, J. et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, 482–517 (2019).
32. Bolouri, H. et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* **24**, 103–112 (2018).
33. Rohrs, S. et al. Hypomethylation and expression of BEX2, IGSF4 and TIMP3 indicative of MLL translocations in acute myeloid leukemia. *Mol. Cancer* **8**, 86 (2009).
34. Landrette, S. F. et al. Plag1 and Plagl2 are oncogenes that induce acute myeloid leukemia in cooperation with Cbfb-MYH11. *Blood* **105**, 2900–2907 (2005).
35. Fu, S. et al. Overexpression of MYCT1 inhibits proliferation and induces apoptosis in human acute myeloid leukemia HL-60 and KG-1a cells in vitro and in vivo. *Front. Pharm.* **9**, 1045 (2018).
36. Arora, D. et al. Expression of protein-tyrosine phosphatases in acute myeloid leukemia cells: FLT3 ITD sustains high levels of DUSP6 expression. *Cell Commun. Signal.* **10**, 19 (2012).
37. Peng, L. et al. Structural maintenance of chromosomes 4 is required for leukemia stem cell maintenance in MLL-AF9 induced acute myeloid leukemia. *Leuk. Lymph.* **59**, 2423–2430 (2018).
38. Andries, V. et al. NBPF1, a tumor suppressor candidate in neuroblastoma, exerts growth inhibitory effects by inducing a G1 cell cycle arrest. *BMC Cancer* **15**, 391 (2015).
39. Korthuis, P. M. et al. CITED2-mediated human hematopoietic stem cell maintenance is critical for acute myeloid leukemia. *Leukemia* **29**, 625–635 (2015).
40. Laverdiere, I. et al. Complement cascade gene expression defines novel prognostic subgroups of acute myeloid leukemia. *Exp. Hematol.* **44**, 1039–1043 (2016).
41. Niu, Y. et al. Distinct prognostic values of Annexin family members expression in acute myeloid leukemia. *Clin. Transl. Oncol.* **21**, 1186–1196 (2019).
42. Chavez-Gonzalez, A. et al. Expression of CD90, CD96, CD117, and CD123 on different hematopoietic cell populations from pediatric patients with acute myeloid leukemia. *Arch. Med. Res.* **45**, 343–350 (2014).
43. Darwish, N. H. et al. Acute myeloid leukemia stem cell markers in prognosis and targeted therapy: potential impact of BMI-1, TIM-3 and CLL-1. *Oncotarget* **7**, 57811–57820 (2016).
44. Hanekamp, D., Cloos, J. & Schuurhuis, G. J. Leukemic stem cells: identification and clinical application. *Int. J. Hematol.* **105**, 549–557 (2017).
45. Dardaei, L., Longobardi, E. & Blasi, F. Prep1 and Meis1 competition for Pbx1 binding regulates protein stability and tumorigenesis. *Proc. Natl Acad. Sci. USA* **11**, 896–905 (2014).
46. Song, L. et al. Protein tyrosine phosphatases receptor type D is a potential tumour suppressor gene inactivated by deoxyribonucleic acid methylation in paediatric acute myeloid leukaemia. *Acta Paediatr.* **105**, 132–141 (2016).
47. Jasielec, J., Saloura, V. & Godley, L. A. The mechanistic role of DNA methylation in myeloid leukemogenesis. *Leukemia* **28**, 1765–1773 (2014).
48. Yin, X. et al. Identification of long non-coding RNA competing interactions and biological pathways associated with prognosis in pediatric and adolescent cytogenetically normal acute myeloid leukemia. *Cancer Cell Int.* **18**, 122 (2018).
49. Hemmati, S., Haque, T. & Gritsman, K. Inflammatory signaling pathways in preleukemic and leukemic stem cells. *Front. Oncol.* **7**, 265 (2017).
50. Li, Z. et al. Growth and differentiation effects of Homer3 on a leukemia cell line. *Asian Pac. J. Cancer Prev.* **14**, 2525–2528 (2013).
51. Stirewalt, D. L. et al. Identification of genes with abnormal expression changes in acute myeloid leukemia. *Genes Chromosomes Cancer* **47**, 8–20 (2008).
52. Jain, M., Bakhshi, S., Shukla, A. A. & Chauhan, S. S. Cathepsins B and L in peripheral blood mononuclear cells of pediatric acute myeloid leukemia: potential poor prognostic markers. *Ann. Hematol.* **89**, 1223–1232 (2010).
53. Pandey, G., Bakhshi, S., Thakur, B., Jain, P. & Chauhan, S. S. Prognostic significance of cathepsin L expression in pediatric acute myeloid leukemia. *Leuk. Lymph.* **59**, 2175–2187 (2018).
54. Alatrash, G. Targeting cathepsin G in myeloid leukemia. *Oncoimmunology* **2**, e23442 (2013).
55. Pandey, G. et al. Prognostic and therapeutic relevance of cathepsin B in pediatric acute myeloid leukemia. *Am. J. Cancer Res.* **9**, 2634–2649 (2019).
56. Liebermann, D. A. & Hoffman, B. Myeloid differentiation (MyD) primary response genes in hematopoiesis. *Oncogene* **21**, 3391–3402 (2002).
57. Wang, L. et al. Abnormal expression of GADD45B in human colorectal carcinoma. *J. Transl. Med.* **10**, 215 (2012).
58. Korycka, A., Lech-Maranda, E. & Robak, T. Novel purine nucleoside analogues for hematological malignancies. *Recent Pat. Anticancer Drug Discov.* **3**, 123–136 (2008).
59. Calle, E. E. & Kaaks, R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat. Rev. Cancer* **4**, 579–591 (2004).
60. Zi, F. et al. Metformin and cancer: an existing drug for cancer prevention and therapy. *Oncol. Lett.* **15**, 683–690 (2018).
61. Aries, I. M. et al. EMP1, a novel poor prognostic factor in pediatric leukemia regulates prednisolone resistance, cell proliferation, migration and adhesion. *Leukemia* **28**, 1828–1837 (2014).
62. Zhang, Z. et al. Integrated bioinformatic analysis of microarray data reveals shared gene signature between MDS and AML. *Oncol. Lett.* **16**, 5147–5159 (2018).
63. Bonardi, F., Fusetti, F., Deelen, P., van Gosliga, D. & Vellenga, E. A proteomics and transcriptomics approach to identify leukemic stem cell (LSC) markers. *Mol. Cell Proteom.* **12**, 626–637 (2013).