Unraveling new markers for prediction of *KMT2A (MLL)* rearrangements

Bruno A. Lopes[1,2], Claus Meyer[3], Rolf Marschalek[3], Mariana Emerenciano[1]

[1]Molecular Cancer Study Group, Division of Clinical Research, Research Center, Instituto Nacional de Câncer, Rio de Janeiro, Brazil

[2]Laboratory of Immunology and Molecular Biology, Institute of Health Sciences, Universidade Federal da Bahia, Salvador, Brazil

[3]Diagnostic Center of Acute Leukemia, Institute of Pharmaceutical Biology, Goethe-University of Frankfurt, Frankfurt/Main, Germany

INTRODUCTION

Chromosomal rearrangements involving the human *MLL/KMT2A* gene are recurrently associated with the disease phenotype of acute leukemia. The presence of distinct *MLL* rearrangements (*MLL-r*) is an independent dismal prognostic factor. The identification of *MLL-r* is routinely based on two main strategies: split-signal fluorescence in situ hybridization (FISH) and RT-PCR. Recently, we stablished a custom NGS panel for the detection of *MLL-r* at nucleotide level. This allowed us to identify a novel breakpoint cluster region within *MLL*, recurrently fused with *USP2* gene. Considering that *MLL-USP2* fusions are generated by a 0.8 Kb inversion within 11q23, we demonstrated that FISH screening fails to detect this alteration in most of the patients. Therefore, immunophenotyping could be an important method for fast screening of *MLL-r*. Although NG2 marker has been associated with *MLL-r*, it has a varying amount of false negative results. Here we evaluate the transcriptome of acute leukemia samples to analyze differential expression and unravel an appropriate marker for diagnostic routine prediction of *MLL-r*.

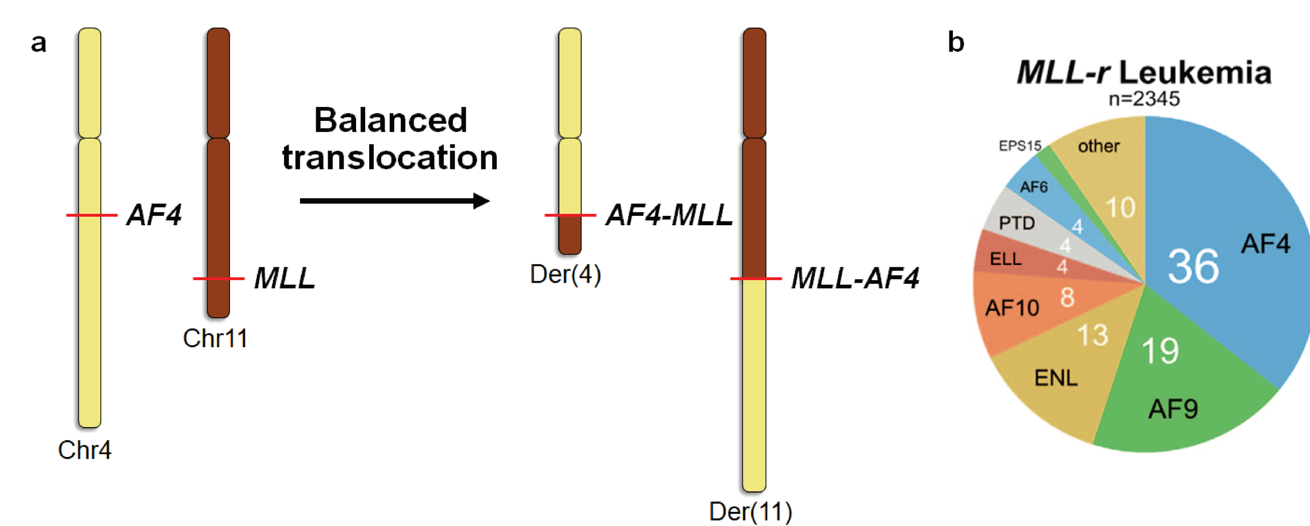

Figure 1. Chromosome 11 alterations leading to *MLL* rearrangements. (A) Schematic representation of t(4;11) and derivative chromosomes with direct (*MLL-AF4*) and reciprocal fusions (*AF4-MLL*); (B) The pie chart reveals the most frequent *MLL* partner genes in overall leukemia (Meyer et al. Leukemia 2017).

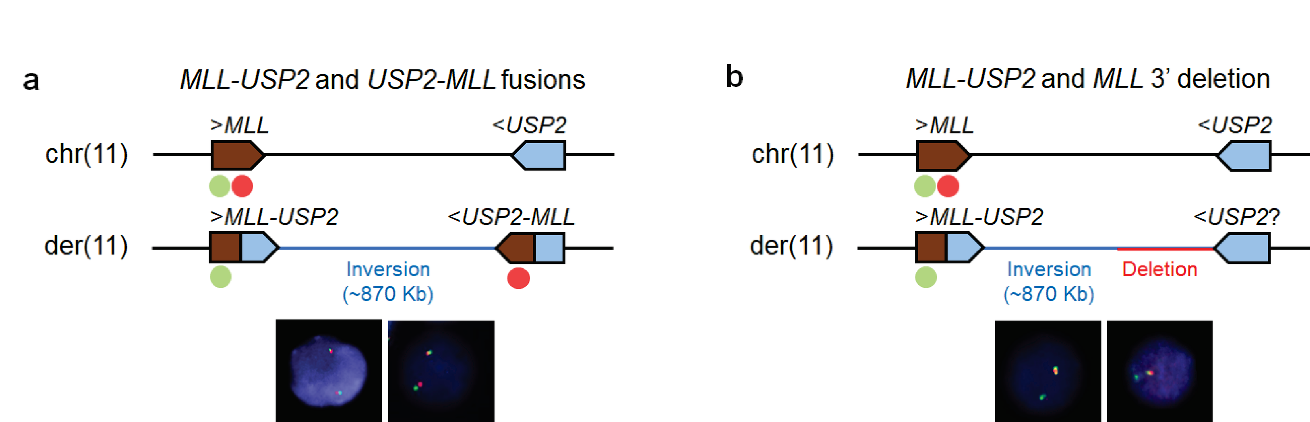

Figure 2. FISH limitation for the assessment of *MLL* status. (a) *MLL-USP2* fusions derive from a small inversion between *MLL* and *USP2*, which sometimes are not evidenced by FISH. However, (b) deletion of *MLL 3'* can also occur in *MLL-USP2* fusions, and the lack of one 3' probe (red) was observed (Meyer and Lopes et al. Leukemia 2019).

METHODS

The TARGET database was used for the identification of a predictive marker for the *MLL-r*. Clinical and RNA-seq data of B-ALL, AML, and MPAL were retrieved from TARGET. The analyses were performed in R statistical environment. The discovery of a predictive marker for *MLL-r* in leukemia was based on differential expression analysis using DESeq. The analysis was performed in two cohorts – B-ALL (ALL Phase II) and AML – of TARGET project. After the determination of candidate genes for prediction of *MLL-r*, their transcript levels were compared among *MLL* status groups in B-ALL, AML, and MPAL. The molecular subtype of B-ALL was also evaluated.

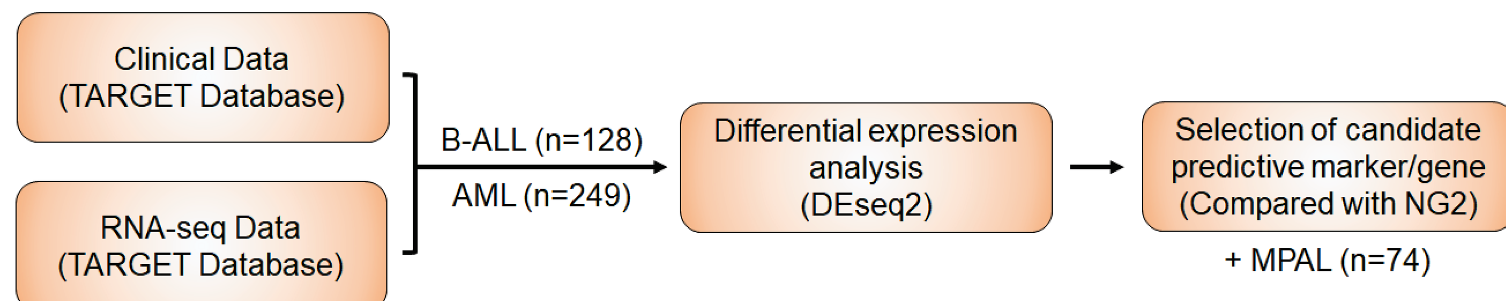

Figure 3. Study flowchart.

RESULTS

A total of 128 B-ALL and 249 AML patients were included, and 52,625 genes were evaluated. A total of 100 and 62 genes were upregulated in B-ALL and AML, respectively [log2(fold-change) > 2, and P>0.01]. *SKIDA1* gene was the most significantly upregulated gene (P=6.4x10^{-7}), while *CSPG4* (coding for NG2) was ranked the 16th position. Among those sixteen most upregulated genes, SKIDA1 also had the highest fold change difference in gene expression while considering both ALL and AML subtypes. In the next step of our analysis, we compared the expression of the top sixteen upregulated genes based on *MLL* status in three subtypes of acute leukemia. This analysis included MPAL, which presented high incidence of *MLL-USP2* fusions. The *CSPG4* presents good performance for estimating *MLL-r*, however, it was not so accurate for discriminating such alterations in patients diagnosed with B-ALL (P=0.058), AML (P=1.8x10^{-7}), and MPAL (P=1.0x10^{-5}) as compared to *SKIDA1* (P=0.007, 1.9x10^{-11}, and 8.0x10^{-6}, respectively).

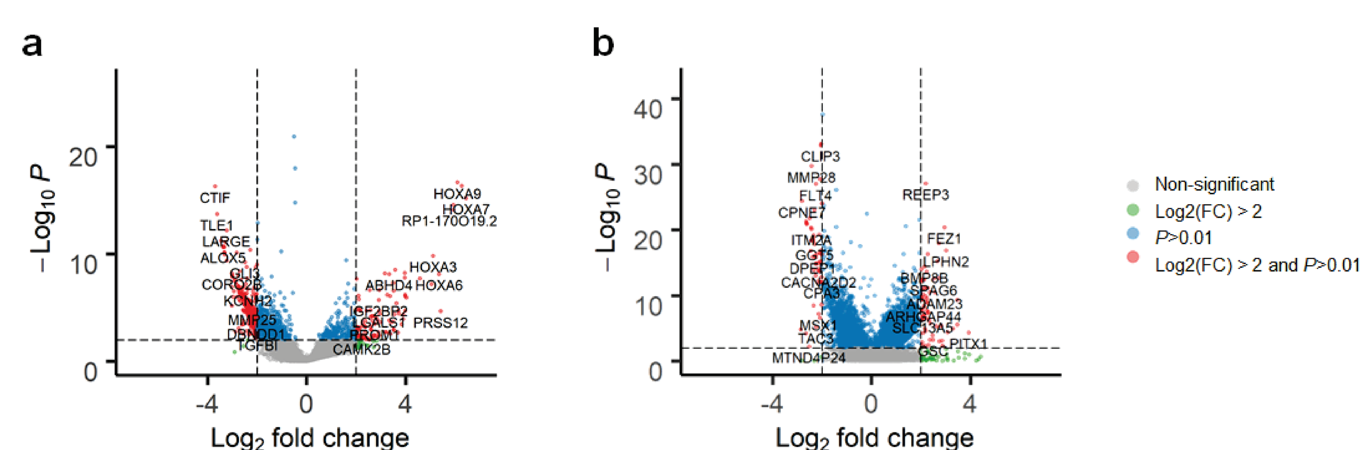

Figure 4. Volcano plot illustrates differentially expressed genes in acute leukemia with *MLL-r*. The top twenty up and down regulated genes were highlighted for (a) B-ALL and (b) AML. Dashed lines indicate log2(fold-change) > 2, and P>0.01.

Table 1. Top upregulated genes in both B-ALL and AML.

#[a]	Genes	Cellular component[b]	B-ALL Counts	B-ALL LFC	B-ALL LFC SE	B-ALL Statistic[c]	B-ALL P-value[c]	B-ALL Adj P-value[c]	AML Counts	AML LFC	AML LFC SE	AML Statistic[b]	AML P-value[c]	AML Adj P-value[c]
1	*SKIDA1*	Cytosol and nucleus	99	3.27	0.61	4.98	6.4E-07	1.6E-04	1233	2.11	0.30	6.84	8.0E-12	1.8E-09
2	*NMUR1*	Plasma membrane	58	2.64	0.62	4.02	5.8E-05	4.3E-03	767	1.83	0.30	5.89	3.8E-09	3.9E-07
3	*VAT1L*	Cytosol	2016	3.41	0.70	3.65	2.7E-04	1.2E-02	632	1.69	0.34	5.86	4.5E-09	4.5E-07
4	*ADCY9*	Plasma membrane; cytosol	3071	1.24	0.40	3.08	2.1E-03	4.9E-02	5100	1.75	0.27	6.32	2.6E-10	3.7E-08
5	*MRPL33*	Mitochondrion	3755	1.16	0.37	3.13	1.7E-03	4.4E-02	11863	1.05	0.18	5.95	2.7E-09	2.9E-07
6	*LINC00092*	NA	72	3.55	0.69	4.09	4.3E-05	3.4E-03	49	1.60	0.29	5.34	9.2E-08	5.6E-06
7	*NUDT7*	Peroxisome and cytosol	224	1.72	0.50	3.41	6.6E-04	2.3E-02	574	1.25	0.23	5.50	3.8E-08	2.7E-06
8	*PQLC2*	Lysosomal membrane	460	0.99	0.33	2.98	2.9E-03	6.1E-02	1672	1.20	0.19	6.17	6.8E-10	8.7E-08
9	*ROPN1L*	Cilium	58	3.59	0.58	5.93	3.0E-09	2.2E-06	425	1.07	0.22	4.93	8.2E-07	3.4E-05
10	*GOLGA8N*	Golgi apparatus	48	1.85	0.65	2.64	8.3E-03	1.2E-01	516	1.67	0.25	6.63	3.4E-11	6.2E-09
11	*AP1S3*	CVM and Golgi apparatus	361	1.56	0.52	2.93	3.4E-03	6.9E-02	641	1.12	0.21	5.26	1.4E-07	8.2E-06
12	*TUBA8*	Cytoskeleton	2401	1.73	0.55	3.08	2.1E-03	5.0E-02	511	1.02	0.21	4.96	7.2E-07	3.0E-05
13	*CAPG*	Flemming body and nucleus	4014	1.60	0.47	3.37	7.5E-04	2.5E-02	14058	1.15	0.25	4.59	4.5E-06	1.4E-04
14	*PARP8*	Intracellular	6685	2.14	0.49	4.31	1.6E-05	1.9E-03	12814	0.86	0.20	4.39	1.1E-05	3.0E-04
15	*ITGA7*	Plasma membrane	113	1.51	0.51	2.93	3.4E-03	6.9E-02	1550	1.33	0.29	4.54	5.6E-06	1.7E-04
16	***CSPG4***	**Plasma membrane**	**677**	**3.91**	**0.70**	**4.07**	**4.7E-05**	**3.6E-03**	**1416**	**1.46**	**0.33**	**4.24**	**2.3E-05**	**5.2E-04**

[a]Upregulated genes in the *MLL*-r (vs. *MLL* wild-type) group were classified based on lower p-value, and ranked for B-ALL and AML. The score calculation was performed by the sum of gene classification in both B-ALL and AML. The same weight was applied for both diseases.
[b]COMPARTMENTS - Subcellular localization database
[c]Wald's test.
CVM, Cytoplasmic vesicle membrane; LFC, log2(fold change); LFC SE, log2(fold change) Standard Error; NA, not available.

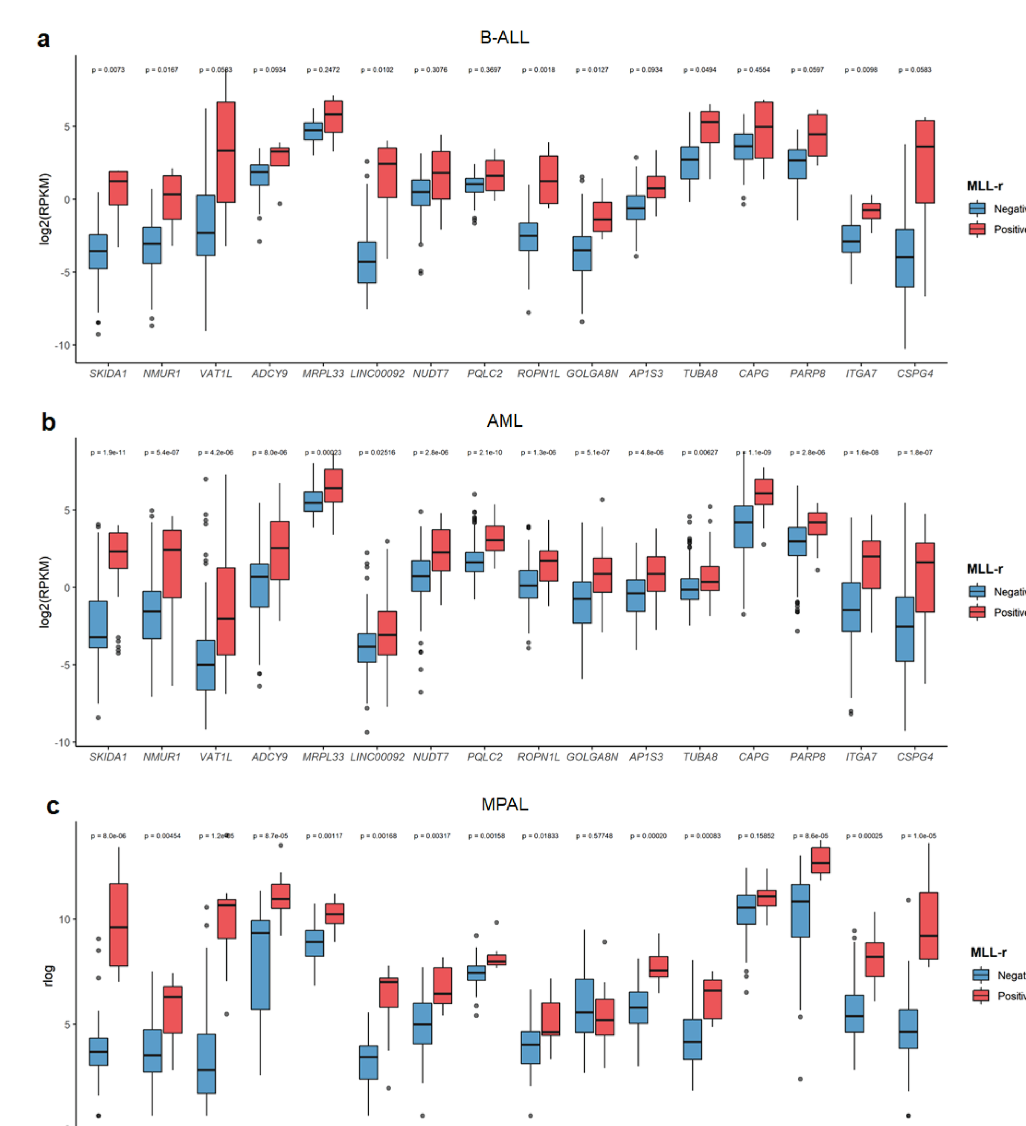

Figure 5. Expression of top 16 upregulated genes according to *MLL* status. Gene expression was assessed for (a) B-ALL, (b) AML, and (c) MPAL.

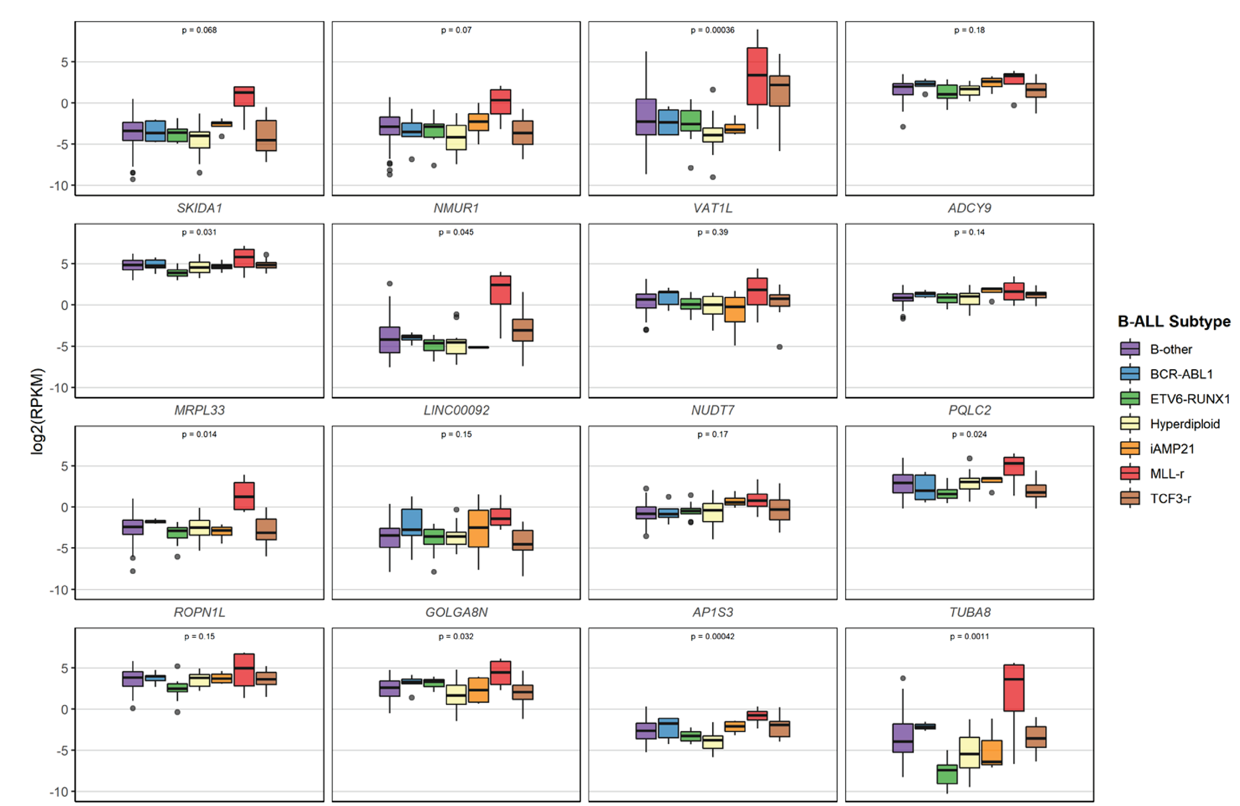

Figure 6. Expression of top 16 upregulated genes according to molecular subtypes of B-ALL.

CONCLUSION

Although we have established a customized NGS assay for the detection of *MLL-r* in acute leukemia, this approach is not available for all laboratories working with leukemia diagnosis. Therefore, the frequency estimation of the main *MLL* partner genes, as well as the development of cheaper and more feasible tools for the determination of *MLL-r* is important for many labs worldwide. Here, we evaluated which flow cytometry marker would be the most appropriate for the diagnosis of *MLL-r*, revealing that SKIDA1 could be applied for the identification of *MLL-r* in different types of acute leukemia.

CAPES · Unterstützt von / Supported by Alexander von Humboldt Stiftung/Foundation · INCA · DCAL · GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Projeto Gráfico: Área de Edição e Produção de Materiais Técnico-Científicos / INCA

INCA · SUS · MINISTÉRIO DA SAÚDE · PÁTRIA AMADA BRASIL GOVERNO FEDERAL