Med Oncol (2013) 30:579
DOI 10.1007/s12032-013-0579-7

ORIGINAL PAPER

Single-nucleotide polymorphism-array improves detection rate of genomic alterations in core-binding factor leukemia

Ana Rosa da Silveira Costa · Anupama Vasudevan · Ana Krepischi · Carla Rosenberg · Maria de Lourdes L. F. Chauffaille

Received: 13 February 2013 / Accepted: 10 April 2013 / Published online: 1 May 2013

Abstract Acute myeloid leukemia (AML) is a group of clonal diseases, resulting from two classes of mutation. Investigation for additional abnormalities associated with a well-recognized subtype, core-binding factor AML (CBF-AML) can provide further understanding and discrimination to this special group of leukemia. In order to better define genetic alterations in CBF-AML and identify possible cooperating lesions, a single-nucleotide polymorphism-array (SNP-array) analysis was performed, combined to KIT mutation screening, in a set of cases. Validation of SNP-array results was done by array comparative genomic hybridization and FISH. Fifteen cases were analyzed. Three cases had microscopic lesions better delineated by arrays. One case had +22 not identified by arrays. Submicroscopic abnormalities were mostly non-recurrent between samples. Of relevance, four regions were more frequently affected: 4q28, 9p11, 16q22.1, and 16q23. One case had an uncovered unbalanced inv(16) due to submicroscopic deletion of 5́MYH11 and 3́CBFB. Telomeric and large copy number neutral loss of heterozygosity (CNN-LOH) regions (>25 Mb), likely representing uniparental disomy, were detected in four out of fifteen cases. Only three cases had mutation on KIT gene, enhancing the role of abnormalities by SNP-array as presumptive cooperating alterations. Molecular karyotyping can add valuable information to metaphase karyotype analysis, emerging as an important tool to uncover and characterize microscopic, submicroscopic genomic alterations, and CNN-LOH events in the search for cooperating lesions.

Keywords Acute myeloid leukemia · Chromosome aberration · Mutation · Loss of heterozygosity · Oligonucleotide array sequence analysis

A. R. S. Costa · M. L. L. F. Chauffaille (✉)
Division of Hematology and Hemotherapy, UNIFESP/Escola Paulista de Medicina, São Paulo, SP CEP: 04037-002, Brazil
e-mail: chauffaille@unifesp.br

A. R. S. Costa
e-mail: aninhabra@yahoo.com.br

A. Vasudevan
Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA

A. Krepischi
CIPE-Hospital AC Camargo, São Paulo, Brazil

C. Rosenberg
Departamento de Genética e Biologia Evolutiva–IBUSP, São Paulo, Brazil

Introduction

Acute myeloid leukemia is a clonal disease resulting from a multistep association of cooperating mutations [1, 2]. Acute myeloid leukemia (AML) whose normal core-binding factor (CBF) transcription complex is disrupting, commonly referred to as CBF-AML: t(8;21)(q22;q22) [*RUNX1-RUNX1T1*] and inv(16)(p13.1q22)/t(16;16)(p13.1;q22) [*CBFB-MYH11*], is a good model for this theory [3]. Animal studies and clinical evidences support the notion that gene rearrangements affecting CBF function are critical, but insufficient to induce AML [1, 2, 4]. A number of chromosomal alterations and *KIT* mutations are strongly associated with this type of leukemia, some of them with prognostic implications [5–7]. However, in some CBF-AML cases, no cooperating abnormality can be identified at a microscopic and molecular level by current methods, which include G-banding karyotype (KT) and direct sequencing [4].

Recent advances in DNA microarrays technology have allowed for the achievement of great resolution and precision [8]. Single-nucleotide polymorphism-arrays (SNP-array) use a high density of polymorphic and non-polymorphic probes to interrogate deletions, gains, and copy number neutral losses of heterozygosity (CNN-LOH) along the whole human genome at a submicroscopic level. This appears to be an exceptional tool to investigate genetic alterations in leukemia [8].

In order to better define genetic abnormalities in CBF-AML at a submicroscopic level and to identify cooperating lesions, we performed SNP-array analysis, associated with *KIT* mutation screening in a set of leukemia patients.

Materials and methods

Patients and samples

Fifteen patients with CBF-AML rearrangements, identified by G-banding karyotype (KT), whose genomic material was available, were retrospectively selected from the year 2000 to 2008. AML diagnosis and reclassification were done as recommended by WHO [9]: bone marrow smear, immunophenotyping, and KT. The diagnosis of CBF-AML had been made based on the presence of t(8;21), inv (16), or t(16;16) by KT. FISH was done for confirmation of rearrangements found by KT in uncertain cases. DNA was extracted from residual cytogenetics samples that were stored in Carnoy's fixative at −20 °C, using the QIAamp DNA purification Kit (Qiagen).

SNP-array experiments and analysis

SNP-array (Genome-Wide Human SNP Array 6.0, Affymetrix) experiments were carried out as recommended by the manufacturer. Copy number data were validated by array comparative genomic hybridization (array-CGH) (SurePrint G3 Custom CGH 4x180K, Agilent Technologies) and interphase FISH (iFISH) with probes for: *BCR* gene, XY (Vysis, Abbott Laboratories), and *CBFB/MYH11* genes (Cytocell). Specific analysis softwares were applied to each platform: Genotyping Console™ (Affymetrix) was used for SNP-arrays and DNA Analytics™ (Agilent Technologies) for array-CGH. In addition, the third-part software, Nexus Copy Number 6.0™ (BioDiscovery), was used to merge the results from both the platforms using a SNP-Rank segmentation algorithm. Only segments with at least 25 consecutive markers and a $\log_2$ratio >0, 2 and <−0, 22 were computed as copy number alterations (gains and losses, respectively). An effective resolution of 100 Kb was established in all platforms. Regions reported in the general population as copy number variations (CNV) by the Database of Genomic Variants (http://projects.tcag.ca/variation/) were disregarded as relevant to the cancer phenotype, as well as CNN-LOH lesions smaller than 8 Mb [10]. CNN-LOH alterations were considered pathologic and a secondary event, likely representing uniparental disomy (UPD), if distal (encompassing the telomeric regions of chromosomes) or longer than 25 Mb in the interstitial chromosomal regions [10]. All CNN-LOH alterations were confirmed as copy number neutral by array-CGH experiments.

KIT mutation analyses

KIT mutations (exons 8 and 17) were retrospectively investigated by direct sequencing as described previously [11], using the same material from the arrays experiments. Direct sequencing of the amplicons was performed using the same primers with the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) on the ABI 3730 Genetic Analyzer (Applied Biosystems). Statistical comparisons using Nexus (Fisher's exact test) were performed to interrogate the association between abnormalities by arrays and the presence of *KIT* mutations.

Results

Patients selected are displayed in Table 1; twelve had t(8;21) and three, inv (16). The mean age was 43.1 years and the median age, 46 years (range from 22–77). Seven out of fifteen were male. Eight cases had no additional alteration detected by KT analysis, beside t(8;21) or inv (16) rearrangements (Table 1).

Relevant microscopic (>5 Mb) alterations only detected after array results were as follows(Table 1): additional material from 1q25q44 region (case 6), interstitial 9q22 deletion, (case 12) and trisomy 8 (case 7). On the other hand, in one case, the array investigation (SNP-array and array-CGH validation) did not detect trisomy 22 (case 3) [12]. iFISH for the 22q11.2 region (*BCR* gene) in this case confirmed one extra copy of chromosome 22 in 17 % of nuclei examined, as showed by KT.

Moreover, many submicroscopic (<5 Mb) chromosomal alterations were detected by SNP-array analysis, but a few could be validated by array-CGH (Table 1). Most of these were non-recurrent. One case of unbalanced inv (16) detected only by arrays displayed a deletion of the 5′*MYH11* and 3′*CBFB* sequences (case 3). iFISH with *CBFB/MYH11* probes demonstrates the presence of this breakpoint deletion and inv(16) variant (Fig. 1).

CNN-LOH lesions (>8 Mb) were found in six cases. Telomeric and large CNN-LOH regions (>25 Mb), likely UPD, were detected in four out of fifteen cases (Table 1).

Table 1 Summary of karyotype, microscopic lesions by SNP-array (> 5 Mb), iFISH validation, submicroscopic lesions by SNP-array (only array-CGH validated), and *KIT* mutation screening

Case	Karyotype	SNP-array (>5 Mb)	iFISH	SNP-array (<5 Mb)- validated by array-CGH	Long CNN-LOH (>8 Mb)	KIT mutation
1	45, X, −Y, t (8;21) (q22;q22) [7] /46,XY, t (8;21) (q22;q22) [8]	−Y	−Y	arr 7p22.3(153,149–277,773)x3; 8p23.1(7,231,610–7,678,486)x3*; 14q21.3(44,893,618-45,082,339)x1*; 14q32.33(105,568,216–105,832,094)x3*	4q[b], 7p, 7q, 9q[b] 14q, 15q, 16p, 16q[b]	D816V
2	46, XY, t (8;21) (q22;q22) [15]	–		arr 16q23.1(74,948,844–75,085,775)x1	–	–
3[a]	47, XX, +8, inv(16) (p13.1q22) [11]/47, XX, inv(16) (p13.1q22), +22 [4]/46,XX [3]	+8	inv(16) variant; +22	arr 16p13.11(15,657,788–16,022,591)x1; 16q22.1(65,590,670–66,241,220)x1; 16q23.1(77,521,373–78,018,855)x1	–	–
4	45, X, −Y, t(8;21) (q22;q22) [8]/46, XY, t(8;21) (q22;q22) [12]	−Y		–	–	–
5	46, XX, t(8;21) (q22;q22) [12]	–		arr 10q26.3(133,849,579–135,374,737)x3*; 15q11.1q11.2(18,276,329–20,232,527)x3*; 15q25.3(83,736,378–83,876,449)x3	1p, 13q(telomeric)[b]	–
6	46, XX, t(8;21) (q22;q22), dup(9)(q13) [9]	70 Mb gain1q: arr 1q25.2q44(176,337,651-247,249,719)x3		arr 2p25.3(1,411,113–1,497,190)x3; 14q32.33(105,058,434–106,005,461)x3*	–	–
7	47, XX, +9, inv(16)(p13q22) [10]	+8		–	–	–
8	46, XY, t(8;21)(q22;q22) [20]	−Y	−Y	–	6p (10 Mb)	–
9	46, XY, inv(16) (q13q22) [5]	–		arr 4q28.3(132,166,536–132,582,395)x4	–	D816F
10	45, X, −Y, t(8;21)(q22;q22) [10]/46, XY, t(8;21)(q22;q22) [10]	−Y		arr 15q11.1q11.2(18,863,219–20,207,987)x3*	Whole 4q[b] (Mosaic)	–
11	46, XX, t (8;21)(q22;q22) [20]	–		arr 10q21.1(56,522,467–56,790,604)x1	1q, 6q, 8q, 9p[b], 9q, 13q[b], 15q, 16q	–
12	45, X, −Y, t(8;21)(q22;q22) [8]/45, X, −Y, t(8;21)(q22;q22), del(9)(q22) [1]/ 46,XY,t(8;21)(q22;q22) [6]	−Y del(9)(q21.31;q31.1), about 22,5 Mb, by visual inspection		arr 7p12.1(52,526,685–52,773,329)x3; 15q11.1q11.2(18,774,193–20,144,774)x3*;	–	–
13	46, XX, t(8;21)(q22;q22) [20]	–		arr 4q28.3(132,144,643–132,579,118)x4; 14q11.1q11.2(18,537,177–19,500,196)x3*	Xp(8 Mb)	N822 K
14	46, XX, t(8;21)(q22;q22) [20]	–		arr 19q13.31(47,814,405–48,524,839)x3*	–	–
15	46, XX, t(8;21)(q22;q22) [2]/46,XX [1]	–		–	–	–

iFISH interphase FISH, *SNP-array* single-nucleotide polymorphism-array, *array-CGH* array comparative genomic hybridization

* likely CNV

[a] Costa et al. (2012)

[b] likely somatic acquired uniparental disomy (UPD

However, this type of abnormality was dispersed to different chromosomes, and there was no predilection to any region. Interestingly, one case negative for the *KIT* mutation (case 10) exhibited a mosaic CNN-LOH event, affecting the entire 4q (Fig. 2).

Four regions were recurrently altered in different samples: (a) 4q28, with a minimal overlapping region of 412 Kb (chr4:132,144,643-132,579,118): high-level gain in cases 9 and 13 (both positive for *KIT* mutation) and CNN-LOH in cases 1 and 10; (b) 9p11, with a minimal overlapping region of 400 Kb (chr9:43,373,695-43,765,864): CNN-LOH in cases 1 and 11;(c) 16q22.1, with a minimal overlapping region of 2.2 Mb (chr16:66,989,302-69,257,789): CNN-LOH in cases 1 and 13; and finally, (d) 16q23, with a minimal overlapping region of 225 Kb (chr16:74,926,441-75,150,825): CNN-LOH in cases 1 and 11 and deletion in case 2.

KIT mutations were only detected in three cases (1, 9, and 13). All mutations were found clustered within exon *17,* which encodes the activation loop (A-loop) in the kinase domain, in two different hotspots: D816 and N822. One case (nr 9) had a novel substitution of the amino acid D826: D816F. Another case (nr 1) had a novel 9 bp *in tandem* duplication within D816, causing an alteration in the reading frame downstream and leading to an amino acid substitution (D816 V), which is a known *KIT* mutation. The statistical analysis using Nexus demonstrated that *KIT* mutation-positive cases were associated with a high-level gain of a 412 kb region at 4q28 and a 2.2 Mb CNN-LOH at 16q22.1 ($p < 0.05$).

Discussion

Cytogenetic abnormalities identified by G-banding karyotype remain the most important prognostic factor in AML [13]. However, AML cases carrying similar karyotypic

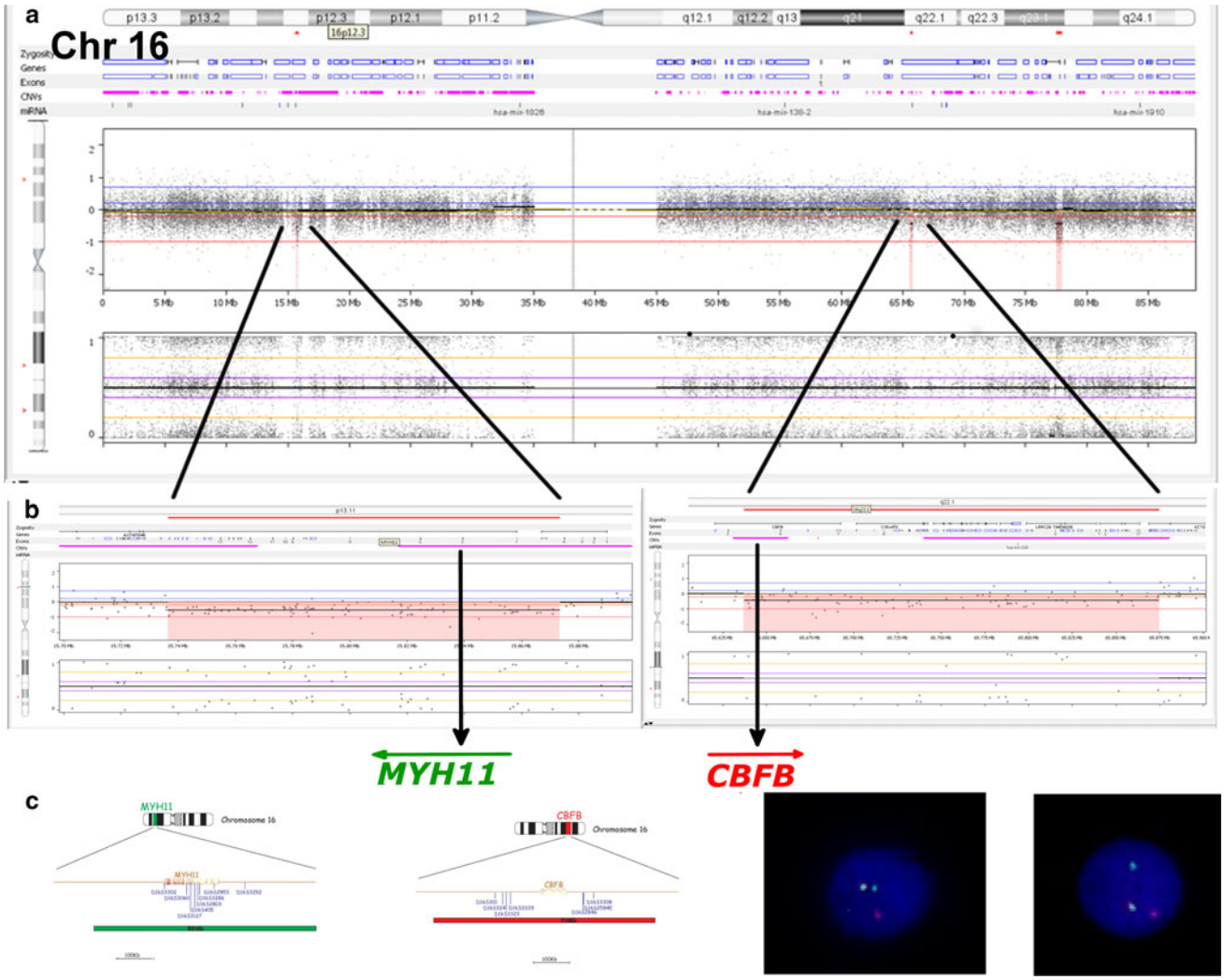

Fig. 1 Array karyotype results to case 3. **a** Chromosome 16 by SNP-array. **b** breakpoint deletion within genes *MYH11* and *CBFB*. **c** iFISH probe design and the three iFISH signal for inv(16) variant in this case

alterations exhibit a wide range of clinical outcome. Compared with other types of AML, CBF-AML shows good prognosis [13]. However, CBF-AML patients have 50 % of overall 5-year survival rate, suggesting the presence of additional genetic mutations leading to diverse outcomes. Many submicroscopic alterations, such as point mutations, have been discovered during the last decade by target methods [13], but the findings are not sufficient to explain all heterogeneity. The emergence of these new genome-wide molecular karyotyping technologies opens up a new perspective of AML studies once cryptic unbalanced copy number alterations and CNN-LOH/UPD can now be detected in a single assay [8].

Molecular karyotyping by SNP-array of these fifteen CBF-AML cases demonstrated additional information regarding microscopic, submicroscopic, and CNN-LOH alterations and that the combination of SNP-array with KT improves the detection rate for abnormalities. In this set of patients, *KIT* mutations were found in just three cases, and no other secondary event was revealed by KT or direct sequencing in at least six patients, enhancing the role of alterations detected by SNP-array as possible cooperating alterations. Although other gene mutations, such as *FLT3* and *RAS* family, were also associated with CBF-AML [4], the *KIT* gene remains the most important genetic driver in this AML subtype.

The present study also calls attention to the prevalence of CNN-LOH lesions, even when more strict criteria were used, such as telomeric or long stretch CNN-LOH [10]: four out of fifteen. As highlighted by some authors [8, 10], CNN-LOH lesions are important alterations for hematological neoplasias, starting with *JAK2* mutation discovery in a CNN-LOH region within 9p in polycythemia vera. Later also explored by Raghavan et al. [14] that showed CNN-LOH prevalence around 20 % in AML using SNP-array technology. More studies should better explore the role of these alterations to treatment and prognosis, as done by Tiu et al. [15].

Regarding to microscopic alterations (>5 Mb), SNP-array analysis (and array-CGH validation) could better define some of them and uncover others, as in case 6, whose array analysis revealed that a 9q13 duplication previously defined by KT was in fact a 1q25.2 gain, or in case 7, whose array analysis revealed trisomy 8, whereas previously, it was interpreted as trisomy 9 by KT. Also, in case 12, the terminal 9q22 deletion was in fact a 22,5 Mb interstitial 9q22 deletion in a low frequency subclone. On the other side, metaphase analysis in case 3 showed +22 not detected by arrays, likely because this subclone is under-represented in the whole DNA sample [12].

In addition, arrays showed a submicroscopic unbalanced inv (16) in case 3 [12]. This was an approximately 130 Kb deletion at 5′ *MYH11* and a 230 Kb deletion at 3′ *CBFB*. Both of these deletions were illustrated by iFISH (Fig. 2) to be variants of inv (16), which seems to be a rare condition whose prognostic implication is still unclear [16]. Another

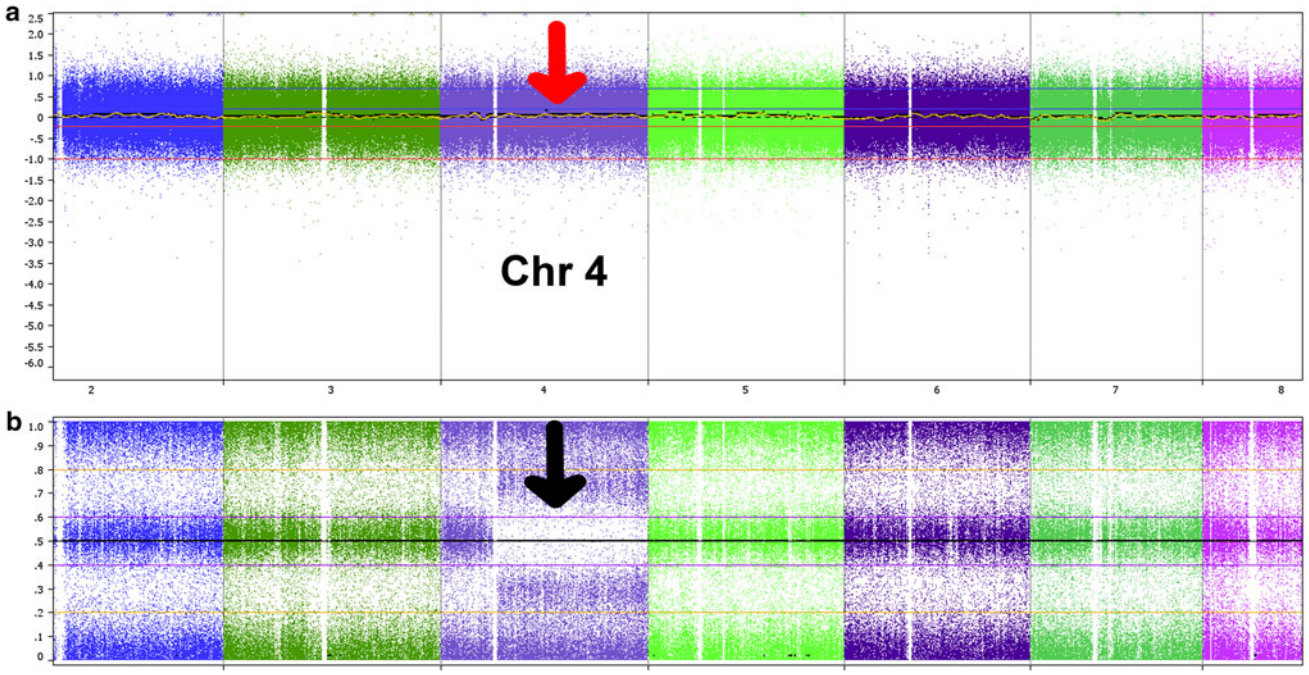

Fig. 2 SNP-array results to case number 10. **a** Copy number plot showing no copy number variation within Chromosome 4 (*light purple*), and *red arrow* corresponds to long arm 4q. **b** B-allele frequency data showing dispersion of probes running from center, but not in extremity as found in classic CNN-LOH lesion, configuring a mosaic CNN-LOH

submicroscopic deletion in chromosome 16 uncovered by arrays in this case was a 500 Kb deletion in 16q23.1 at the 3′ portion of the *WWOX* gene. Many studies imply the *WWOX* gene as an important tumor suppressor gene [17]. The *WWOX* gene resides in the very same region as the common chromosomal fragile site 16D (FRA16D), which is one of the most sensitive common chromosomal fragile site loci in the human genome [18, 19]. Of clinical relevance, this case (Case 3) was negative for the *KIT* mutation and had an unexpectedly bad outcome.

Concerning to submicroscopic lesions (<5 Mb), the recurrence was low, in accordance with previous studies, and located in different regions from that previously described [20]. Maybe this data represent the submicroscopic heterogeneity between samples, which could account for the diverse clinical outcome between CBF-AML patients. Although many submicroscopic alterations were detected by SNP-array in all cases, the array-CGH platform was not able to validate some of them, likely due to differences in the coverage and density of probes, which are lower in the array-CGH platform as compared to the SNP-array. Submicroscopic lesions not confirmed using array-CGH need to be validated by other methodology before being considered true copy number alterations. Yet, many alterations were found in CNV regions (and disregarded in this study), and their implications in oncogenesis could only be assessed by a paired analysis with germline samples.

This study showed that some regions were more affected by alterations. Remarkably, the confluent sites 4q28, 9p11, 16q22.1, and 16q23 exhibited copy number alterations and CNN-LOH in distinct cases. As well, the known minimal deleted region (MDR) 9q21 associated with t(8;21) [19] was present in two cases: as an CNN-LOH (case 11) and as an interstitial deletion (case 12), emphasizing all these regions as possible sites for leukemia-related genes. The 4q28 region, marked by high gain and CNN-LOH in different samples, has no gene described in the NCBI gene database (reference GRCh37/hg19 assembly). Even those gene-poor regions can affect gene regulation in many aspects, and little is known about the influence of copy number alterations to health or disease [21]. Regions 9p11 and 16q23, marked by deletion and CNN-LOH alterations, have two members of the contactin associated protein family gene: *CNTNAP3B* and *CNTNAP4,* respectively. This gene family has unclear significance for cancer. However, this 16q23 region is close to FRA16D, found in case 3 and discussed above. In opposition, the 16q22.1 region has dozens of genes described in the reference assembly, and some of them have already been implicated in oncogenesis, such as *CDH1* and *CDH3,* whose loss of function is involved in tumor progression.

Two cases carrying t(8;21) and one case with inv (16) had *KIT* mutations within exon 17. In one patient (case 9) with t(8;21), we describe a novel mutation that led to the replacement of one amino acid in the A-loop. The statistically significant association between 4q28 and 16q22.1 lesions and the *KIT* mutation-positive cases were not explored here because they should first be confirmed in a larger sample.

After this investigation, two cases (nr 14 and 15) had no additional abnormalities, beyond rearrangements, revealed by all these methods: KT, molecular arrays, and *KIT* mutation screening (Table 1).

In conclusion, molecular karyotyping can add valuable information to metaphase KT analysis, even in cases of apparently balanced translocations such as CBF-AML, emerging as an important tool to uncover and characterize microscopic, submicroscopic genomic alterations, and CNN-LOH events in the search for cooperating lesions.

Acknowledgments We thank to Mônica Castro Varela, PhD and Roseli Zanelato from Instituto de Biociências da USP/SP for their support with *KIT* mutation study. This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

Conflict of interest All authors have no conflict of interest to report.

References

1. Gilliland DG. Hematologic malignancies. Curr Opin Hematol. 2001;8(4):189–91.
2. Kelly LM, Gilliland DG. Genetics of myeloid leukemias. Annu Rev Genomics Hum Genet. 2002;3:179–98.
3. Downing JR. The core-binding factor leukemias: lessons learned from murine models. Curr Opin Genet Dev. 2003;13(1):48–54.
4. Peterson LF, Boyapati A, Ahn EY, Biggs JR, Okumura AJ, Lo MC, et al. Acute myeloid leukemia with the 8q22;21q22 translocation: secondary mutational events and alternative t(8;21) transcripts. Blood. 2007;110(3):799–805.
5. Schlenk RF, Benner A, Krauter J, et al. Individual patient data-based meta-analysis of patients aged 16 to 60 years with core binding factor acute myeloid leukemia: a survey of the German Acute Myeloid Leukemia Intergroup. J Clin Oncol. 2004;22(18):3741–50.
6. Appelbaum FR, Kopecky KJ, Tallman MS, et al. The clinical spectrum of adult acute myeloid leukaemia associated with core binding factor translocations. Br J Haemato. 2006;135(2):165–73.
7. Paschka P. Core binding factor acute myeloid leukemia. Semin Oncol. 2008;35(4):410–7.
8. Maciejewski JP, Mufti GJ. Whole genome scanning as a cytogenetic tool in hematologic malignancies. Blood. 2008;112(4):965–74.
9. Swerdlow SH, Campo E, Harris NL, Pileri SA, Stein H, Thiele J, et al. World Health Organization classification of tumours of haematopoietic and lymphoid tissues. Lyon: IARC Press; 2008.
10. O'Keefe C, McDevitt MA, Maciejewski JP. Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. Blood. 2010;115(14):2731–9.
11. Pollard JA, Alonzo TA, Gerbing RB, Ho PA, Zeng R, Ravindranath Y, et al. Prevalence and prognostic significance of KIT mutations in pediatric patients with core binding factor AML

enrolled on serial pediatric cooperative trials for de novo AML. Blood. 2010;115(12):2372–9.
12. Costa AR, Belangero SI, Melaragno MI, Chauffaille ML. Additional chromosomal abnormalities detected by array comparative genomic hybridization in AML. Med Oncol. 2012;29(3):2083–7.
13. Grimwade D, Hills RK. Independent prognostic factors for AML outcome. Hematol Am Soc Hematol Educ Program. 2009;2009: 385–95.
14. Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, Foot NJ, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. Cancer Res. 2005;65(2):375–8.
15. Tiu R, Gondek L, O′Keefe C, Huh J, Sekeres M, Elson P, et al. New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid eulemia. J Clin Oncol. 2009;27(31):5219–26.
16. Kelly J, Foot NJ, Conneally E, Enright H, Humphreys M, Saunders K, et al. 3′CBFbeta deletion associated with inv(16) in acute myeloid leukemia. Cancer Genet Cytogenet. 2005;162(2): 122–6.
17. Abdeen SK, Salah Z, Maly B, Smith Y, Tufail R, Abu-Odeh M, et al. Wwox inactivation enhances mammary tumorigenesis. Oncogene. 2011;30(36):3900–6.
18. Ekizoglu S, Muslumanoglu M, Dalay N, Buyru N. Genetic alterations of the WWOX gene in breast cancer. Med Oncol. 2012;29:1529–35.
19. O'Keefe LV, Richards RI. Common chromosomal fragile sites and cancer: focus on FRA16D. Cancer Lett. 2006;232(1):37–47.
20. Kühn MW, Radtke I, Bullinger L, Goorha S, Cheng J, et al. High-resolution genomic profiling of adult and pediatric core-binding factor acute myeloid leukemia reveals new recurrent genomic alterations. Blood. 2012;119(10):e67–75.
21. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464(7289):704–12.