

## Article

# A Bioinformatics View on Acute Myeloid Leukemia Surface Molecules by Combined Bayesian and ABC Analysis

Michael C. Thrun <sup>1,\*</sup>, Elisabeth K. M. Mack <sup>2,†</sup>, Andreas Neubauer <sup>2</sup>, Torsten Haferlach <sup>3</sup>, Miriam Frech <sup>2</sup>, Alfred Ultsch <sup>1</sup> and Cornelia Brendel <sup>2</sup>

<sup>1</sup> Department of Mathematics and Computer Science, Philipps-University Marburg, Hans-Meerwein-Straße, 35032 Marburg, Germany

<sup>2</sup> Department of Hematology, Oncology and Immunology, Philipps-University Marburg, 35043 Marburg, Germany

<sup>3</sup> MLL Münchner Leukämielabor GmbH, Max Lebsche Platz 31, 81377 Munich, Germany

\* Correspondence: mthrun@informatik.uni-marburg.de

† These authors contributed equally to this work.

**Abstract:** “Big omics data” provoke the challenge of extracting meaningful information with clinical benefit. Here, we propose a two-step approach, an initial unsupervised inspection of the structure of the high dimensional data followed by supervised analysis of gene expression levels, to reconstruct the surface patterns on different subtypes of acute myeloid leukemia (AML). First, Bayesian methodology was used, focusing on surface molecules encoded by cluster of differentiation (CD) genes to assess whether AML is a homogeneous group or segregates into clusters. Gene expressions of 390 patient samples measured using microarray technology and 150 samples measured via RNA-Seq were compared. Beyond acute promyelocytic leukemia (APL), a well-known AML subentity, the remaining AML samples were separated into two distinct subgroups. Next, we investigated which CD molecules would best distinguish each AML subgroup against APL, and validated discriminative molecules of both datasets by searching the scientific literature. Surprisingly, a comparison of both omics analyses revealed that CD339 was the only overlapping gene differentially regulated in APL and other AML subtypes. In summary, our two-step approach for gene expression analysis revealed two previously unknown subgroup distinctions in AML based on surface molecule expression, which may guide the differentiation of subentities in a given clinical–diagnostic context.

**Keywords:** Bayesian machine learning; gene expressions; leukemia; cluster of differentiation genes; CD genes



**Citation:** Thrun, M.C.; Mack, E.K.M.; Neubauer, A.; Haferlach, T.; Frech, M.; Ultsch, A.; Brendel, C. A Bioinformatics View on Acute Myeloid Leukemia Surface Molecules by Combined Bayesian and ABC Analysis. *Bioengineering* **2022**, *9*, 642. <https://doi.org/10.3390/bioengineering9110642>

Academic Editor: Luca Mesin

Received: 26 September 2022

Accepted: 22 October 2022

Published: 3 November 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

High-throughput molecular analyses are becoming increasingly affordable in clinical medicine, and have significantly improved our understanding of diseases such as cancer [1,2]. However, overwhelming amounts of “big omics data” still provoke the challenge of extracting meaningful information with clinical benefit. Deep molecular insights can provide a basis to shift organ/tissue-centered approaches to tumor diagnosis and therapy towards precision oncology, i.e., the integration of genetic and genomic data to estimate a patient’s prognosis and guide treatment decisions [3]. In fact, molecular classification systems have been proposed already for some entities such as breast cancer [4] or acute myeloid leukemia (AML) [5]. However, large-scale sequencing or gene expression studies are of limited practical value in specific clinical settings concerning both diagnostics and therapy. Notably, specific low-complexity biomarkers can be determined in less time than system-wide profiles, and molecular therapeutics need precisely defined structures rather than superordinate patterns as their particular sites of action. Therefore, rapid diagnostics and effective molecular therapies require knowledge about “the important few” features that low-throughput testing procedures or targeted compounds can address.

Diagnosis of acute myeloid leukemia (AML) is of great medical importance in the field of hematology, as one of its subtypes, specifically, acute promyelocytic leukemia (APL), is associated with a substantial risk of early death from bleeding complications [6]. Although diagnosis of AML can be established in many cases by cytology, discrimination of APL and non-promyelocytic AML (np-AML) is sometimes challenging and requires further examination. Throughout the past two decades, flow cytometry has become an additional powerful tool to rapidly characterize AML and distinguish this entity from other acute leukemias or myeloid neoplasms based on the immunophenotyping of blasts [7,8]. Up to now, single distinct AML-specific cell surface markers have not been identified, thus the European Leukemia Net (ELN) guidelines recommend a broad antibody panel for diagnostic testing [9].

In this work, we applied an unbiased mathematical approach that combines Bayesian and computed ABC analysis [10] on gene expression data in order to identify the most informative cluster of differentiation (CD) genes on AML blasts that could distinguish APL from np-AML. Specifically, we analyzed microarray gene expression profiles for a series of 390 primary patient samples comprising 266 np-AML, 15 APL, and 109 control cases (healthy/non-leukemia patients) [11], and RNA-Seq-data from 135 np-AML and 15 APL patient samples from the TCGA LAML project [12].

## 2. Materials and Methods

### 2.1. Gene Expression Datasets

Two AML gene expression datasets were analyzed in this study. The first one (provided by T.H) was a subset of a previously published microarray dataset [11] and included gene expression profiles of 281 diagnostic AML samples (266 np-AML, 15 APL) and 109 non-leukemia/healthy individuals that had been obtained using Affymetrix HG-U133 Plus 2.0 microarrays. The ethics board approved the analysis of this dataset for the purpose of this study at the faculty of medicine, University of Marburg (No. 138/16). The second dataset comprised RNA-Seq data from the TCGA LAML project for 135 np-AML and 15 APL patients [12], and can be downloaded from the Broad GDAC Firehose platform ([http://firebrowse.org/?cohort=LAML&download\\_dialog=true](http://firebrowse.org/?cohort=LAML&download_dialog=true), accessed on 3 April 2017). It can also be downloaded from the GDC Data Portal for each case separately (<https://portal.gdc.cancer.gov/> (accessed on 3 April 2017). Clinical annotations for the LAML dataset included FAB subgroups as the most diverse established classification system for AML, while more recent classifications such as WHO 2016 [8] or ELN 2017/2022 classifications [7,13] were not provided. Thus, we chose FAB classes as the reference for our calculations, which we also considered appropriate given that our bioinformatic explorations were directed towards the cell surface. A list of 417 CD genes that were examined for expression in np-AML, APL, and non-leukemia samples (available only in the microarray dataset) was compiled manually using the list of CD molecules provided by the HUGO gene nomenclature committee (<https://www.genenames.org/data/genegroup/#!/group/471>, accessed on 30 November 2019) and a query to the NCBI gene database (<https://www.ncbi.nlm.nih.gov/gene>, accessed on 2 November 2017; query term “CDxxx AND Homo Sapiens [Organism]”).

### 2.2. Calculation of Bayesian Decision Borders for Expressed and Unexpressed Genes

Initial exploratory data analysis with techniques taken from [14] indicated that the distribution of the log-transformed expression data for each of the two complete sample sets followed a multimodal distribution. Mixture models are the standard statistical tool for such applications [15,16], and further preprocessing was not required. Gaussian mixture models (GMM) were fitted to the overall distribution of the log-transformed microarray and the RNA-Seq gene expression data and subjected to Bayesian analysis using the R package “AdaptGauss” available on the Comprehensive R Archive Network (CRAN) [15]. In brief, the procedure incorporated the following definitions and operations: In Gaussian Mixture

models (GMM), the probability density of the measurements,  $GMM(x)$ , is represented as a weighted sum of Gaussians

$$GMM(x) = \sum_{i=1}^M w_i N(x|m_i, s_i) = \sum_{i=1}^M w_i \cdot \frac{1}{\sqrt{2\pi s_i}} \cdot e^{-\frac{(x-m_i)^2}{2s_i^2}} \tag{1}$$

where  $N(x|m_i, s_i)$  denotes Gaussian normal distributions with mean  $m_i$  and standard deviation  $s_i$ ; the weights,  $w_i$ , which add up to 1, indicate the relative contribution of each component; and  $M$  denotes the number of components in the mixture. For each dataset, a suitable value for  $M$  was selected based on the Akaike information criterion [17], and the GMM was adapted to the data using an expectation–maximization algorithm. This approach resulted in a GMM with  $M = 2$ . Chi-square tests [18] were applied to estimate the probability that the GMM did not adequately describe the logarithmized gene expression data. The chi-square tests yielded  $p$ -values of  $p < 0.001$  for both GMMs. Posteriors for each GMM were calculated as follows:

$$p(c_i|x) = \frac{p(c_i) * p(x|c_i)}{\sum_{i=1}^M p(c_i) * p(x|c_i)} \tag{2}$$

with

$$\sum_{i=1}^M p(c_i) * p(x|c_i) = \sum_{i=1}^M w_i * N(m_i, s_i) = GMM(x) \tag{3}$$

in which  $p(x|c_i)$  is the likelihood to generate data in this class (conditional probability of being in the mode of gene expression values  $x$ );  $p(c_i)$  is the probability of choosing a class (prior); and  $p(c_i|x)$  is the posterior.

Bayesian decision borders between unexpressed and expressed genes, defined as the posteriors of  $p = 0.5$  with log expression values of  $x = 2.96$  (microarray) and  $x = 1.48$  (RNA-Seq), were calculated, for which the back transformation yielded expression values of 912 and 30.2. In detail, the left Gaussian contained low gene expression values and the right Gaussian contained high gene expression values. The posteriors defined the probability of a gene expression value either belonging to the left or to the right Gaussian. By exploiting the GMM, the expression datasets were automatically normalized to the closed interval  $[0, 1]$  using the Bayes posteriors. A value of  $p(c_i|x = gene_a) = 0$  indicates definitive underexpression of  $gene_a$ , and a value of  $p(c_i|x = gene_a) = 1$  denotes definitive overexpression of  $gene_a$ .

### 2.3. Identification of AML Subgroups

The subgroups  $G_k$   $k = 1, 2, 3$  were identified with the Ward algorithm accessible in the R package “FCPS” available on CRAN [19]. As the algorithm was applied to the Bayes posteriors  $p(c_i|x)$ , no normalization was necessary. The inspection of the dendrogram indicated optimal clustering. Here, significant changes in fusion levels of the ultrametric portion of the Euclidean distance in the Ward algorithm (y-axis) indicated the best cut [20].

### 2.4. Selection of Differentially Expressed Genes (Deg)

In the next step, the probability  $|p_{deg}|$  that  $gene_a$  is differentially expressed is defined in Equation (4) as the difference of group average Bayes posteriors  $\overline{p}_{G_k} = \frac{1}{N_k} \sum_{u \in G_k} p_u(c_i|gene_a)$  with  $p_u$  being the posterior value of a subject’s case  $u \in G_k$  of a subgroup  $G_k$  with its cardinality  $N_k = |G_k|$ , as follows:

$$p_{deg}(gene_a) = \overline{p}_{G_k}(c_i|gene_a) - \overline{p}_{G_j}(c_i|gene_a) \tag{4}$$

with  $G_k$  and  $G_j$  (where by  $k = 1, 2, 3, j = 1, 2, 3, j \neq k$ ) being the two groups of comparison (e.g., APL vs. AML) and the sign (+, −) defining the direction. A value of  $p_{deg}(gene_a) = -2$  indicates definitive underexpression, and a value of  $p_{deg}(gene_a) = +2$  denotes definitive overexpression in relation to the compared groups.

Differences in gene expression levels were analyzed for the dichotomies APL vs. AML1, APL vs. AML2, AML1 vs. AML2, APL vs. normal, AML1 vs. normal, and AML2 vs. normal in the microarray dataset, and APL vs. AML1, APL vs. AML2, and AML1 vs. AML2 in the RNA-Seq dataset, using Cohen’s D with varying variances and group sizes [21] as a measure of the effect size. The most important effect sizes, i.e., the most significant differences in gene expression, were identified by computed ABC analysis [10] using the R package “ABCanalysis” available on CRAN (<http://cran.r-project.org/web/packages/ABCanalysis/index.html> (accessed on 3 April 2017)). The ABC curve can visualize the data by graphically representing the cumulative distribution function closely related to the Lorenz curve. Using the ABC curve, the algorithm calculates the optimal limits by exploiting the mathematical properties pertaining to the distribution of analyzed items. Recursively applying ABC analysis three times on all probabilities  $|p_{deg}|$  yielded the used thresholds: genes were considered to be expressed if their difference exceeded the threshold of 0.3 for RNA-Seq (0.7 for microarray data), because genes with values above the threshold were in group A (A and B respectively) in the third computed ABC analysis, and considered to be the most important ones.

### 3. Results

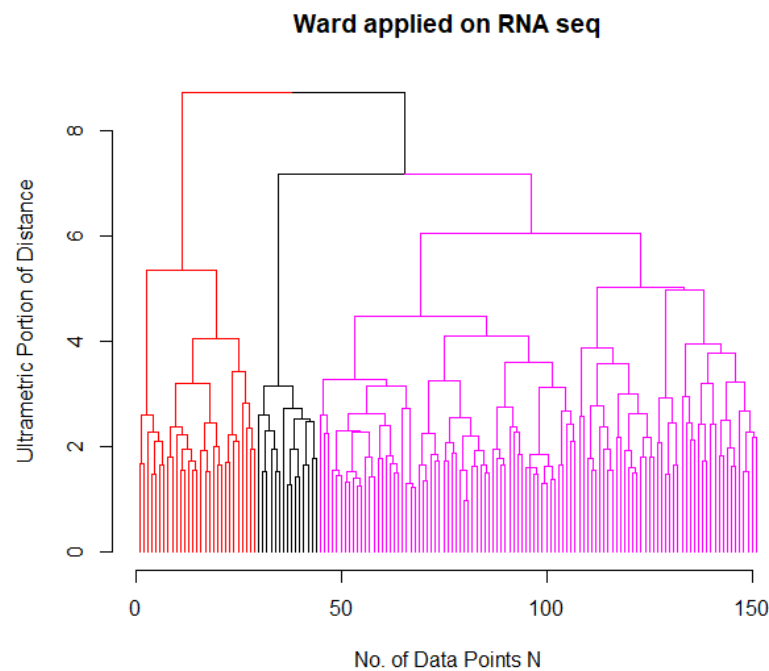
At first, the gene expressions in both datasets were logarithmized. Distribution analysis [14] revealed multimodal distributions for the CD genes. Therefore, distributions were modeled with Gaussian mixtures as described in the methods section. The models allowed for calculating Bayesian posteriors, which were used for further analysis.

#### 3.1. Subgroups of AML Based on the Expression of CD Genes

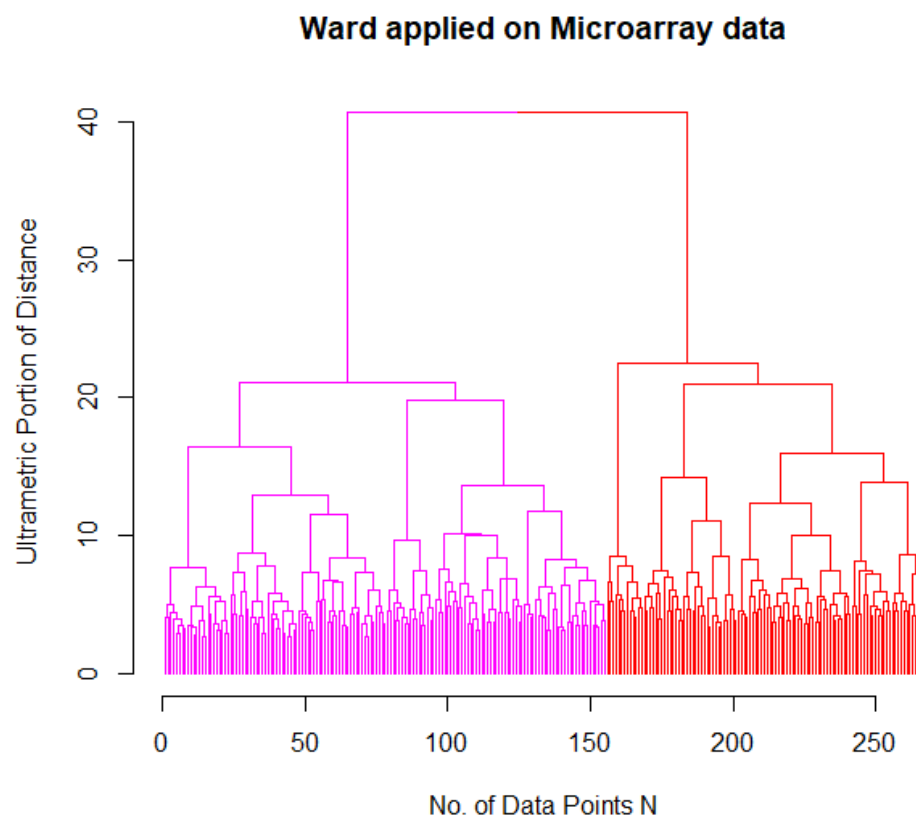
In the next step, hierarchical clustering, using the Ward algorithm, of the RNA-Seq data revealed three major clusters marked in the dendrogram in black in Figure 1, consisting solely of APL/AML M3 cases (Ward cluster 3). Beyond this separation, AML had two subgroups (magenta and red). In order to compare the RNA-expression-based AML clustering with the historical classification of the French–American–British group [22], both classifications were aligned in Table 1. Within the first cluster of Ward, M1, M2, and M0 were predominant, and within the second cluster, M5. M4 was equally distributed, and for M6 to M8, only a few cases were available. FAB M3 is historically defined as APL, and was only found in Ward cluster 3. Hence, the Ward dendrogram for the AML patients without APL of the microarray dataset was also investigated. In line with the clustering results from the RNA-Seq data, the non-APL group revealed a cluster structure of two subgroups, as depicted in Figure 2.

**Table 1.** Contingency table comparing Ward clustering of the RNA-Seq dataset and FAB classification [22]. Abbreviations: M0, undifferentiated acute myeloblastic leukemia; M1, acute myeloblastic leukemia with minimal maturation; M2, acute myeloblastic leukemia with maturation; M3, acute promyelocytic leukemia (APL); M4, acute myelomonocytic leukemia; M5, acute monocytic leukemia; M6, acute erythroid leukemia; M7, acute megakaryoblastic leukemia.

Ward Cluster	M0	M1	M2	M3	M4	M5	M6	M7
1 (magenta)	15	34	36	0	17	2	2	1
2 (red)	0	3	1	0	12	13	0	0
3 (black)	0	0	0	15	0	0	0	0



**Figure 1.** Ward clustering applied to the RNA-Seq data revealed a cluster structure of three groups  $G_k$ ,  $k = 1, 2, 3$ , one consisting only of APL and two of AML with the FAB classification of Table 1 which also depicts the colors of this figure. Clustering and dendrogram generation were computed with the R package “FCPS” available on the Comprehensive R Archive Network (CRAN) [19].



**Figure 2.** Dendrogram of Ward clustering applied to microarray np-AML patient samples revealed two subgroups (magenta and red). Clustering and dendrogram generation were computed with the R package “FCPS”, available on CRAN. The dataset provided normal controls in addition to APL. Both diagnostic entities were not used in the clustering.

A dichotomy of np-AML is neither recognized in the present WHO classification [7] nor in the stratification for treatment strategies. The separation of np-AML to APL is in agreement with previous research [23]. In addition, it is mentioned in one prior publication that AML divides into one APL cluster and two AML clusters with principal component analysis using 37 microarray samples [24]. Alignment of the two clusters to FAB classification could either be interpreted as a myeloid vs. monocytic differentiation pattern or as an immature (M0–2 and M6/M7) vs. a more mature (M4–M5) biologic nature of the leukemia blasts, as suggested by Goardon et al. [25]. To corroborate the latter hypothesis, we performed a detailed inspection of specific antigen distribution patterns.

### 3.2. Differential Expression of CD Genes in AML Subtypes

To determine the CD genes that most accurately discriminate the subgroups of AML (APL, AML1, AML2) in the microarray analysis, we determined the differences in mean expression values for each dichotomy by calculating Cohen's D and then performed computed ABC analyses. In total, 23 genes significantly discriminated two AML subgroups in at least one dichotomy in the microarray dataset (Table 2). Moreover, 10 genes were differentially expressed in APL vs. AML1, and 19 genes in APL vs. AML2, with six co-regulated genes in APL vs. both groups. Only one gene (CD18) was differentially expressed in all three pairs of AML subgroups. Hence, a comparison of APL vs. a mixture of both AML groups would have underscored the expression of considerable many potential target genes.

**Table 2.** The most important differentially expressed CD genes in AML subtypes in the microarray dataset. Cohen's D values are indicated for each dichotomous comparison (group A on the left vs. group B on the right). Positive values indicate higher expression in group A. Negative values indicate higher expression in group B. Values < 0.7 were set to 0.

	Gene	APL vs. AML1	APL vs. AML2	AML1 vs. AML2
1	CD3D	0.9	0.92	0
2	CD9	0.89	0	0
3	CD11b	0	−1.17	−0.94
4	CD14	0	−0.99	0
5	CD15	0	−0.77	0
6	CD18	−0.97	−1.69	−0.72
7	CD24	0	−1.29	−0.79
8	CD37	0	−1.09	0
9	CD50	0	−0.98	0
10	CD52	−0.96	−1.25	0
11	CD55	0.78	0	0
12	CD62L	−1.04	−1.44	0
13	CD66b	0	−1.49	−0.82
14	CD66c	0	−1.06	0
15	CD83	−0.7	−1.4	0
16	CD87	1.08	0	0
17	CD88	0	−1.1	−1.09
18	CD98	0.93	1.05	0
19	CD210A	0	−1	−0.79
20	CD282	0	−0.81	−0.75
21	CD339	0.9	0.75	0
22	CD354	0	0	−0.75
23	CD369	0	−1.04	−0.94

Of six genes that were correspondingly differentially expressed between APL and each AML1 and AML2, three (CD3D, CD98, CD339) were higher expressed in APL, and three (CD52, CD62L, CD83) were higher in AML1/2. In addition, the CD18 gene was also higher in APL vs. AML1/2, and differentially expressed in all three pairs of AML subgroups. Finally, one gene (CD354) discriminated AML1 from AML2, but not AML1 and AML2 from APL.

In the RNA-Seq dataset, 14 genes significantly discriminated between AML subgroups in at least one dichotomy (Table 3). Three genes (CD84, CD148, CD339) separated APL from each AML1 and AML2, with all of these genes showing higher expression in APL. In addition, three genes (CD91, CD197, CD261) discriminated between AML1 and AML2. However, CD197 did not discriminate in the microarray data. Of note, CD84, CD148, CD197, and CD261, as well as all other genes that significantly discriminated between two AML subgroups only in one dichotomy (APL vs. AML1, APL vs. AML2, AML1 vs. AML2), were only found in one dataset and showed a Cohen's D-value of 0 in the other and CD91 was not expressed in the microarray dataset (Table S1). Thus, overexpression of CD339 in APL was the only consistent finding in the two datasets.

**Table 3.** The most important differentially expressed CD genes in AML subtypes in the RNA-Seq dataset. Cohen's D values are indicated for each dichotomous comparison (group A vs. group B). Positive values indicate higher expression in group A. Negative values indicate higher expression in group B. Values < 0.3 were set to 0.

	Gene	APL vs. AML1	APL vs. AML2	AML1 vs. AML2
1	CD1D	0	0.42	0
2	CD7	0	0.52	0
3	CD13	0	0.39	0
4	CD44	0	0.53	0
5	CD79A	−0.39	0	0
6	CD84	0.32	0.32	0
7	CD88	0	0.34	0
8	CD91	0	0	−0.32
9	CD148	0.88	0.89	0
10	CD197	0	0	0.32
11	CD227	0	−0.36	0
12	CD230	0	0.32	0
13	CD261	0	0.55	0.52
14	CD339	0.45	0.42	0

Together, the comparison of APL vs. AML2 revealed a diverse expression of well-known differentiation marker molecules, such as CD11b, CD13, CD14, CD15, and CD24, whereas fewer surface molecules were differentially expressed between APL and AML1. The minor expression of specialized functional molecules inside the cell and on the surface is a feature of “stemness”, hereby indicating that AML2 might represent a more differentiated type of AML, and that AML1 is more stem-cell-like. Moreover, only CD79a was overexpressed in AML1 vs. APL. CD79a is considered a B-cell antigen, but is often expressed in immature blast crisis acute leukemias following CML (chronic myeloid leukemia) [26].

### 3.3. Differential Expression of CD Genes in AML Compared to Normal Samples

To further substantiate the differential expression of specific CD genes in AML1, AML2, and APL, we compared these entities to normal samples separately. The results are presented in Table 4. As TCGA RNA-Seq data did not include healthy/normal controls, these

analyses were performed only in the microarray dataset. ABC analyses revealed differential regulation between at least one AML subgroup and normal controls of 48 CD genes.

Two genes (CD99, CD135) were upregulated in each AML subgroup compared to healthy/normal samples, and five genes (CD24, CD66b, CD66c, CD218b, CD233) were expressed significantly lower. One gene (CD88) showed opposing differential regulation in APL and AML1 compared to AML2 relative to healthy bone marrow. Three genes were differentially expressed only between APL and normal, with two genes (CD3D, CD339) being upregulated and one gene (CD50) being downregulated. Furthermore, 20 genes were expressed higher and 6 genes lower in AML2, but no gene was discriminative for AML1 alone or AML1/2 compared to normal (Table 3). Four genes (CD99, CD135, CD218b, and CD233) separated any of the AML subgroups from normal bone marrow without discriminating between them. On the other hand, 21 genes (CD3D, CD9, CD11b, CD14, CD18, CD24, CD37, CD50, CD52, CD62L, CD66b, CD66c, CD83, CD87, CD88, CD98, CD210A, CD282, CD339, CD354, and CD369) were differentially expressed between both at least two different AML subgroups and one AML subgroup compared to normal samples (Tables 2 and 3). Thus, ABC analysis using normal samples as a reference confirmed the differential expression of CD genes in different subgroups of AML.

### 3.4. Validation of Target Genes by Literature Screening

In order to corroborate the target CD antigens with a focus on the distinction between APL and np-AML that were derived from both microarray and RNA-Seq datasets, we screened the literature for reports from other AML groups [Table 5].

In sum, 36 genes (14 in the RNA-Seq, 23 in the microarray dataset) were expressed differentially in APL and np-AML (Table 5). Overexpression of CD339/JAG1 in APL was the only consistent finding in the two datasets, in line with previous reports including both gene expression data from RNA-Seq and microarray analyses [27] or flow cytometry studies [28]. Of note, differential expression in APL and other AML subtypes cannot be deduced from these earlier studies for all CD genes identified by our calculations, such as CD84 [29], CD88 [30], CD148 [23,31], CD210A/IL10RA, and CD227 [32,33]. However, all of these genes have been detected in cells from all hematopoietic lineages or AML blasts, and CD148 and CD210 have even been recently suggested as novel immunotherapeutic targets in AML [34], except for CD261/TNFRSF10A which we did not find in previous publications pointing to implications in AML or APL.

Several genes that discriminated AML or one of its major subgroups from controls are known diagnostic markers and/or targets for investigational therapies in AML, for example CD11b, CD13, CD14, CD24, CD33 [35,36], CD71 [37], CD81 [38] and CD99 [39], while CD132 has not been investigated concerning its potential role in AML.

Therefore, current knowledge of CD gene expression in distinct subgroups of AML strongly supports the validity of our mathematical approach to extract clinically important information from gene expression data, specifically in the context of AML, for diagnostic purposes.

**Table 4.** The most important differentially expressed CD genes between AML subtypes and normal controls in the microarray dataset. Cohen's D values are indicated for each dichotomous comparison (group A/left vs. group B/right). Positive values indicate higher expression in group A. Negative values indicate higher expression in group B. Values < 0.7 were set to 0.

	Gene	APL vs. Normal	AML1 vs. Normal	AML2 vs. Normal
1	CD1D	0	0	0.39
2	CD3D	0.72	0	0
3	CD9	1	0	0.31
4	CD11b	−1.24	−1	0
5	CD13	0	0	0.33
6	CD14	0	0	0.49



Table 4. Cont.

	Gene	APL vs. Normal	AML1 vs. Normal	AML2 vs. Normal
7	CD18	−1.18	0	0.52
8	CD24	−1.83	−1.33	−0.54
9	CD32	0	0	0.34
10	CD33	0	0	0.46
11	CD37	0	0	0.97
12	CD44	0	0	0.45
13	CD50	−0.73	0	0
14	CD52	−0.86	0	0.39
15	CD62L	−0.98	0	0.46
16	CD66b	−1.91	−1.24	−0.42
17	CD66c	−1.8	−1.43	−0.74
18	CD71	0	0	−0.44
19	CD81	0.82	0	0.53
20	CD83	−0.84	0	0.56
21	CD87	0.8	0	0.4
22	CD88	−0.8	−0.79	0.3
23	CD97	0	0	0.73
24	CD98	1.01	0	0
25	CD99	1.44	1.14	1
26	CD107b	0	0	0.41
27	CD114	0	0	0.55
28	CD119	0	0	0.39
29	CD120b	0	0	0.5
30	CD132	0.72	0	0.35
31	CD135	0.7	0.82	0.6
32	CD157	0	0	0.37
33	CD210A	0	0	0.68
34	CD218b	−0.79	−0.7	−0.52
35	CD225	0	0	−0.39
36	CD230	0	0	0.36
37	CD233	−1.29	−0.83	−1.09
38	CD235a	−0.87	0	−0.81
39	CD235b	0	0	−0.53
40	CD236	0	0	−0.41
41	CD240D	0	0	−0.44
42	CD241	0	0	−0.31
43	CD256	0	0	0.4
44	CD257	0	0	0.53
45	CD282	0	0	0.68
46	CD339	0.92	0	0
47	CD354	0	0	0.75
48	CD369	0	0	0.4

**Table 5.** Supporting evidence from the literature for differential expression of CD genes in APL and other AML subtypes. If differential expression could not be deduced from previous reports, references describing expression in AML are indicated.

Index in Table 2/Table 3	Gene	Evidence
21,14	CD339	[27,28]
1	CD3D	[40]
2	CD9	[41–44]
3	CD11b	[41–44]
4	CD14	[19,43]
5	CD15	[45,46]
6	CD18	[47]
7	CD24	[48]
8	CD37	[49]
9	CD50	[50]
10	CD52	[51]
11	CD55	[52]
12	CD62L	[53]
13	CD66b	[54]
14	CD66c	[54]
15	CD83	[55,56]
16	CD87	[57]
17	CD88	[30]
18	CD98	[58,59]
19	CD210A	[60,61]
20	CD282	[57]
23	CD369	[62]
1	CD1D	[63]
2	CD7	[64,65]
3	CD13	[45,46]
4	CD44	[60,66]
5	CD79A	[67,68]
6	CD84	[29,69]
7	CD88	[30]
9	CD148	[31]
11	CD227	[24,32,33]
12	CD230	[70]
13	CD261	x

#### 4. Discussion

Characteristic expression patterns of cell surface molecules provide the basis for rapid diagnosis of AML by flow cytometry [9]. Here, we used combined Bayesian and ABC analysis to identify subgroups of AML based on characteristic cell surface patterns reconstructed from gene expression data for CD genes. We explored associations of AML CD subgroups with an established morphologic classification system for AML, precisely,

FAB classification. Moreover, we extracted the most important discriminating CD genes that distinguish the AML subgroups from each other and normal samples. As the focus of our work was more closely related to potential diagnostic applications, the following discussion primarily refers to current experimentally confirmed knowledge on CD gene expression in AML. Yet, clinical implications are clearly emerging as a future perspective, given the growing spectrum of monoclonal antibodies targeting AML and related myeloid diseases [71].

Intriguingly, using two independent AML expression datasets obtained by microarray or RNA-Seq analyses, we found that AML separated into only two natural clusters in both datasets based on the expression of CD genes. In the RNA-Seq dataset, we identified two subgroups of AML and were able to separate APL from AML. One of these clusters is APL, the only subclass of AML that has been previously defined, and is managed therapeutically differently from all other AML subclasses in the clinic [72]. Np-AML, on the other hand, is less diverse based on CD gene expression than, for example, based on morphology or genetic findings. Employing the Ward algorithm on 394 CD genes, AML clustered into two subgroups, which we here pragmatically designated AML1 and AML 2. In the RNA-Seq dataset, AML1 was associated with FAB classes M1 and M2, while AML2 corresponded to FAB class M5. Both AML subgroups contained FAB class M4. In the microarray dataset, the FAB classification was not accessible. This association supports the notion that the AML1 cluster refers to an immature immunophenotype, while the AML2 cluster characterizes leukemia cells with the expression of maturation markers, i.e., AML1 cells may derive from a pluripotent stem cell and AML2 from a progenitor cell, respectively.

In order to further investigate whether the classification of AML by CD gene expression may be exploited for AML diagnosis by flow cytometry in addition to previously recommended immunophenotype targets, we selected the most important discriminating genes for each subgroup compared to the other subgroups and to normal controls. Performing ABC analysis on the microarray and the RNA-Seq datasets predominantly yielded non-overlapping results. Only the CD339 gene, which corresponds to the Jagged-1 protein, was expressed higher in APL than in AML1 and AML2 in both datasets, and was also distinctive for APL compared to normal samples, which is in line with previous analyses of publicly available gene expression datasets [27]. However, CD339 is not included in previously reported antibody panels for the identification of APL by flow cytometry [41,42,73]. Therefore, it is tempting to speculate that our unbiased analysis of the AML surface can inform the design of improved diagnostic tools.

On the other hand, the only other gene that was somehow discriminating between AML subgroups, CD88, was underexpressed in AML2 compared to APL in the RNA-Seq dataset but overexpressed relative to both APL and AML1 in the microarray dataset. CD88 (C5AR1) has been reported to be expressed in AML cell lines and primary patient blasts and to contribute to cell motility [30], illustrating its potential role in this disease. We attribute the conflicting results for CD88 expression, as well as the overall lack of consistent observations, to the different molecular techniques for gene expression measurement [74], as Bayesian analysis based on GMMs has been shown to yield stable and consistent results on noisy data of various types, such as income, heat or cold pain, GDP time series or nitrate concentrations in streams [15,75,76]. Of note, the dataset of only 15 cases of APL samples is small, but in line with the reported frequency of this subgroup. Consequently, consolidation with a larger dataset would be desirable to strengthen the results reported here. Of non-overlapping genes that were differentially expressed between AML subgroups or AML and normal samples, several CDs are either known targets (such as CD33) or have already been reported to be implicated in AML.

Novel treatment options for AML cases with an unfavorable prognosis are highly demanded and monoclonal antibodies that target AML surface molecules are currently evaluated in clinical trials ([71,77]). Although we do not have detailed information on cytogenetics or survival in both of our datasets, our study reveals potential target molecules, particularly if AML1 is indeed a stem cell near a subgroup with inferior prognosis.

In summary, we conclude that our unbiased mathematical approach was highly successful in deducing knowledge from complex biological datasets. “The important few” AML-specific CD genes identified by combined Bayesian and calculated ABC analysis are widely congruent with actual knowledge in the field, independent from the particular method of gene expression analysis. Thus, the work described here may serve as a blueprint for translating relevant information from high-dimensional biological data into clinical applications.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/bioengineering9110642/s1>, Table S1: Overview of all significant differentially expressed CD genes in both datasets.

**Author Contributions:** Conceptualization, A.U. and C.B.; Data curation, T.H. and M.F.; Formal analysis, M.C.T., E.K.M.M. and C.B.; Investigation, M.C.T., E.K.M.M. and A.U.; Methodology, M.C.T. and A.U.; Project administration, A.N. and C.B.; Resources, M.F.; Software, M.C.T.; Supervision, A.N.; Validation, E.K.M.M.; Writing—original draft, M.C.T., E.K.M.M. and C.B.; Writing—review & editing, M.C.T., E.K.M.M., A.N., T.H., M.F., A.U. and C.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** E.K.M.M. and A.N. were funded by the Deutsche José-Carreras Leukämie-Stiftung (Grant Numbers AR 07/2017 and AH 06-01 to A.N.). E.K.M.M. was supported by the Clinician Scientist program of the Philipps University of Marburg and the University Hospital of Gießen and Marburg.

**Institutional Review Board Statement:** The ethics board approved analysis of the datasets for the purpose of this study at the faculty of medicine, University of Marburg (No. 138/16).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** RNA-Seq data can be downloaded from the GDC Data Portal (<https://portal.gdc.cancer.gov/>) (accessed on 3 April 2017). In addition, the RNA-seq data can be provided upon request. (accessed on 3 April 2017)). The microarray dataset was published by Haferlach et al. [11].

**Acknowledgments:** The authors thank André Marquardt for help with downloading the TCGA dataset.

**Conflicts of Interest:** The authors declare no conflict of interest related to this work.

## References

1. Kandoth, C.; McLellan, M.D.; Vandin, F.; Ye, K.; Niu, B.; Lu, C.; Xie, M.; Zhang, Q.; McMichael, J.F.; Wyczalkowski, M.A.; et al. Mutational landscape and significance across 12 major cancer types. *Nature* **2013**, *502*, 333–339. [[CrossRef](#)] [[PubMed](#)]
2. Sanchez-Vega, F.; Mina, M.; Armenia, J.; Chatila, W.K.; Luna, A.; La, K.C.; Dimitriadoy, S.; Liu, D.L.; Kantheti, H.S.; Saghafinia, S.; et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **2018**, *173*, 321–337.e310. [[CrossRef](#)] [[PubMed](#)]
3. Garraway, L.A. Genomics-Driven Oncology: Framework for an Emerging Paradigm. *J. Clin. Oncol.* **2013**, *31*, 1806–1814. [[CrossRef](#)] [[PubMed](#)]
4. Dawson, S.-J.; Rueda, O.M.; Aparicio, S.; Caldas, C. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J.* **2013**, *32*, 617–628. [[CrossRef](#)]
5. Papaemmanuil, E.; Gerstung, M.; Bullinger, L.; Gaidzik, V.I.; Paschka, P.; Roberts, N.D.; Potter, N.E.; Heuser, M.; Thol, F.; Bolli, N.; et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* **2016**, *374*, 2209–2221. [[CrossRef](#)]
6. Sanz, M.A.; Fenaux, P.; Tallman, M.S.; Estey, E.H.; Löwenberg, B.; Naoe, T.; Lengfelder, E.; Döhner, H.; Burnett, A.K.; Chen, S.-J.; et al. Management of acute promyelocytic leukemia: Updated recommendations from an expert panel of the European LeukemiaNet. *Blood* **2019**, *133*, 1630–1643. [[CrossRef](#)]
7. Döhner, H.; Estey, E.; Grimwade, D.; Amadori, S.; Appelbaum, F.R.; Büchner, T.; Dombret, H.; Ebert, B.L.; Fenaux, P.; Larson, R.A.; et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **2017**, *129*, 424–447. [[CrossRef](#)]
8. Arber, D.A.; Orazi, A.; Hasserjian, R.; Thiele, J.; Borowitz, M.J.; Le Beau, M.M.; Bloomfield, C.D.; Cazzola, M.; Vardiman, J.W. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **2016**, *127*, 2391–2405. [[CrossRef](#)]
9. Schuurhuis, G.J.; Heuser, M.; Freeman, S.; Béné, M.-C.; Buccisano, F.; Cloos, J.; Grimwade, D.; Haferlach, T.; Hills, R.K.; Hourigan, C.S.; et al. Minimal/measurable residual disease in AML: A consensus document from the European LeukemiaNet MRD Working Party. *Blood* **2018**, *131*, 1275–1291. [[CrossRef](#)]

10. Ultsch, A.; Lotsch, J. Computed ABC Analysis for Rational Selection of Most Informative Variables in Multivariate Data. *PLoS ONE* **2015**, *10*, e0129767. [[CrossRef](#)]
11. Haferlach, T.; Kohlmann, A.; Wiczorek, L.; Basso, G.; Kronnie, G.T.; Béné, M.-C.; De Vos, J.; Hernández, J.M.; Hofmann, W.-K.; Mills, K.I.; et al. Clinical Utility of Microarray-Based Gene Expression Profiling in the Diagnosis and Subclassification of Leukemia: Report from the International Microarray Innovations in Leukemia Study Group. *J. Clin. Oncol.* **2010**, *28*, 2529–2537. [[CrossRef](#)] [[PubMed](#)]
12. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **2013**, *368*, 2059–2074. [[CrossRef](#)] [[PubMed](#)]
13. Döhner, H.; Wei, A.H.; Appelbaum, F.R.; Craddock, C.; DiNardo, C.D.; Dombret, H.; Ebert, B.L.; Fenaux, P.; Godley, L.A.; Hasserjian, R.P.; et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **2022**, *140*, 1345–1377. [[CrossRef](#)] [[PubMed](#)]
14. Thrun, M.C.; Gehlert, T.; Ultsch, A. Analyzing the fine structure of distributions. *PLoS ONE* **2020**, *15*, e0238835. [[CrossRef](#)] [[PubMed](#)]
15. Ultsch, A.; Thrun, M.C.; Hansen-Goos, O.; Lötsch, J. Identification of Molecular Fingerprints in Human Heat Pain Thresholds by Use of an Interactive Mixture Model R Toolbox (AdaptGauss). *Int. J. Mol. Sci.* **2015**, *16*, 25897–25911. [[CrossRef](#)]
16. Fraley, C.; E Raftery, A. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611–631. [[CrossRef](#)]
17. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
18. Smirnov, N. Table for Estimating the Goodness of Fit of Empirical Distributions. *Ann. Math. Stat.* **1948**, *19*, 279–281. [[CrossRef](#)]
19. Thrun, M.C.; Stier, Q. Fundamental clustering algorithms suite. *SoftwareX* **2021**, *13*, 100642. [[CrossRef](#)]
20. Thrun, M.C. *Projection Based Clustering through Self-Organization and Swarm Intelligence*; Springer: Heidelberg, Germany, 2018.
21. Hedges, L.V.; Olkin, I. *Statistical Methods for Meta-Analysis*; Academic Press: New York, NY, USA, 1985.
22. Bennett, J.M.; Catovsky, D.; Daniel, M.-T.; Flandrin, G.; Galton, D.A.G.; Gralnick, H.R.; Sultan, C. Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. *Br. J. Haematol.* **1976**, *33*, 451–458. [[CrossRef](#)]
23. Haferlach, T.; Kohlmann, A.; Schnittger, S.; Dugas, M.; Hiddemann, W.; Kern, W.; Schoch, C. AML M3 and AML M3 variant each have a distinct gene expression signature but also share patterns different from other genetically defined AML subtypes. *Genes Chromosom. Cancer* **2005**, *43*, 113–127. [[CrossRef](#)] [[PubMed](#)]
24. Schoch, C.; Kohlmann, A.; Schnittger, S.; Brors, B.; Dugas, M.; Mergenthaler, S.; Kern, W.; Hiddemann, W.; Eils, R.; Haferlach, T. Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10008–10013. [[CrossRef](#)] [[PubMed](#)]
25. Goardon, N.; Marchi, E.; Atzberger, A.; Quek, L.; Schuh, A.; Soneji, S.; Woll, P.S.; Mead, A.; Alford, K.A.; Rout, R.; et al. Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia. *Cancer Cell* **2011**, *19*, 138–152. [[CrossRef](#)] [[PubMed](#)]
26. Matutes, E.; Morilla, R.; Farahat, N.; Carbonell, F.; Swansbury, J.; Dyer, M.; Catovsky, D. Definition of acute biphenotypic leukemia. *Haematologica* **1997**, *82*, 64–66. [[PubMed](#)]
27. Grieselhuber, N.; Klco, J.; Verdoni, A.M.; Lamprecht, T.; Sarkaria, S.; Wartman, L.D.; Ley, T.J. Notch signaling in acute promyelocytic leukemia. *Leukemia* **2013**, *27*, 1548–1557. [[CrossRef](#)]
28. Kanamori, E.; Itoh, M.; Tojo, N.; Koyama, T.; Nara, N.; Tohda, S. Flow cytometric analysis of Notch1 and Jagged1 expression in normal blood cells and leukemia cells. *Exp. Ther. Med.* **2012**, *4*, 397–400. [[CrossRef](#)]
29. Zaiss, M.; Hirtreiter, C.; Rehli, M.; Rehm, A.; A Kunz-Schughart, L.; Andreesen, R.; Hennemann, B. CD84 expression on human hematopoietic progenitor cells. *Exp. Hematol.* **2003**, *31*, 798–805. [[CrossRef](#)]
30. Ismail, A.A.; Borkowska-Rzeszotek, S.; Kubiś, E.; Bujko, K.; Brzeźniakiewicz-Janus, K.; Bolkun, L.; Kloczko, J.; Moniuszko, M.; Basak, G.W.; Wiktor-Jedrzejczak, W.; et al. Activation of the complement cascade enhances motility of leukemic cells by downregulating expression of HO-1. *Leukemia* **2016**, *31*, 446–458. [[CrossRef](#)]
31. Kresinsky, A.; Bauer, R.; Schnöder, T.M.; Berg, T.; Meyer, D.; Ast, V.; König, R.; Serve, H.; Heidel, F.H.; Böhmer, F.-D.; et al. Loss of DEP-1 (Ptpn) promotes myeloproliferative disease in FLT3-ITD acute myeloid leukemia. *Haematologica* **2018**, *103*, e505–e509. [[CrossRef](#)]
32. Brossart, P.; Schneider, A.; Dill, P.; Schammann, T.; Grünebach, F.; Wirths, S.; Kanz, L.; Bühring, H.J.; Brugger, W. The epithelial tumor antigen MUC1 is expressed in hematological malignancies and is recognized by MUC1-specific cytotoxic T-lymphocytes. *Cancer Res.* **2001**, *61*, 6846–6850.
33. Fatrai, S.; Schepers, H.; Tadema, H.; Vellenga, E.; Daenen, S.M.; Schuringa, J.J. Mucin1 expression is enriched in the human stem cell fraction of cord blood and is upregulated in majority of the AML cases. *Exp. Hematol.* **2008**, *36*, 1254–1265. [[CrossRef](#)] [[PubMed](#)]
34. Köhnke, T.; Liu, X.; Haubner, S.; Bücklein, V.; Hänel, G.; Krupka, C.; Solis-Mezarino, V.; Herzog, F.; Subklewe, M. Integrated multiomic approach for identification of novel immunotherapeutic targets in AML. *Biomark. Res.* **2022**, *10*, 43. [[CrossRef](#)] [[PubMed](#)]
35. Pollard, J.A.; Loken, M.; Gerbing, R.B.; Raimondi, S.C.; Hirsch, B.A.; Aplenc, R.; Bernstein, I.D.; Gamis, A.S.; Alonzo, T.A.; Meshinchi, S. CD33 Expression and Its Association With Gemtuzumab Ozogamicin Response: Results From the Randomized Phase III Children’s Oncology Group Trial AAML0531. *J. Clin. Oncol.* **2016**, *34*, 747–755. [[CrossRef](#)] [[PubMed](#)]

36. Amadori, S.; Suci, S.; Selleslag, D.; Aversa, F.; Gaidano, G.; Musso, M.; Annino, L.; Venditti, A.; Voso, M.T.; Mazzone, C.; et al. Gemtuzumab Ozogamicin Versus Best Supportive Care in Older Patients with Newly Diagnosed Acute Myeloid Leukemia Unsuitable for Intensive Chemotherapy: Results of the Randomized Phase III EORTC-GIMEMA AML-19 Trial. *J. Clin. Oncol.* **2016**, *34*, 972–979. [[CrossRef](#)] [[PubMed](#)]
37. Liu, Q.; Wang, M.; Hu, Y.; Xing, H.; Chen, X.; Zhang, Y.; Zhu, P. Significance of CD71 expression by flow cytometry in diagnosis of acute leukemia. *Leuk. Lymphoma* **2013**, *55*, 892–898. [[CrossRef](#)]
38. Boyer, T.; Guihard, S.; Roumier, C.; Peyrouze, P.; Gonzales, F.; Berthon, C.; Quesnel, B.; Preudhomme, C.; Behal, H.; Duhamel, A.; et al. Tetraspanin CD81 is an adverse prognostic marker in acute myeloid leukemia. *Oncotarget* **2016**, *7*, 62377–62385. [[CrossRef](#)]
39. Kang, L.C.; Dunphy, C.H. Immunoreactivity of MIC2 (CD99) and Terminal Deoxynucleotidyl Transferase in Bone Marrow Clot and Core Specimens of Acute Myeloid Leukemias and Myelodysplastic Syndromes. *Arch. Pathol. Lab. Med.* **2006**, *130*, 153–157. [[CrossRef](#)]
40. Pereira, M.S.; Silva, M.L.M.; Valente, A.N.; De Souza, M.H.O.; Abdelhay, E. Translocation of CD3D gene in an acute myeloid leukemia (M5) with t(11;17)(q23;21). *Cancer Genet. Cytogenet.* **1993**, *71*, 173–175. [[CrossRef](#)]
41. Ren, F.; Zhang, N.; Xu, Z.; Xu, J.; Zhang, Y.; Chen, X.; Tan, Y.; Chang, J.; Wang, H. The CD9<sup>+</sup>CD11b<sup>-</sup>HLA-DR<sup>-</sup> immunophenotype can be used to diagnose acute promyelocytic leukemia. *Int. J. Lab. Hematol.* **2018**, *41*, 168–175. [[CrossRef](#)]
42. Dong, H.Y.; Kung, J.X.; Bhardwaj, V.; McGill, J. Flow Cytometry Rapidly Identifies All Acute Promyelocytic Leukemias With High Specificity Independent of Underlying Cytogenetic Abnormalities. *Am. J. Clin. Pathol.* **2011**, *135*, 76–84. [[CrossRef](#)]
43. Chen, Z.; Li, Y.; Tong, Y.; Gao, Q.; Mao, X.; Zhang, W.; Xia, Z.; Fu, C. Stepwise discriminant function analysis for rapid identification of acute promyelocytic leukemia from acute myeloid leukemia with multiparameter flow cytometry. *Int. J. Hematol.* **2016**, *103*, 306–315. [[CrossRef](#)] [[PubMed](#)]
44. Ferrero, D.; Carlesso, N.; Gallo, E.; Pregno, P.; de Fabritiis, P.; Petti, M.C.; Mandelli, F. CD9 antigen on acute non-lymphoid leukemia cells: Preferential expression by promyelocytic (M3) subtype. *Leuk. Res.* **1991**, *15*, 457–461. [[CrossRef](#)]
45. Kárai, B.; Habók, M.; Reményi, G.; Rejtő, L.; Ujfalusi, A.; Kappelmayer, J.; Hevessy, Z. A novel flow cytometric method for enhancing acute promyelocytic leukemia screening by multidimensional dot-plots. *Ann. Hematol.* **2019**, *98*, 1413–1420. [[CrossRef](#)]
46. Orfao, A.; Chillón, M.C.; Bortoluci, A.M.; López-Berges, M.C.; García-Sanz, R.; Gonzalez, M.; Taberero, M.D.; García-Marcos, M.A.; Rasillo, A.I.; Hernández-Rivas, J. The flow cytometric pattern of CD34, CD15 and CD13 expression in acute myeloblastic leukemia is highly characteristic of the presence of PML-RARalpha gene rearrangements. *Haematologica* **1999**, *84*, 405–412. [[PubMed](#)]
47. Zhou, Y.; Jorgensen, J.L.; Wang, S.A.; Ravandi, F.; Cortes, J.; Kantarjian, H.M.; Medeiros, L.J.; Konoplev, S. Usefulness of CD11a and CD18 in Flow Cytometric Immunophenotypic Analysis for Diagnosis of Acute Promyelocytic Leukemia. *Am. J. Clin. Pathol.* **2012**, *138*, 744–750. [[CrossRef](#)] [[PubMed](#)]
48. Raife, T.J.; Lager, D.J.; Kemp, J.D.; Dick, F.R. Expression of CD24 (BA-1) Predicts Monocytic Lineage in Acute Myeloid Leukemia. *Am. J. Clin. Pathol.* **1994**, *101*, 296–299. [[CrossRef](#)] [[PubMed](#)]
49. Kamashev, D.; Vitoux, D.; De Thé, H. PML–RARA–RXR Oligomers Mediate Retinoid and Reginoid/cAMP Cross-Talk in Acute Promyelocytic Leukemia Cell Differentiation. *J. Exp. Med.* **2004**, *199*, 1163–1174. [[CrossRef](#)]
50. Khaznadar, Z.; Boissel, N.; Agaugué, S.; Henry, G.; Cheok, M.; Vignon, M.; Geromin, D.; Cayuela, J.-M.; Castaigne, S.; Pautas, C.; et al. Defective NK Cells in Acute Myeloid Leukemia Patients at Diagnosis Are Associated with Blast Transcriptional Signatures of Immune Evasion. *J. Immunol.* **2015**, *195*, 2580–2590. [[CrossRef](#)]
51. Li, S.-W.; Tang, D.; Ahrens, K.P.; She, J.-X.; Braylan, R.C.; Yang, L. All-trans-retinoic acid induces CD52 expression in acute promyelocytic leukemia. *Blood* **2003**, *101*, 1977–1980. [[CrossRef](#)]
52. Guc, D.; Canpinar, H.; Kucukaksu, C.; Kansu, E. Expression of complement regulatory proteins CR1, DAF, MCP and CD59 in haematological malignancies. *Eur. J. Haematol.* **2000**, *64*, 3–9. [[CrossRef](#)]
53. Jambrovics, K.; Uray, I.P.; Keresztessy, Z.; Keillor, J.W.; Fésüs, L.; Balajthy, Z. Transglutaminase 2 programs differentiating acute promyelocytic leukemia cells in all-trans retinoic acid treatment to inflammatory stage through NF-κB activation. *Haematologica* **2019**, *104*, 505–515. [[CrossRef](#)] [[PubMed](#)]
54. Ismail, M.; Zaghloul, A.; Abdulateef, N.; Morsi, H. Membranous Expression of pan CD66, CD66a, CD66b, and CD66c and their Clinical Impact in Acute Leukemia: Cross Sectional Longitudinal Cohort Study in Saudi Arabia. *J. Leuk.* **2017**, *5*, 230–239. [[CrossRef](#)]
55. Shrestha, B.; Walton, K.; Reff, J.; Sagatys, E.M.; Tu, N.; Boucher, J.C.; Ghafoor, B.T.; Pidala, J.; Blazar, B.R.; Anasetti, C.; et al. Human CD83 Targeted Chimeric Antigen Receptor T Cell for the Prevention of Graft Versus Host Disease and Treatment of Myeloid Leukemia. *Blood* **2019**, *134*, 196. [[CrossRef](#)]
56. Rigolin, G.M.; Della Porta, M.; Bigoni, R.; Tieghi, A.; Cuneo, A.; Castoldi, G. Dendritic cells in acute promyelocytic leukaemia. *Br. J. Haematol.* **2001**, *114*, 830–833. [[CrossRef](#)]
57. Atfy, M.; Eissa, M.; Salah, H.E.; El Shabrawy, D.A. Role of urokinase plasminogen activator receptor (CD87) as a prognostic marker in acute myeloid leukemia. *Med. Oncol.* **2011**, *29*, 2063–2069. [[CrossRef](#)]
58. Bajaj, J.; Konuma, T.; Lytle, N.K.; Kwon, H.Y.; Ablack, J.N.; Cantor, J.M.; Rizzieri, D.; Chuah, C.; Oehler, V.G.; Broome, E.H.; et al. CD98-Mediated Adhesive Signaling Enables the Establishment and Propagation of Acute Myelogenous Leukemia. *Cancer Cell* **2016**, *30*, 792–805. [[CrossRef](#)]

59. Bixby, D.; Wieduwilt, M.J.; Akard, L.P.; Khoury, H.J.; Becker, P.S.; Van Der Horst, E.H.; Ho, W.; Cortes, J.E. A Phase I Study of IGN523, a Novel Anti-CD98 Monoclonal Antibody in Patients with Relapsed or Refractory Acute Myeloid Leukemia (AML). *Blood* **2015**, *126*, 3809. [[CrossRef](#)]
60. Coustan-Smith, E.; Song, G.; Shurtleff, S.; Yeoh, A.E.-J.; Chng, W.J.; Chen, S.P.; Rubnitz, J.E.; Pui, C.-H.; Downing, J.R.; Campana, D. Universal monitoring of minimal residual disease in acute myeloid leukemia. *JCI Insight* **2018**, *3*, 98561. [[CrossRef](#)]
61. Chen, N.; Xu, Y.; Mou, J.; Rao, Q.; Xing, H.; Tian, Z.; Tang, K.; Wang, M.; Wang, J. Targeting of IL-10R on acute myeloid leukemia blasts with chimeric antigen receptor-expressing T cells. *Blood Cancer J.* **2021**, *11*, 144. [[CrossRef](#)]
62. Zhang, H.; Nakauchi, Y.; Köhnke, T.; Stafford, M.; Bottomly, D.; Thomas, R.; Wilmot, B.; McWeeney, S.K.; Majeti, R.; Tyner, J.W. Integrated analysis of patient samples identifies biomarkers for venetoclax efficacy and combination strategies in acute myeloid leukemia. *Nat. Cancer* **2020**, *1*, 826–839. [[CrossRef](#)]
63. Metelitsa, L.S.; I Weinberg, K.; Emanuel, P.; Seeger, R.C. Expression of CD1d by myelomonocytic leukemias provides a target for cytotoxic NKT cells. *Leukemia* **2003**, *17*, 1068–1077. [[CrossRef](#)] [[PubMed](#)]
64. Sobas, M.; Montesinos, P.; Boluda, B.; Bernal, T.; Vellenga, E.; Nomdedeu, J.; González-Campos, J.; Chillón, M.; Holowiecka, A.; Esteve, J.; et al. An analysis of the impact of CD56 expression in de novo acute promyelocytic leukemia patients treated with upfront all-trans retinoic acid and anthracycline-based regimens. *Leuk. Lymphoma* **2018**, *60*, 1030–1035. [[CrossRef](#)]
65. Del Poeta, G.; Stasi, R.; Venditti, A.; Cox, C.; Aronica, G.; Masi, M.; Bruno, A.; Simone, M.D.; Buccisano, F.; Papa, G.; et al. CD7 Expression in Acute Myeloid Leukemia. *Leuk. Lymphoma* **1995**, *17*, 111–119. [[CrossRef](#)] [[PubMed](#)]
66. Legras, S.; Günther, U.; Stauder, R.; Curt, F.; Oliferenko, S.; Kluin-Nelemans, H.; Marie, J.; Proctor, S.; Jasmin, C.; Smadja-Joffe, F. A Strong Expression of CD44-6v Correlates With Shorter Survival of Patients With Acute Myeloid Leukemia. *Blood* **1998**, *91*, 3401–3413. [[CrossRef](#)] [[PubMed](#)]
67. Bhargava, P.; Kallakury, B.V.; Ross, J.S.; Azumi, N.; Bagg, A. CD79a Is Heterogeneously Expressed in Neoplastic and Normal Myeloid Precursors and Megakaryocytes in an Antibody Clone-Dependent Manner. *Am. J. Clin. Pathol.* **2007**, *128*, 306–313. [[CrossRef](#)] [[PubMed](#)]
68. Arber, D.A.; A Jenkins, K.; Slovak, M.L. CD79 alpha expression in acute myeloid leukemia. High frequency of expression in acute promyelocytic leukemia. *Am. J. Pathol.* **1996**, *149*, 1105–1110.
69. Çelik, H.; Lindblad, K.E.; Popescu, B.; Gui, G.; Goswami, M.; Valdez, J.; DeStefano, C.; Lai, C.; Thompson, J.; Ghannam, J.Y.; et al. Highly multiplexed proteomic assessment of human bone marrow in acute myeloid leukemia. *Blood Adv.* **2020**, *4*, 367–379. [[CrossRef](#)]
70. Leung, K.K.; Nguyen, A.; Shi, T.; Tang, L.; Ni, X.; Escoubet, L.; MacBeth, K.J.; DiMartino, J.; Wells, J.A. Multiomics of azacitidine-treated AML cells reveals variable and convergent targets that remodel the cell-surface proteome. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 695–700. [[CrossRef](#)]
71. Gallazzi, M.; Ucciero, M.A.M.; Faraci, D.G.; Mahmoud, A.M.; Al Essa, W.; Gaidano, G.; Mouhssine, S.; Crisà, E. New Frontiers in Monoclonal Antibodies for the Targeted Therapy of Acute Myeloid Leukemia and Myelodysplastic Syndromes. *Int. J. Mol. Sci.* **2022**, *23*, 7542. [[CrossRef](#)]
72. Lo-Coco, F.; Avvisati, G.; Vignetti, M.; Thiede, C.; Orlando, S.M.; Iacobelli, S.; Ferrara, F.; Fazi, P.; Cicconi, L.; Di Bona, E.; et al. Retinoic Acid and Arsenic Trioxide for Acute Promyelocytic Leukemia. *N. Engl. J. Med.* **2013**, *369*, 111–121. [[CrossRef](#)]
73. Ma, D.; Zhong, S.; Liu, X.; Mai, H.; Mai, G.; Xu, C.; Zhou, F. CD3D and PRKCQ work together to discriminate between B-cell and T-cell acute lymphoblastic leukemia. *Comput. Biol. Med.* **2016**, *77*, 16–22. [[CrossRef](#)] [[PubMed](#)]
74. Guo, Y.; Sheng, Q.; Li, J.; Ye, F.; Samuels, D.C.; Shyr, Y. Large Scale Comparison of Gene Expression Levels by Microarrays and RNAseq Using TCGA Data. *PLoS ONE* **2013**, *8*, e71462. [[CrossRef](#)] [[PubMed](#)]
75. Thrun, M.C. Cluster Analysis of Per Capita Gross Domestic Products. *Entrep. Bus. Econ. Rev.* **2018**, *7*, 217–231. [[CrossRef](#)]
76. Thrun, M.; Ultsch, A. Models of Income Distributions for Knowledge Discovery. In *European Conference on Data Analysis*; University of Essex: Colchester, UK, 2015; pp. 136–137.
77. Gbadamosi, M.O.; Shastri, V.M.; Elsayed, A.H.; Ries, R.; Olabige, O.; Nguyen, N.H.K.; De Jesus, A.; Wang, Y.-C.; Dang, A.; Hirsch, B.A.; et al. A ten-gene DNA-damage response pathway gene expression signature predicts gemtuzumab ozogamicin response in pediatric AML patients treated on COGAAML0531 and AAML03P1 trials. *Leukemia* **2022**, *36*, 2022–2031. [[CrossRef](#)] [[PubMed](#)]