Identification of specific cluster of differentiation genes in acute myeloid leukemia by combined Bayesian and ABC analysis

Cornelia Brendel[1*] Elisabeth Mack[1*], Miriam Frech[1], Andreas Neubauer[1], Torsten Haferlach[2] and Alfred Ultsch[3] ([*]equal contribution)

[1]Department of Internal Medicine and Hematology, Oncology and Immunology, Philipps University Marburg, Medical Center of the University Gießen and Marburg, Marburg, Germany;
[2] MLL Münchner Leukämie Labor GmbH, Max Lebsche Platz 31, 81377 Munich, Germany
[3] Data Bionics Research, Philipps University Marburg, Hans-Meerwein Strasse, 35043 Marburg, Germany

Introduction

Acute Myeloid Leukemia (AML), particularly Acute Promyelocytic Leukemia (APL) is a medical emergency that requires urgent initiation of treatment. Immunophenotyping by flow cytometry enables rapid diagnosis of AML including separation of APL from non-APL AML (npAML) and monitoring minimal/measurable residual disease. However, leukemia associated immunophenotypes are described by complex patterns of cell surface molecules rather than by a few unique markers. Here, we applied an unbiased combinatorial mathematical approach in order to identify specific cluster of differentiation (CD) genes that discriminate APL, npAML and normal controls. Moreover, we aimed to uncover combinations of CD molecules that could serve as targets for novel therapies involving chimeric antigen receptor T-cells (CARTs).

Methods

We combined Bayesian and calculated ABC analysis to compare microarray gene expression profiles of 266 primary npAML and 15 APL patient samples to 109 healthy controls. Differential expression of selected genes was confirmed by qRT-PCR and flow cytometry in the AML cell lines lines NB4, MV4-11, U937 and HL-60. Associations of CD cluster expression with patient overall survival in eight independent AML datasets were assessed by PRECOG analysis

Figures

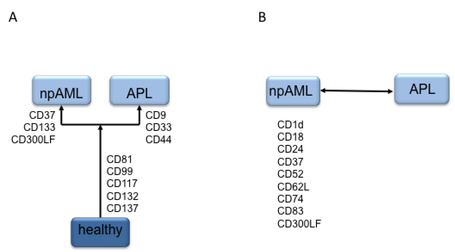

Figure 1

Up-regulation of CD genes according to combined Bayesian and ABC analysis of microarray gene expression data in a cohort of 281 AML samples including 15 APL samples compared to 109 healthy individuals.
Three dichotomies were chosen: npAML versus healthy and APL versus healthy (A) and APL versus npAML (B). Comparison of npAML or APL versus healthy revealed in 11 differently regulated genes. Direct comparison of APL versus NPAML samples revealed 9 differently expressed CD genes.

AML: Acute myeloid leukemia; APL: Acute promyelocytic leukemia; npAML: Non-APL AML

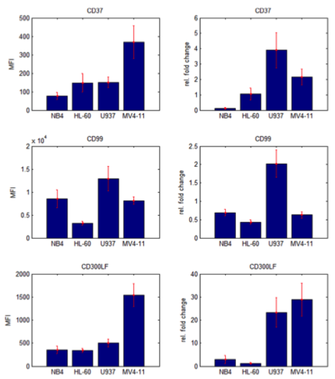

Figure 2

RNA and surface protein expression of CD37, CD99 and CD300LF in four AML cell lines.
Gene expression of CD37, CD99 and CD300LF was determined by RT-qPCR (left) and surface protein was measured by flow cytometry. RNA expression of CD molecules is indicated as fold change relative to a reference RNA calibrator (not shown): Flow cytometry results are depicted as mean fluorescent intensities. CD37 was significantly lower expressed in NB4 compared to HL-60, U937 and MV4-11, while CD300LF was low in NB4 and HL-60 against U937 and MV4-11 ($p<0.05$; student´s t-test). CD99 expression was similar for NB4 and other AML cell lines with highest expression of both RNA and protein in U937.
Error bars indicate standard deviations from three independent experiments.

MFI: mean fluorescence intensity; rel fold change: gene expression level determined with RTqPCR

Table 1

Functional Abstraction of biological processes in AML.
After Over-Representation Analysis (ORA) of 281 AML versus 109 healthy samples gene ontology (GO) annotations were performed. Subsequently, Functional Abstraction was applied and 12 significant terms from the hierarchy of 308 total relevant terms were identified.

GO term description	GO term ID	p value
transport	GO:0006810	0.00002
cell motility	GO:0048870	0.016
biological adhesion	GO:0022610	0.007
cell communication	GO:0007154	0.03
immune system process	GO:0002376	0.003
response to stimulus	GO:0050896	0.0003
cell differentiation	GO:0030154	0.004
cell death	GO:0008219	0.01
growth	GO:0040007	0.02
cellular component organization	GO:0016043	0.008
biological regulation	GO:0065007	0.01
single-organism metabolic process	GO:0044710	0.0003

Table 2

Prediction performance for dual specific CAR therapies.
ABC analysis of all pairs of the 371 CD genes was performed and revealed seven significant combinations of concomitantly overexpressed CD molecules. A combination of CD99 and CD103 excluded all healthy samples and was applicable in 68% of npAML samples.

No	CDA	CDB	alias	NPAML Accuracy	Healthy Accuracy	Total Accuracy
1	CD99	CD63		90	84	88
2	CD99	CD113t-C	HBB	89	82	87
3	CD99	CD184	CXCR4	77	87	80
4	CD99	CD103	ITGAE	68	100	77
5	CD99	CD50	ICAM3	68	94	76
6	CD99	CD257	TNFSF13B	64	94	73
7	CD99	CD42	MT2A	61	97	71

Results

The Microarrays contained measurements for 317 CD genes that qualified for further analysis. We found up-regulation of CD37, CD81, CD99, CD117/KIT, CD132/IL2RG, CD133/PROM1, CD137/TNFRSF9 and CD300LF in npAML versus healthy and up-regulation of CD9, CD33, CD44, CD81, CD99, CD117, CD132 and CD137 in APL compared to controls. Differential expression of CD37 and CD300LF in npAML and APL was confirmed experimentally in four AML cell lines by quantitative PCR and flow cytometry and by direct *in silico* comparison of expression profiles. Both CD37 and CD300LF were associated with inferior patient survival by PRECOG analysis. Moreover, our data suggest CD1d, CD18/ITGB2, CD24, CD52, CD62L, CD74, CD83 as novel markers discriminating APL and npAML. We did not identify promising targets for potential CART therapies in AML as all pairs of CDs that were AML-specific contained CD99, which is also a T cell antigen and not solely expressed on AML blasts.

Conclusions

Our unbiased straight-forward mathematical approach to highly complex biological data is a powerful tool to establish sound hypotheses for preclinical and clinical studies evaluating diagnostic markers and therapeutic targets in AML and other tumor entities.

Disclosures

The authors declare no competing interest.