Close Correlation of Copy Number Aberrations Detected by Next-Generation Sequencing with Results from Routine Cytogenetics in Acute Myeloid Leukemia

Sebastian Vosberg[1,2,3,4], Tobias Herold[2,3,5], Luise Hartmann[1,2,3,4], Martin Neumann[2,3,6], Sabrina Opatz[2,3,5], Klaus H. Metzeler[2,3,5], Stephanie Schneider[5], Alexander Graf[7], Stefan Krebs[7], Helmut Blum[7], Claudia D. Baldus[6], Wolfgang Hiddemann[1,2,3,4], Karsten Spiekermann[1,2,3,4], Stefan K. Bohlander[8], Ulrich Mansmann[2,3,9] and Philipp A. Greif[1,2,3,4]*

[1] Experimental Leukemia and Lymphoma Research (ELLF), Department of Internal Medicine III, University Hospital of the Ludwig-Maximilians-Universität (LMU) München, Munich, Germany
[2] German Cancer Research Center (DKFZ), Heidelberg, Germany
[3] German Cancer Consortium (DKTK), Heidelberg, Germany
[4] Clinical Cooperative Group Leukemia, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany
[5] Laboratory for Leukemia Diagnostics, Department of Internal Medicine III, University Hospital of the Ludwig-Maximilians-Universität (LMU) München, Munich, Germany
[6] Department of Hematology and Oncology, Charité University Hospital, Berlin, Germany
[7] Laboratory for Functional Genome Analysis (LAFUGA), Gene Center of the Ludwig-Maximilians-Universität (LMU) München, Munich, Germany
[8] Molecular Medicine and Pathology, The University of Auckland, Auckland, New Zealand
[9] Institute for Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität (LMU) München, Munich, Germany

Correspondence to: Philipp A. Greif, MD, Max-Lebsche-Platz 30, 81377 Munich, Germany
Tel: +49-89-4400-43982, Fax: +49-89-4400-43970, Email: pgreif@med.uni-muenchen.de; p.greif@dkfz.de

Short title: CNA Analysis in NGS Data from AML Patients

Supported by: the German Cancer Aid (grant 109031 to P.A.G. and S.K.B.), the Wilhelm-

Sander-Stiftung (grant 2014.162.1 to P.A.G. and H.B.)

High throughput sequencing approaches, including the analysis of exomes or gene panels, are widely used and established to detect tumor-specific sequence variants such as point mutations or small insertions/deletions. Beyond single nucleotide resolution, sequencing data also contain information on changes in sequence coverage between samples and thus allow the detection of somatic copy number alterations (CNAs) representing gain or loss of genomic material in tumor cells arising from aneuploidy, amplifications, or deletions. To test the feasibility of CNA detection in sequencing data we analyzed the exomes of 25 paired leukemia/remission samples from acute myeloid leukemia (AML) patients with well-defined chromosomal aberrations, detected by conventional chromosomal analysis and/or molecular cytogenetics assays. Thereby, we were able to confirm chromosomal aberrations including trisomies, monosomies, and partial chromosomal deletions in 20 out of 25 samples. Comparison of CNA detection using exome, custom gene panel, and SNP array analysis showed equivalent results in five patients with variable clone size. Gene panel analysis of AML samples without matched germline control samples resulted in confirmation of cytogenetic findings in 18 out of 22 cases. In all cases with discordant findings, small clone size (<33%) was limiting for CNA detection. We detected CNAs consistent with cytogenetics in 83% of AML samples including highly correlated clone size estimation (R=0.85), while six out of 65 cytogenetically normal AML samples exhibited CNAs apparently missed by routine cytogenetics. Overall, our results show that high throughput targeted sequencing data can be reliably used to detect copy number changes in the dominant AML clone.

INTRODUCTION

Copy number alterations (CNAs) arise from deletion, duplication, or amplification of chromosomal regions or whole chromosomes and occur frequently as disease-initiating events in various human tumor entities (Beroukhim et al., 2010). In acute myeloid leukemia (AML), more than 50% of adult patients harbor chromosome aberrations including translocations or gain/loss of whole chromosomes or chromosomal parts (Grimwade, 2001; Mrózek et al., 2004). Chromosomal alterations define clinical subgroups of AML according to the WHO classification (Swerdlow et al., 2008).

In diagnostic routine, chromosomal aberrations are detected using conventional metaphase karyotyping and/or molecular cytogenetics, e.g. interphase fluorescence in situ hybridization (FISH). The findings provide valuable prognostic information, which is essential for therapeutic decisions. The presence or absence of chromosomal aberrations predicts response to induction chemotherapy, risk of relapse, and overall survival (Grimwade et al., 2001; Farag et al., 2006; Fröhling et al., 2006). Some aberrant karyotypes, especially monosomies of chromosomes 5 and 7, as well as monosomal and complex karyotypes are associated with unfavorable outcome (Ghanem et al., 2012). Cytogenetic risk stratification is challenging as AML is a very heterogeneous disease and a sample can present with multiple subclones harboring different alterations (Cancer Genome Atlas Research Network, 2013). The analysis of subclone estimation is limited in routine cytogenetics as by conventional karyotyping typically only ~20 metaphase cells are scored (Döhner et al., 2010).

Array based techniques and more recently whole genome sequencing are used as methods of choice for genome wide detection of CNAs in hereditary disease and cancer genomics (Carter, 2007; Rausch et al., 2012) including AML genomics (Walter et al., 2009; Ding et al., 2012). In contrast, targeted sequencing of AML patients including whole exome sequencing (WES) and gene panel sequencing (GPS) is commonly used to detect somatic

sequence alterations that might contribute to leukemogenesis (Yan et al., 2011; Greif et al., 2012; Cancer Genome Atlas Research Network, 2013). Recent advances in next generation sequencing (NGS) and target enrichment have led to a more sensitive and cost effective detection of somatic variants in tumor DNA (Ilyas et al., 2015) including copy number alterations (Bolli et al., 2015). Of note, targeted sequencing is preceded by enrichment of genomic DNA for a well-defined target region and varying enrichment efficiency between regions and samples leads to an unequal coverage distribution. A linear regression model was used to describe the test coverage as a linear function of the control coverage, normalizing for enrichment efficiency and enabling the comparison of zero coverage regions (Rigaill et al., 2012). Furthermore, the degree of sequence coverage difference reflects the size of the subclone harboring the CNA and thus gives information about the clonal architecture of a tumor.

Analysis of CNAs in AML patients gives us the opportunity to compare NGS data with multiple standard techniques in tumor biology like conventional karyotyping, molecular cytogenetics or SNP array profiling. We selected 25 AML patient samples with matched germline control samples for WES and subsequent detection of CNAs. A total of 22 AML patient samples with deletion on the long arm of chromosome 9 (del(9q)) were selected for GPS and CNAs were detected by comparison to a control cohort of 21 AML patient samples negative for del(9q). Five AML del(9q) samples with varying clone size harboring the deletion were analyzed with WES, GPS and SNP array profiling.

MATERIALS AND METHODS

Patient Characteristics

Patient samples were characterized centrally according to the German AML cooperative group (AMLCG) study protocols. Informed consent was obtained according to the declaration of Helsinki. Study protocols were approved by the institutional review boards of the participating centers. Cases were selected based on availability of high throughput sequencing data and cytogenetics reports. For WES, 90 paired leukemia and complete remission samples were included, with the leukemia sample taken either at initial diagnosis or at relapse (Table 1). Chromosomal alterations were detected in 25 AML samples, analyzed by conventional metaphase karyotyping (n=22) and/or by interphase FISH (n=15), with an overlap of 12 samples subject to both techniques. AML samples with poor karyogram quality and/or complex karyotype were additionally analyzed using 24-color multiplex-FISH (M-FISH, n=3). AML blast counts as reported by routine cytomorphology were available for 22 patients (88%). The clone size harboring the mutation was estimated as $mean(\frac{\#\ positive\ metaphases}{\#\ all\ metaphases}, \frac{\#\ positive\ interphases}{\#\ all\ interphases})$. A total of 65 AML samples were reported without any cytogenetic alteration by routine diagnostics (cytogenetically normal; CN-AML). Five AML patients with very similar alterations (partial deletion of the long arm of chromosome 9; del(9q)) but variable clonal architecture (6%-90%) were selected for additional SNP array profiling. Diagnostic samples from 43 AML patients were selected for GPS including 22 samples with del(9q) as sole unbalanced aberration of chromosome 9, with five samples already used for WES and SNP array profiling (Table 2), and 21 control samples without any known unbalanced cytogenetic aberration of chromosome 9. AML blast cell counts were available for 16 AML del(9q) patients (73%). As all 9q deletions were reported by conventional cytogenetics, the clone size was estimated as $\frac{\#\ positive\ metaphases}{\#\ all\ metaphases}$. Routine cytogenetics is performed following the guidelines of the European Leukemia Net-

Workpackage Cytogenetics (Haferlach et al., 2007). For metaphase karyotyping, cultured heparinized bone marrow aspirates and/or peripheral blood specimen were treated with colcemid and stained with Giemsa (G-banding). Single chromosome labelling by M-FISH (Speicher et al., 1996) was performed using the 24XCyte Human Multicolor FISH Probe Kit (MetaSystems, Altlussheim, Germany). The results were reported according to the International System for Human Cytogenetics Nomenclature guidelines (Shaffer et al., 2009). Interphase FISH analysis was performed on EDTA- or heparinized bone marrow smears treated with methanol:acetic acid (glacial) (3:1). Denatured, fluorescently labeled FISH probes were hybridized on the slide for 3 minutes at 76°C and over night at 37°C (all FISH probes used in this study are listed in Supplementary Table 1).

Exome Sequencing

The protein coding regions of the fragmented genomic DNA were captured according to the SureSelect Human All Exon 50Mb target enrichment protocol (Agilent, Santa Clara, CA) as described before (Opatz et al., 2013). Paired-end sequencing was performed either on an Illumina Genome Analyzer IIx (GA IIx, n=8 AML samples) resulting in 2 x 80 bp paired-end sequence reads or on an Illumina HiSeq 2500 (n=17 AML samples) resulting in 2 x 101 bp paired-end sequence reads. Depending on the sequencing platform, samples were sequenced with a mean coverage of 32x (range 28x-38x) or 104x (range 89x-127x). Leukemia samples and matched germline control samples from each patient were sequenced together on one instrument run to ensure technical comparability. Raw sequence data were processed as described previously and aligned to the human genome 19 (GRCh37) reference genome assembly obtained from the UCSC Genome Browser (Kent et al., 2002). Detailed exome sequencing metrics for each sample are summarized in Supplementary Table 2.

Targeted Amplicon Sequencing

Samples were prepared as described before (Herold et al., 2014) using a custom Haloplex gene panel (Agilent, Santa Clara, CA) including 140 AML related genes and mutation hotspots including 14 candidate genes located in the commonly deleted region (CDR) on the long arm of chromosome 9 resulting in a total target sequence of 492 kb, summarized in Supplementary Table 3. Paired-end sequencing with a read length of 250 bp was performed on an Illumina MiSeq instrument, and sequence reads were mapped to the human genome 19 (GRCh37) reference genome assembly. Detailed gene panel sequencing metrics for each patient are summarized in Supplementary Table 4.

SNP Array Profiling

SNP array profiling was performed using a CytoScan HD Array and analyzed using the Chromosome Analysis Suite 2.0 software (Affymetrix, Santa Clara, CA). Regions with altered gene copy number were defined by a ratio of 0.8 as cutoff for genomic loss and 1.2 as cutoff for genomic gain representing a variant allele frequency of 20% as applied previously to the detection of somatic sequence variants (Greif et al., 2012; Opatz et al., 2013). The minimum size of a CNA to be called was set to 5 Mb, as the resolution of metaphase karyotyping is about 5-10 Mb (Gelehrter et al., 1998). Neighboring CNA regions were merged if the interval between was less than 5 Mb and the copy number change differed less than 10%. For comparison with WES and GPS data, SNP positions were limited to coding exon regions.

CNA Calling in Sequencing Data

The mean coverage of each exon was computed using the DepthOfCoverage program from the Genome Analysis Toolkit for each sample separately (McKenna et al., 2010) with default parameters. Insufficient mean exon coverage values in both test and control samples

were replaced with overall mean coverage (minimum 10x for WES data, minimum 30x for GPS data). A linear regression model was applied to each chromosome separately in order to normalize for target enrichment efficiency and to model the test sample coverage as a linear function of the control sample coverage based on exon-by-exon comparisons. An exact segmentation algorithm separated regions of equal coverage from regions of differing coverage between test and control samples (Rigaill et al., 2012). We defined a minimum number of two exons per region and a maximum of five regions per chromosome. The minimum size of an altered region to be called as CNA was set to 5 Mb, corresponding to the resolution of SNP array analysis and classical cytogenetics. To detect CNAs in the dominant clone, the minimum coverage fold change was set to 0.8 for deletions and 1.2 for amplifications. If the CNA was reported in a cytogenetic subclone with an estimated clone size <50%, based on routine data, the minimum fold change for CNA calling was set to 0.9 and 1.1, respectively. Assuming heterozygous aberrations, the clone size containing the CNA was estimated as $\left|\frac{test\ coverage}{control\ coverage} - 1\right| \times number\ of\ wild\ type\ chromosomes \times 100\%$. Using WES data, AML test samples were compared with control samples from complete remission. Using GPS data, del(9q)-positive AML samples were compared pairwise to del(9q)-negative AML samples, resulting in 21 comparisons per test sample limited to chromosome 9, as patients of the control cohort had various alterations on other chromosomes. Start and stop positions of a CNA were defined as the median start and stop positions of all CNAs detected by comparisons. CNAs not called in the majority of the comparisons (n>10) were regarded as false positives and excluded from further analysis. Finally, we identified a commonly deleted region (CDR) on chromosome 9 by selecting overlapping CNAs that were called in the majority of AML del(9q)-positive patients (n≥11) by the median start and stop positions of CNAs detected in individual patients.

RESULTS

Detection of Chromosomal Aberrations by Exome Sequencing

CNA profiling based on WES data sets from 25 paired AML and remission samples (P1-P25, Table 1) showed somatically acquired chromosomal aberrations consistent with findings from routine cytogenetic diagnostics in 20 out of 25 samples (80%, Fig. 1). We were able to confirm various types of chromosomal abnormalities in AML patients such as aneuploidy (e.g. trisomy 8, P8; Fig. 2A) and partial amplifications or deletions of chromosomes (e.g. deletion on the long arm of chromosome 5, P15; Fig. 2B). All CNAs detected by WES of aberrant AML samples are listed in Table 1 (CNA plots for all AML samples are shown in Supplementary Fig. 1). In order to test for the detection of submicroscopic CNAs (<5 Mb) using WES, we selected a patient carrying an unbalanced *KMT2A/MLLT3* rearrangement (formerly known as *MLL/AF9*, P14) resulting from a translocation of chromosomes 9 and 11 with partial loss of the 3'-part of the rearranged *KMT2A* gene as reported by molecular cytogenetics using a *KMT2A* break apart FISH probe. We detected a significant exon copy number change in a region of 351 exons downstream of the breakpoint in *KMT2A* corresponding to a size of 931.8 kb (Fig. 2C). While *KMT2A* coding exons 1 to 10 show a similar coverage between leukemia and remission samples, from exon 11 onwards, the coverage in the leukemic sample is reduced by 47% indicating a heterozygous deletion affecting 95% of the cells. Thus, we were able to narrow down the position of the breakpoint to a region of 3.6 kb between exons 10 and 11 (genomic positions chr11:118,355,691-118,359,328; Supplementary Fig. 2). We detected chromosomal aberrations in four out of eight samples with low coverage exome data based on GA IIx sequencing (50%), and in 16 out of 17 samples with high coverage exome data sequenced on a HiSeq 2500 instrument (94%). AML samples in which CNAs could not be detected had significantly smaller aberrant clone sizes (mean 18%, range 8%-31%)

compared with AML samples with confirmed alterations (mean 67%, range 23%-95%; *P*=0.0016, Mann-Whitney test; Fig. 3A) as reported by routine cytogenetics (Table 1). Clone size estimation based on detected copy number change in WES reached high correlation with clone size estimation based on routine diagnostics (R=0.85, Pearson correlation; Fig. 1). CNA calling in WES data confirmed 95% of numerical aberrations that each were reported as dominant clone by routine diagnostics. In one sample (P5) with an estimated clone size of 66%, based on routine data, a trisomy of chromosome 8 could be confirmed by adjusting the minimum fold change from 1.2 to 1.1.

Interestingly, we detected additional aberrations not reported by routine karyotyping in four out of 25 AML samples (16%). In sample P9, we detected an additional deletion on the short arm of chromosome 11 with a clone size of 77%. In samples P16 and P17, representing diagnostic and relapse samples from the same patient, a deletion on the long arm of chromosome 7 could be detected with identical start and stop positions of the CNA in both samples and clone sizes of 93% and 74%, respectively. Further, a copy number gain on the long arm of chromosome 13 was detected in both samples with nearly identical start positions (neighboring exons) and identical stop positions of the CNA and clone sizes of 99% and 100%, respectively. In sample P20, we detected a partial duplication of the long arm of chromosome 1 with a clone size of 84%, while routine diagnostics reported a duplication of the short arm of chromosome 1 in conjunction with a derivative chromosome 18. Furthermore, we detected another deletion on the short arm of chromosome 17 comprising the *TP53* locus, with a clone size of 87%. In order to validate CNAs detected exclusively by WES, we performed FISH analysis. Based on the availability of appropriate sample material and FISH probes, four out of seven CNAs could be tested, with four confirmed (in P16 and P20). The results of the validation are summarized in Table 3.

Results of CNA calling in 65 CN-AML samples were consistent with a normal karyotype in 59 cases (91%). In total, we detected 12 CNAs in six CN-AML samples, with one patient

harboring seven CNAs (C15), including deletions, duplications and one trisomy (Table 3). Assuming heterozygous aberrations, the mean clone size harboring a CNA was 72% (range 47%-91%); deletions and duplications had a mean size of 26.1 Mb (range 5.5-81.2 Mb). A pair of diagnosis and relapse samples from the same patient (C5 and C6) harbored a partial duplication of chromosome Y with identical start but different stop positions, 3.76 Mb apart. In CN-AMLs with detectable CNAs based on WES, on average 21 metaphases were karyotyped in routine diagnostics (range 16-25) and cytomorphology detected on average 62.5% blasts (range 23%-84%). CNA plots for all CN-AML samples are shown in Supplementary Fig. 3. Routine diagnostics data of CN-AML samples with detected CNAs are summarized in Supplementary Table 5. Validation of CNA findings in CN-AML samples by FISH was possible only for three out of 14 CNAs due to limited availability of appropriate material and FISH probes, with one CNA confirmed. Therefore, we performed SNP array profiling of five CN-AML samples with CNA findings and confirmed nine out of eleven alterations (82%) (Table 3). CNAs detected with WES and SNP arrays show a very close correlation with regards to start/stop positions (median deviation 44 kb, range 5 to 626 kb) and copy number state (median deviation 0.13 copy numbers, range 0.01 to 0.43) (Fig. 4).

Comparative Assessment of AML del(9q) by SNP Array and Targeted Sequencing

SNP array profiling of five AML del(9q) patients with variable aberrant clone sizes (range 8%-95%, P21-P25) resulted in consistent findings with CNA calling based on WES, as both approaches were able to detect a deletion on the long arm of chromosome 9 in two out of five patients. In one patient (P24), the SNP array showed a reduced copy number count of 1.8 in a region of 33.49 Mb starting at the centromere ranging from chr9:71,007,230 to chr9:104,500,266 suggesting a heterozygous deletion on the long arm of chromosome 9 in 40% of the cells (Fig. 5A). The size of the deletion as well as the change in copy number was consistent with the CNA called using WES extending from chr9:71,080,009 to

chr9:105,757,679 with a clone size of 38% (Fig. 5B). In another patient (P25), the SNP array detected a deletion on chromosome 9 ranging from genomic positions chr9:71,007,230 to chr9:88,961,400 (17.95 Mb) affecting 80% of the cells (Fig. 5C). The size of the deletion and the copy number change was consistent with results from WES with a CNA called at chr9:71,080,009-88,968,114 affecting 84% of the cells (Fig. 5D). Patient samples without any CNA detected by WES also did not show any detectable CNA in SNP array profiling.

Targeted GPS and subsequent CNA calling resulted in the detection of the 9q deletion in 18 out of 22 patients (82%; Fig. 1). Patients already used for WES and SNP array profiling show highly consistent results in CNA calling based on GPS. In patients P21-P23, no del(9q) could be detected, patient P24 showed a heterozygous deletion on chromosome 9 with a mean clone size of 45%, and patient P25 showed a heterozygous deletion on chromosome 9 with a mean clone size of 89%. In all patients positive for del(9q) by GPS, the deletion affected the exons of all genes located on the q arm of chromosome 9 that were included in the gene panel, except for two genes located near the telomere (*DBC1* and *NOTCH1*). Thus, we detected a CDR of at least 8.4 Mb ranging from genomic positions chr9:79,229,486 to chr9:87,636,352 (Fig. 6). Interestingly, in seven AML patients several exons of neighboring genes (*GKAP1*, *KIF27*, *C9ORF64*, and *HNRNPK*) within the CDR show a relatively lower decrease of gene copy counts compared with surrounding genes, suggesting a partial rescue from the deletion. Results of CNA calling based on GPS for individual samples are displayed in Supplementary Fig. 4. Similar to the results from WES, AML samples negative for del(9q) according to GPS had significantly smaller clones harboring the deletion detected by chromosomal analysis (mean 18%, range 8%-31%) compared with AML samples positive for del(9q) according to GPS (mean 81%, range 24%-100%; P=0.0017, Mann-Whitney test; Fig. 3B).

As an additional control, we compared each of the test samples against each remaining test samples. In this analysis, no CNA was called on chromosome 9. Furthermore, we tested

each of the control samples against all other control samples. Interestingly, this resulted in the identification of two AML samples within the control cohort to be del(9q)-positive based on CNA calling by GPS (Table 3). One sample (C67) was diagnosed with a translocation of chromosomes 3 and 21, based on 15 metaphases in routine cytogenetics and had a blast count of 70%. CNA calling by GPS revealed a deletion on chromosome 9 ranging from chr9:79,229,486 to chr9:87,636,352 with a clone size of 74%. In a second sample (C69), we detected a deletion on chromosome 9 ranging from chr9:82,188,604 to chr9:86,617,779 with a clone size of 68%. Routine diagnostics revealed a highly complex karyotype with more than 20 aberrations based on a low number of metaphases (n=8) as well as a low blast cell count (30%). Both CNAs were confirmed by SNP array profiling. Results of CNA calling based on GPS in control samples are displayed in Supplementary Fig. 5 and routine diagnostics data of control samples with detected CNAs by GPS is summarized in Supplementary Table 5.

DISCUSSION

Our study shows that targeted sequencing approaches can be efficiently used not only to identify somatic mutations with single nucleotide resolution, but also to detect somatic CNAs in AML. Exome sequencing offers the possibility to detect CNAs on any chromosome with high resolution. We were able to confirm findings from routine cytogenetics such as aneuploidies and partial deletions of chromosomes. AML is particularly well-suited to compare NGS-based CNA calling with diagnostic results, as conventional cytogenetics is essential in clinical evaluation of hematologic malignancies.

Although the low detection rate of CNA in low coverage exons sequenced on an Illumina Genome Analyzer IIx suggests that increasing coverage facilitates CNA calling in exome data, the most important confounder for false negative results was the clonal architecture of the tumor sample and the size of the clone harboring the CNA. Given an adequate clone

size, CNAs could be detected in exomes from both sequencing instruments, and the impact of the overall coverage or the proportion of low-coverage target regions was rather low.

By comparing exomes from initial diagnosis or relapse of AML to exomes from matching germline control samples obtained at complete remission we were able to confirm the results of routine cytogenetics if the alteration was observed in more than one third of the cells. Samples in which we could not detect previously reported chromosomal aberrations either had a reduced tumor load or smaller subclones harboring the alteration as estimated from the number of aberration positive metaphases and/or interphase nuclei. Alterations in clones representing less than 30% of the sample were detected in about half of the cases. Furthermore, we observed a close correlation of clone size estimation based on routine diagnostics and the clone size estimation resulting from CNA calling using both WES or GPS data (R=0.85) and an average difference of the estimated clone size of 15% (range 0%-47%; Fig. 1).

The highly correlated results obtained with both WES and SNP array profiling for a set of five AML del(9q) patients with varying clonal architecture demonstrate that our NGS based approach is as sensitive and precise as array based techniques for genome wide CNA calling. Despite the experimental variability of target enrichment in WES, we confirmed the size, position, and frequency of deletions on chromosome 9 in two out of five patients with both methods. Apparently, the clonal architecture of the tumor and the size of the CNA harboring clone are limiting the sensitivity of both methods.

As matched germline control samples from complete remission or from other tissues are not necessarily available for every AML sample, especially at the time of initial diagnosis, during therapy, or in refractory AML, we tested CNA calling in sequencing data without germline control samples. Although a germline control specimen might be sampled by taking a buccal swab or a skin biopsy, these procedures add to the workload in clinical routine and may cause unnecessary stress for the patient. Again, it is quite obvious that clonal

architecture is limiting the power to detect CNAs using GPS. All samples with cytogenetically detected deletions but negative for CNAs by GPS had an estimated clone size of <25%. Moreover, GPS revealed highly consistent results compared with WES and SNP array profiling in all five patients analyzed by these methods. We were able to show that even without a matched germline control sample, targeted sequencing of CNA hotspots (e.g. long arm of chromosome 9) is a versatile method to detect chromosomal deletions. We used a selected control cohort of AML patients reported as wild type for the specific allele, but beyond that, a unified and standardized control, either pooled from an adequate number of healthy individuals or derived in silico, is most likely sufficient to detect alterations with high specificity.

The detection of additional deletions, duplications and trisomies in ten AML samples (cytogenetically aberrant, n=4; CN-AML, n=6) using WES, as well as the detection of deletions on chromosome 9 in del(9q)-negative AML samples (n=2) using GPS points towards limitations of conventional cytogenetics. FISH and SNP array analyses revealed a high confirmation rate of CNAs missed by routine cytogenetics (15/17, 88%). Validation could not be performed for four CNAs due to inadequate sample material and/or FISH probe availability. Two of these CNAs in a patient's relapse sample with insufficient material for validation were confirmed in the corresponding diagnostic sample, suggesting additional true positive calls as identical loci were altered in both samples. Two partial duplications of chromosome Y could not be confirmed. The sequence homology between chromosomes X and Y, pseudo-autosomal regions, and the increased amount of repetitive sequences on chromosome Y (Skaletsky et al., 2003) might increase the risk of false positive CNA calling on sex chromosomes.

It remains unclear why routine diagnostics missed certain CNAs. The median size of additional alterations detected by WES was 39.0 Mb and the coverage ratios indicated aberrations of the dominant AML clone in 90% of the AML samples. Highly complex

karyotypes as well as low numbers of available metaphases and/or low tumor load may hamper the detection of subchromosomal alterations in routine cytogenetics. In addition, the potential heterogeneity of a patient's AML may result in different findings when comparing different specimens from the same patient (e.g. consecutively aspirated EDTA-bone marrow for DNA extraction and heparin-bone marrow for cytogenetics). Further, the mitotic index of the malignant clone might be lower than that of the normal bone marrow cells resulting in aberrant clone estimates based on conventional cytogenetics. On the other hand, the mitotic index of the aberrant clone might be higher resulting in an increased sensitivity of conventional cytogenetics to detect small subclones harboring CNAs. However, single cell resolution lends unique characteristics to cytogenetics, while CNA calling based on standard NGS applications so far is limited to bulk sample profiling. Single-cell sequencing protocols have recently been established (Zong et al., 2012). As the cost for sequencing is still falling at an exponential rate, single cell technology will open up new avenues to capture the clonal diversity of tumor cell populations in the future. Currently, WES and GPS offer fast turnaround times and represent cost-effective approaches to characterize tumor DNA. However, such targeted sequencing approaches are commonly focused on protein coding regions, and, thus, do not capture alterations in non-coding regions of the genome. Therefore targeted NGS based techniques may complement established cytogenetics in order to characterize leukemia samples more comprehensively in the future.

Estimation of the tumor architecture based on metaphase count and blast count remains challenging as the overall number of metaphases evaluated per patient is rather low (20 to 30). In addition, specimens used for sequencing commonly undergo purification by Ficoll gradient centrifugation further limiting the comparability of clone size estimations based on sequencing and cytogenetics. The increased tumor load in AML samples prepared for NGS is obvious when comparing clone size estimations based on WES and routine data. As expected, the clone size reported by routine diagnostics is negatively correlated with the

effect of amplification by Ficoll gradient centrifugation. Smaller subclones generally show a stronger increase in size while dominant AML clones tend to be slightly overrepresented in routine karyotyping (Fig. 1). AML samples harboring the alteration in the primary clone show highly correlated clone size estimations using both methods, while smaller subclones harboring an alteration might be missed in NGS-based CNA calling. This is similar to the sensitivity of Sanger sequencing as the standard method for the identification of sequence mutations in routine diagnostics (Chin et al., 2013). Nevertheless, we found a high correlation between clone size estimation based on sequence coverage difference in NGS samples and conventional cytogenetics and/or FISH.

Complex karyotypes with more than three cytogenetic aberrations affect 10%-12% of all AML patients and are strongly associated with adverse outcome (Mrózek, 2008). Isolated copy number alterations, e.g. on chromosome 5 and 7, also have prognostic significance (Grimwade et al., 2010). So far, aneuploidy and specific gene fusions (e.g. *PML*/*RARA*) are screened for in routine diagnostics using specific FISH probes (Fröhling et al., 2002; Lugthart et al., 2008). Targeted sequencing of point mutation hotspots and recurrent chromosomal alterations with prognostic significance in a single step using a customized AML gene panel may reduce costs, time to diagnosis, and even manpower in routine diagnostic laboratories in the future. As we have shown, this will be possible even without the requirement of matched germline control samples. The variable coverage between different target regions could be viewed as an obstacle to detect CNAs. However, the variable coverage mainly depends on the enrichment efficiency, which is similar in normal and aberrant samples and thus it should be relatively straightforward to establish a standard enrichment profile as control.

In summary, we show that CNA detection in targeted sequencing data may complement classical and molecular cytogenetics in routine diagnostics, especially for the detection of autosomal CNAs in the dominant AML clone. Our approach enables us to detect alterations

missed by routine diagnostics and the upscalable coverage of gene panel sequencing may facilitate the detection of CNAs even in small subclones of a heterogeneous tumor cell population. However, these approaches are limited to a defined target region. While the present study demonstrates the detection of known chromosomal aberrations by targeted sequencing as a proof of concept, the detection and confirmation of small CNAs below the resolution of cytogenetics in sequencing data is a challenge ahead.

ACKNOWLEDGMENTS

The authors thank Bianka Ksyenzik and Kathrin Bräundl for technical support and the Genomics and Proteomics Core Facility at the German Cancer Research Center (DKFZ) for sequencing services. We thank Michael Bonin and Alexandra Dangel (IMGM laboratories) for SNP array analysis.

REFERENCES

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M. 2010. The landscape of somatic copy-number alteration across human cancers. Nature 463:899-905

Bolli N, Manes N, McKerrell T, Chi J, Park N, Gundem G, Quail MA, Sathiaseelan V, Herman B, Crawley C, Craig JI, Conte N, Grove C, Papaemmanuil E, Campbell PJ, Varela I, Costeas P, Vassiliou GS. 2015. Characterization of gene mutations and copy number changes in acute myeloid leukemia using a rapid target enrichment protocol. Haematologica 100:214-222

Cancer Genome Atlas Research Network. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 368:2059-2074

Carter NP. 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 39:S16-21

Chin EL, da Silva C, Hegde M. 2013. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. BMC Genet 14:6

Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC,

Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature 481:506-510

Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz MA, Sierra J, Tallman MS, Löwenberg B, Bloomfield CD; European LeukemiaNet. 2010. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood 115:453-474

Farag SS, Archer KJ, Mrózek K, Ruppert AS, Carroll AJ, Vardiman JW, Pettenati MJ, Baer MR, Qumsiyeh MB, Koduru PR, Ning Y, Mayer RJ, Stone RM, Larson RA, Bloomfield CD. 2006. Pretreatment cytogenetics add to other prognostic factors predicting complete remission and long-term outcome in patients 60 years of age or older with acute myeloid leukemia: results from Cancer and Leukemia Group B 8461. Blood 108:63-73

Fröhling S, Skelin S, Liebisch C, Scholl C, Schlenk RF, Döhner H, Döhner K; Acute Myeloid Leukemia Study Group, Ulm. 2002. Comparison of cytogenetic and molecular cytogenetic detection of chromosome abnormalities in 240 consecutive adult patients with acute myeloid leukemia. J Clin Oncol 20:2480-2485

Fröhling S, Schlenk RF, Kayser S, Morhardt M, Benner A, Döhner K, Döhner H; German-Austrian AML Study Group. 2006. Cytogenetics and age are major determinants of outcome in intensively treated acute myeloid leukemia patients older than 60 years: results from AMLSG trial AML HD98-B. Blood 108:3280-3288

Gelehrter TD, Collins FS, Ginsburg D. 1998. Principles of medical genetics. 2nd edition. Baltimore: Williams & Wilkins. 153-194

Ghanem H, Tank N, Tabbara IA. 2012. Prognostic implications of genetic aberrations in acute myelogenous leukemia with normal cytogenetics. Am J Hematol 87:69-77

Greif PA, Dufour A, Konstandin NP, Ksienzyk B, Zellmeier E, Tizazu B, Sturm J, Benthaus T, Herold T, Yaghmaie M, Dörge P, Hopfner KP, Hauser A, Graf A, Krebs S, Blum H, Kakadia PM, Schneider S, Hoster E, Schneider F, Stanulla M, Braess J, Sauerland MC, Berdel WE, Büchner T, Woermann BJ, Hiddemann W, Spiekermann K, Bohlander SK. 2012. GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. Blood 120:395-403

Grimwade D. 2001. The clinical significance of cytogenetic abnormalities in acute myeloid leukaemia. Best Pract Res Clin Haematol 14:497- 529

Grimwade D, Walker H, Harrison G, Oliver F, Chatters S, Harrison CJ, Wheatley K, Burnett AK, Goldstone AH; Medical Research Council Adult Leukemia Working Party. 2001. The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. Blood 98:1312-1320

Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, Wheatley K, Harrison CJ, Burnett AK; National Cancer Research Institute Adult Leukaemia Working Group. 2010. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood 116:354-365

Haferlach C, Rieder H, Lillington DM, Dastugue N, Hagemeijer A, Harbott J, Stilgenbauer S, Knuutila S, Johansson B, Fonatsch C. 2007. Proposals for standardized protocols for

cytogenetic analyses of acute leukemias, chronic lymphocytic leukemia, chronic myeloid leukemia, chronic myeloproliferative disorders, and myelodysplastic syndromes. Genes Chromosomes Cancer 46:494-499

Herold T, Metzeler KH, Vosberg S, Hartmann L, Röllig C, Stölzel F, Schneider S, Hubmann M, Zellmeier E, Ksienzyk B, Jurinovic V, Pasalic Z, Kakadia PM, Dufour A, Graf A, Krebs S, Blum H, Sauerland MC, Büchner T, Berdel WE, Woermann BJ, Bornhäuser M, Ehninger G, Mansmann U, Hiddemann W, Bohlander SK1, Spiekermann K, Greif PA. 2014. Isolated trisomy 13 defines a homogeneous AML subgroup with high frequency of mutations in spliceosome genes and poor prognosis. Blood 124:1304-1311

Ilyas AM, Ahmad S, Faheem M, Naseer MI, Kumosani TA, Al-Qahtani MH, Gari M, Ahmed F. 2015. Next generation sequencing of acute myeloid leukemia: influencing prognosis. BMC Genomics 16:S5

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res 12:996-1006

Lugthart S, van Drunen E, van Norden Y, van Hoven A, Erpelinck CA, Valk PJ, Beverloo HB, Löwenberg B, Delwel R. 2008. High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. Blood 111:4329-4337

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303

Mrózek K. 2008. Cytogenetic, molecular genetic, and clinical characteristics of acute myeloid leukemia with a complex karyotype. Semin Oncol 35:365-377

Mrózek K, Heerema NA, Bloomfield CD. 2004. Cytogenetics in acute leukemia. Blood Rev

18:115-136

Opatz S, Polzer H, Herold T, Konstandin NP, Ksienzyk B, Zellmeier E, Vosberg S, Graf A, Krebs S, Blum H, Hopfner KP, Kakadia PM, Schneider S, Dufour A, Braess J, Sauerland MC, Berdel WE, Büchner T, Woermann BJ, Hiddemann W, Spiekermann K, Bohlander SK, Greif PA. 2013. Exome sequencing identifies recurring FLT3 N676K mutations in core-binding factor leukemia. Blood 122:1761-1769

Rausch T, Jones DT, Zapatka M, Stütz AM, Zichner T, Weischenfeldt J, Jäger N, Remke M, Shih D, Northcott PA, Pfaff E, Tica J, Wang Q, Massimi L, Witt H, Bender S, Pleier S, Cin H, Hawkins C, Beck C, von Deimling A, Hans V, Brors B, Eils R, Scheurlen W, Blake J, Benes V, Kulozik AE, Witt O, Martin D, Zhang C, Porat R, Merino DM, Wasserman J, Jabado N, Fontebasso A, Bullinger L, Rücker FG, Döhner K, Döhner H, Koster J, Molenaar JJ, Versteeg R, Kool M, Tabori U, Malkin D, Korshunov A, Taylor MD, Lichter P, Pfister SM, Korbel JO. 2012. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 148:59-71

Rigaill GJ, Cadot S, Kluin RJ, Xue Z, Bernards R, Majewski IJ, Wessels LF. 2012. A regression model for estimating DNA copy number applied to capture sequencing data. Bioinformatics 28:2357-2365

Shaffer LG, Slovak ML, Campbell, LJ. 2009. An International System for Human Cytogenetic Nomenclature: Recommendations of the International Standing Committee on Human Cytogenetic Nomenclature (Karger Publishers, Berlin, Germany)

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK,

Rozen S, Page DC. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837

Speicher MR, Gwyn Ballard S, Ward DC. 1996. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet 12:368-375

Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW. 2008. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, 4th ed. Lyon, France: IARC Press

Walter MJ, Payton JE, Ries RE, Shannon WD, Deshmukh H, Zhao Y, Baty J, Heath S, Westervelt P, Watson MA, Tomasson MH, Nagarajan R, O'Gara BP, Bloomfield CD, Mrózek K, Selzer RR, Richmond TA, Kitzman J, Geoghegan J, Eis PS, Maupin R, Fulton RS, McLellan M, Wilson RK, Mardis ER, Link DC, Graubert TA, DiPersio JF, Ley TJ. 2009. Acquired copy number alterations in adult acute myeloid leukemia genomes. Proc Natl Acad Sci U S A 106:12950-12955

Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. 2011. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet 43:309-315

Zong C, Lu S, Chapman AR, Xie XS. 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338:1622-1626

TABLE 1. Routine Diagnostics Data of Patients Used for Exome Sequencing and Results from CNA Calling

ID	% BC	Routine cytogenetics	Routine FISH	CNA by WES	CN change
P1	95%	47,XY,+4[19]/46,XY[1]	-	chr4:53,383-191,013,434[c]	+0.82
P2[d]	27%	47,XY,+8[10]/46,XY[15]	nuc ish 8cen(D8Z2x3)[7/108]	chr8:116,086-146,279,543[c]	+0.30
P3[d]	35%	-	nuc ish 8cen(D8Z2x3)[35/105]	chr8:116,086-146,279,543[c]	+0.33
P4[e]	60%	47,XY,+8[18]/46,XY[7]	nuc ish 8cen(D8Z2x3)[87/108]	chr8:116,086-146,279,543[c]	+0.95
P5[e]	60%	-	nuc ish 8cen(D8Z2x3)[86/130]	chr8:116,086-139,688,878[a]	+0.19
P6[f]	90%	47,XY,+8[4]/46,XY[17]	nuc ish 8cen(D8Z2x3)[10/100]	-	-
P7[f]	84%	47,XY,+8[12]/46,XY[13]	nuc ish 8cen(D8Z2x3)[4/109]	chr8:116,086-146,279,543[c]	+0.59
P8	94%	47,XY,+8[23]/46,XY[2]	nuc ish 8cen(D8Z2x3)[90/105]	chr8:116,086-146,279,543[c]	+0.93
P9	NA	47,XY,+8[8]/46,XY[6]	-	chr8:116,086-146,279,543[c] chr11:30,032,264-43,333,746	+0.80 -0.77
P10	60%	47,XY,+13[4]/46,XY[16]	nuc ish 13q14(RBx3)[7/100]	-	-
P11	90%	47,XY,+13[18]/46,XY[3]	nuc ish 13q14(RBx3)[72/100]	chr13:19,748,003-115,091,756[c]	+0.88
P12	57%	46,XY,t(9;11)(p22;q23)[4]/ 47,idem,+21[17]/46,XY[4]	nuc ish 21q22.13-q22.2(D21S342, D21S341,D21S259x2)[59/109]	chr21:10,906,905-48,084,286[c]	+0.66
P13	70%	45,XY,-7[15]/46,XY[10]	-	chr7:193,200-158,937,463[c]	-0.74
P14	90%	46,XY,t(9;11)(p22;q23)[15]/46,XY[5]	nuc ish 11q23(MLL5'x2,MLL3'x1)[98/100]	chr11:118,359,329-119,291,096	-0.95
P15	65%	46,XY,del(5)(q13q31)[13]/46,XY[10]	nuc ish 5p15.2(D5S721,D5S23x2), 5q31(EGR1x1)[61/115]	chr5:136,957,787-170,648,844	-0.78
P16[g]	98%	46,XY,t(9;22)(q34;q11),del(20)(q11)[25]	nuc ish 20q13.2(ZNF217x1)[108/128]	chr7:119,914,687-158,937,463[b] chr13:72,049,273-115,091,756[b] chr20:31,022,235-54,579,227	-0.93 +0.99 -0.94
P17[g]	57%		nuc ish 20q13.2(ZNF217x1)[95/102]	chr7:119,914,687-158,937,463[b] chr13:72,014,787-115,091,756[b] chr20:31,022,235-54,579,227	-0.74 +1.00 -0.95
P18[h,i]	82%	46,XY,t(8;12)(p11;q11)[8]/ 46,idem,der(5)t(1;5)(p31;q2?)[13]/ 47,idem,+der(12)t(8;12)(p11;q11)[2]	nuc ish 5p15.2(D5S721,D5S23x2), 5q31(EGR1x1)[75/114]	chr1:190,067,148-249,212,562[b] chr5:85,915,170-180,795,226[b]	+0.68 -0.66
P19[h,i]	91%	46,XY,der(5)t(1;5)(q31;q2?), t(8;12)(p11;q11)[9]/46,XY[4]	nuc ish 5p15.2(D5S721,D5S23x2), 5q31(EGR1x1)[102/112]	chr1:190,067,148-249,212,562[b] chr5:85,913,873-180,795,226[b]	+0.83 -0.90
P20[i]	74%	46,XX,der(18)t(1;18)(p?;q23)[17]/46,XX[2]	-	chr1:149,857,810-249,212,562[b] chr17:6,011-7,951,883[a] chr18:70,205,411-78,005,231[b]	+0.84 -0.87 -0.95
P21	NA	46,XX,del(9)(q22)[2]/46,XX[23]	-	-	-
P22	85%	46,XX,del(9)(q22)[8]/46,XX[18]	-	-	-
P23	90%	46,XX,del(9)(q22q34)[5]/46,XX[20]	-	-	-
P24	NA	46,XX,del(9)(q21q34)[6]/46,XX[16]	-	chr9:71,080,009-105,757,679	-0.38
P25	95%	46,XY,del(9)(q22q34)[19]/46,XY[1]	-	chr9:71,080,009-88,968,114	-0.84

BC, blast count; FISH, fluorescence in situ hybridization; CNA, copy number alteration; WES, whole exome sequencing; CN, copy number.

[a] CNA starts at first exon of chromosome.

[b] CNA stops at last exon of chromosome.

[c] CNA covers the complete chromosome.

[d,e,f,g,h] Samples represent diagnostic and relapse samples from the same patient.

[i] Samples were additionally karyotyped with multiplex FISH.

TABLE 2. Routine Diagnostics Data of AML del(9q) Patients Used for Gene Panel Sequencing and Results From CNA Calling

ID	% BC	Routine cytogenetics	CNA by GPS	CN change
P21	NA	46,XX,del(9)(q22)[2]/46,XX[23]	-	-
P22	85%	46,XX,del(9)(q22)[8]/46,XX[18]	-	-
P23	90%	46,XX,del(9)(q22q34)[5]/46,XX[20]	-	-
P24	NA	46,XX,del(9)(q21q34)[6]/46,XX[16]	chr9:21,970,901-87,636,352	-0.22
P25	95%	46,XY,del(9)(q22q34)[19]/46,XY[1]	chr9:21,968,724-87,636,352	-0.44
P26	80%	48-51,XY,+8,del(9)(q22),+1-4mar[cp17]/46,XY[6]	chr9:82,941,560-87,636,352	-0.39
P27	90%	46,XX,del(9)(q22q34)[16]/46,XX[4]	chr9:82,321,662-86,294,952	-0.31
P28	80%	46,XY,del(9)(q22)[6]/46,XY[19]	chr9:82,268,974-87,367,000	-0.14
P29	80%	46,XX,del(9)(q22q34)[19]/46,XX[1]	chr9:21,970,901-87,636,352	-0.39
P30	NA	46,XY,del(9)(q22q34)[20]	chr9:79,229,486-87,486,732	-0.37
P31	50%	46,XY,del(9)(q11q32)[19]/46,XY[1]	chr9:21,970,901-87,636,352	-0.34
P32	90%	46,XX,del(9)(q21)[20]	chr9:21,970,901-87,636,352	-0.39
P33	90%	47,XX,+6,del(9)(q21)[20]	chr9:79,229,486-87,636,352	-0.44
P34	95%	46,XX,del(9)(q21q34)[15]	chr9:79,253,102-87,486,732	-0.37
P35	75%	46,XY,del(9)(q13q32)[13]/47,XY,+21[5]/46,XY[2]	chr9:79,229,486-87,636,352	-0.26
P36	NA	47,XY,+10[7]/46,XY,del(9)(q22)[3]/46,XY[11]	-	-
P37	25%	46,XY,t(8;21)(q22;q22),del(9)(q22)[20]	chr9:79,229,486-86,354,659	-0.37
P38	95%	46,XX,del(9)(q22)[4]/46,XX[13]	-	-
P39	NA	46,XY,del(9)(q22q32)[22]/46,XY[3]	chr9:79,229,486-87,636,352	-0.36
P40	25%	46,XY,t(8;21)(q22;q22),del(9)(q22)[18]/46,XY[2]	chr9:79,229,486-87,636,352	-0.33
P41	NA	46,XY,t(8;21)(q22;q22),del(9)(q22)[20]	chr9:79,229,486-86,308,773	-0.36
P42	75%	46,XY,t(8;21)(q22;q22)[1]/ 46,idem,del(9)(q22)[12]/ 45,idem,-Y,del(9)(q22)[4]	chr9:79,465,379-86,593,167	-0.35

BC, blast count; CNA, copy number alteration; GPS, gene panel sequencing; CN, copy number.

TABLE 3. CNAs Exclusively Called Using Targeted Sequencing Approaches and Validation by FISH and/or SNP Array Profiling

ID	CNA by NGS	CN change by NGS	FISH	CNA by SNP array	CN change by SNP array
P9[a]	chr11:30,032,264-43,333,746	-0.77	NA	NA	NA
P16[a,d]	chr7:119,914,687-158,937,463[f]	-0.93	nuc ish 7q22(CUTL1x2), 7q36(D7S2419,D7S688, D7S2640x1)[98/111]	NA	NA
	chr13:72,049,273-115,091,756[f]	+0.99	nuc ish 13q14 (D13S319x2,D13S25x2), 13q34(D13S1825x3)[106/132]	NA	NA
P17[a,d]	chr7:119,914,687-158,937,463[f]	-0.74	NA	NA	NA
	chr13:72,014,787-115,091,756[f]	+1.00	NA	NA	NA
P20[a]	chr1:149,857,810-249,212,562[f]	+0.84	nuc ish 1p36(EGFL3, TP73x2),1q25(ANGPTL1, ABL2x3)[108/116]	NA	NA
	chr17:6,011-7,951,883[e]	-0.87	nuc ish 17p13(TP53x1), 17cen(D17Z1x2)[96/114]	NA	NA
C5[b]	chrY:150,855-5,605,983[e]	+0.69	nuc ish Xcen(DXZ1x1), Yp11.3(SRYx1)[131]	-	-
C6[b]	chrY:150,855-9,368,097[e]	+0.75	nuc ish Xcen(DXZ1x1), Yp11.3(SRYx1)[163]	-	-
C14[b]	chr21:19,732,074-26,960,101	+0.47	NA	NA	NA
C15[b]	chr1:168,032,859-249,212,562[f]	+0.91	NA	chr1:168,027,535-249,224,684	+1.0
	chr6:73,713,631-114,179,023	-0.67	NA	chr6:73,348,560-113,715,029	-0.66
	chr7:193,200-8,791,399[e]	-0.84	NA	chr7:43,360-8,878,169	-0.87
	chr7:102,937,907-129,520,811	-0.87	NA	chr7:102,928,126-129,565,937	-1.0
	chr7:143,771,313-158,937,463[f]	-0.81	NA	chr7:143,748,883-159,119,707	-1.0
	chr16:71,094,407-90,142,318[f]	-0.84	NA	chr16:71,071,488-90,155,062	-1.0
	chr20:42,088,411-50,273,689	-0.57	NA	chr20:42,045,797-50,250,884	-1.0
C24[b]	chr6:73,713,631-139,695,089	-0.53	nuc ish 6q21(SEC63x1), 6q23(MYBx1)[38/110]	chr6:73,087,915-140,071,912	-0.57
C37[b]	chr8:116,086-146,279,543[g]	+0.65	NA	chr8:158,048-145,798,536	+1.0
C67[c]	chr9:79,229,487-87,636,352	-0.37	NA	chr9:68,665,170-110,894,060	-1.0
C69[c]	chr9:82,188,604-86,617,779	-0.29	NA	chr9:79,971,000-86,921,410	-0.68

CNA, copy number alteration; NGS, next generation sequencing; CN, copy number; FISH, fluorescence in situ hybridization; SNP, single nucleotide polymorphism; NA, not available due to limited availability of appropriate material/FISH probe.

[a] Test patients for exome sequencing diagnosed as cytogenetically aberrant.

[b] Control patients for exome sequencing diagnosed as cytogenetically normal.

[c] Control patients for gene panel sequencing diagnosed as del(9q)-negative.

[d] Samples represent diagnostic and relapse sample from the same patient.

[e] CNA starts at first exon of chromosome.

[f] CNA stops at last exon of chromosome.

[g] CNA covers the complete chromosome.

FIGURE 1: Results of CNA detection using WES, GPS, SNP array profiling and comparison of clone size estimation with routine diagnostics. Symbols represent patient samples according to the methods used to detect CNAs. WES, whole exome sequencing; GPS, gene panel sequencing; NGS, next generation sequencing.

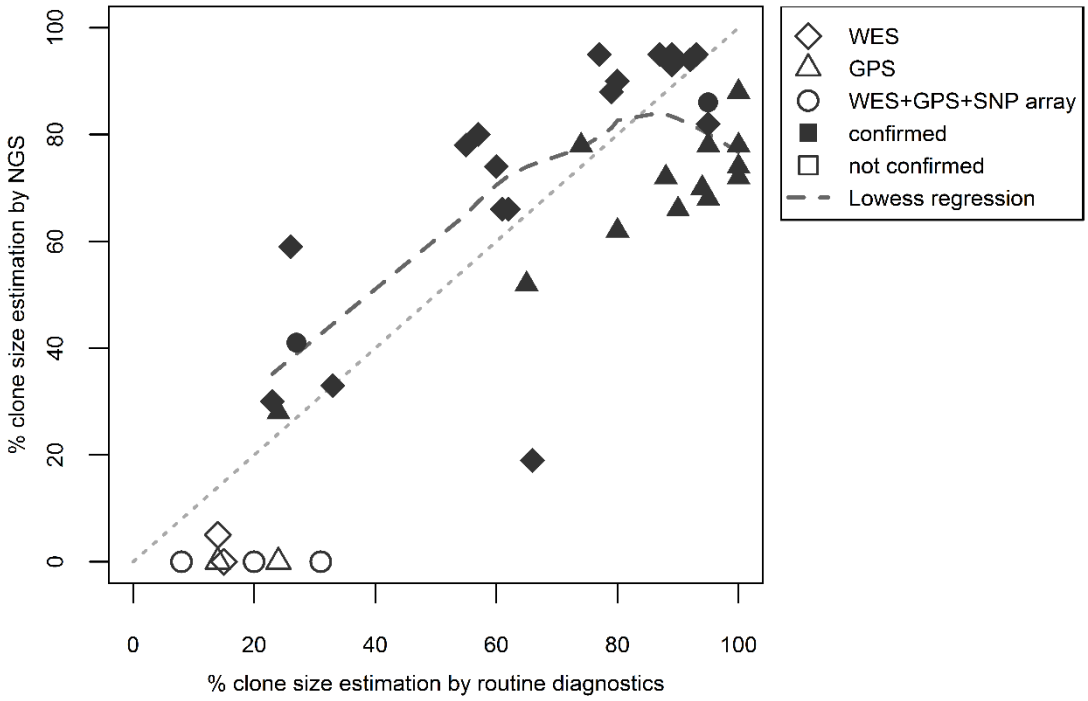

FIGURE 2: Detection of copy number alterations by exome sequencing. (A) Trisomy 8 (P8); (B) Partial deletion of the long arm of chromosome 5 (P15); (C) Deletion on the long arm of chromosome 11 including partial deletion of the 3' part of the *KMT2A* gene and downstream exons (P14). Chromosomal regions are plotted on the X-axis, size of chromosomes is proportional to exon count. Sequence coverage ratio of leukemia and remission samples is indicated on the Y-axis and displayed by black bars. Shading of chromosomal regions depends on copy number change. Horizontal dashed lines, gain or loss of one gene copy in 100% of cells; Vertical dotted lines, chromosome boundaries.

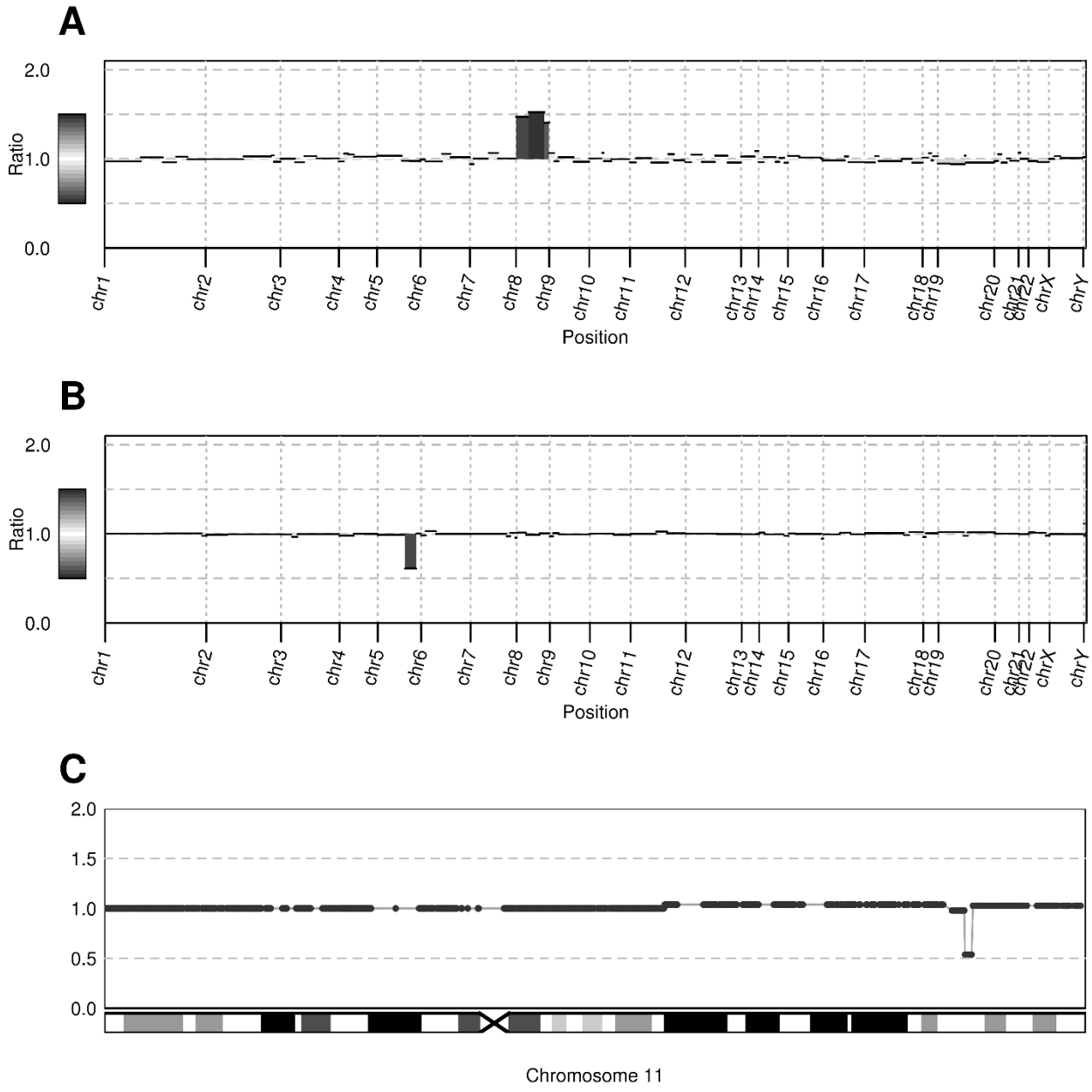

FIGURE 3: Comparison of clone sizes estimated by routine diagnostics in AML samples with confirmation status of CNAs by targeted sequencing. (A) AML samples used for exome sequencing. (B) AML samples used for gene panel sequencing.

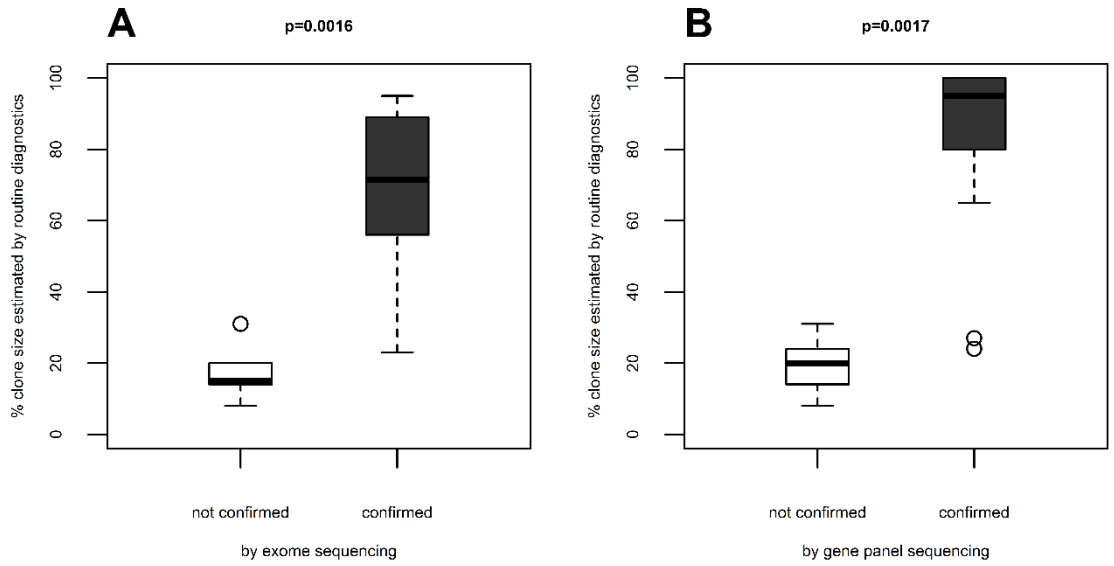

FIGURE 4: Validation of CNAs detected in cytogenetically normal AML samples by exome sequencing using SNP array profiling. Chromosome positions are plotted on the X-axis. Copy number state is plotted on the Y-axis. Dashed gray line, CNAs called by WES; dotted black line, CNAs called by SNP array profiling.

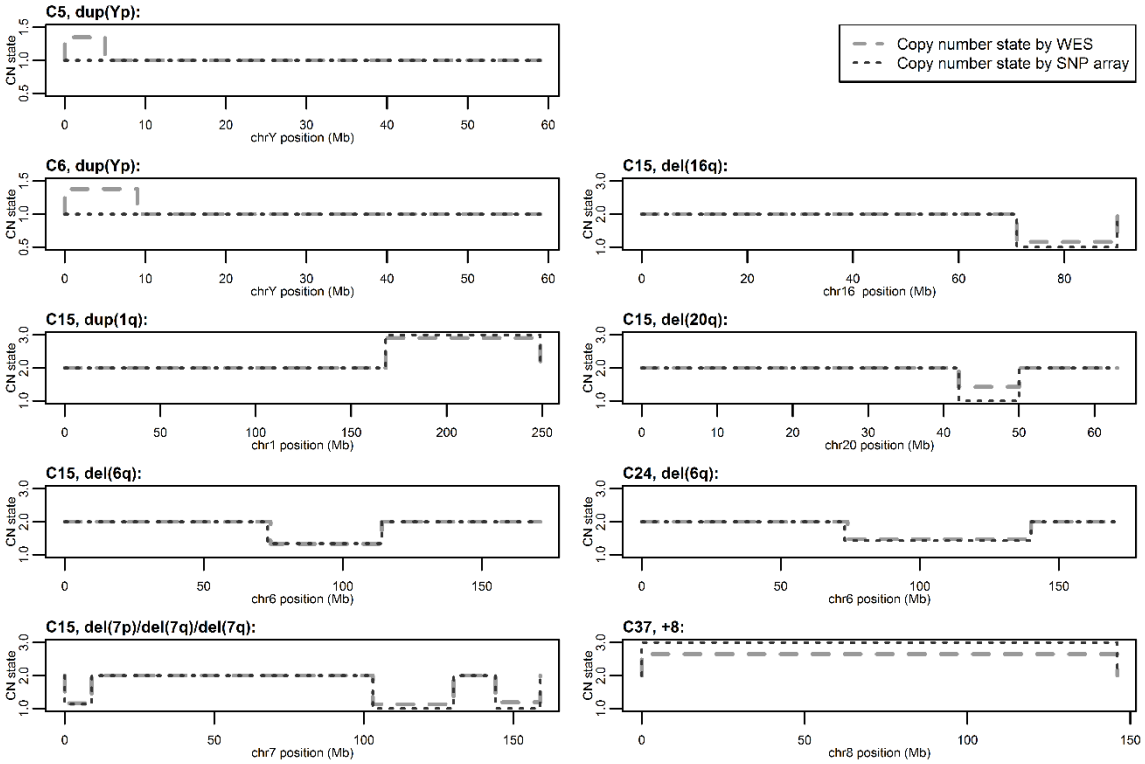

FIGURE 5: Detection of a partial deletion on the long arm of chromosome 9 using WES and SNP array profiling. (A)/(B): patient P24; (C)/(D): patient P25. (A)/(C): CNA calling based on SNP array profiling; (B)/(D): CNA calling using WES data.

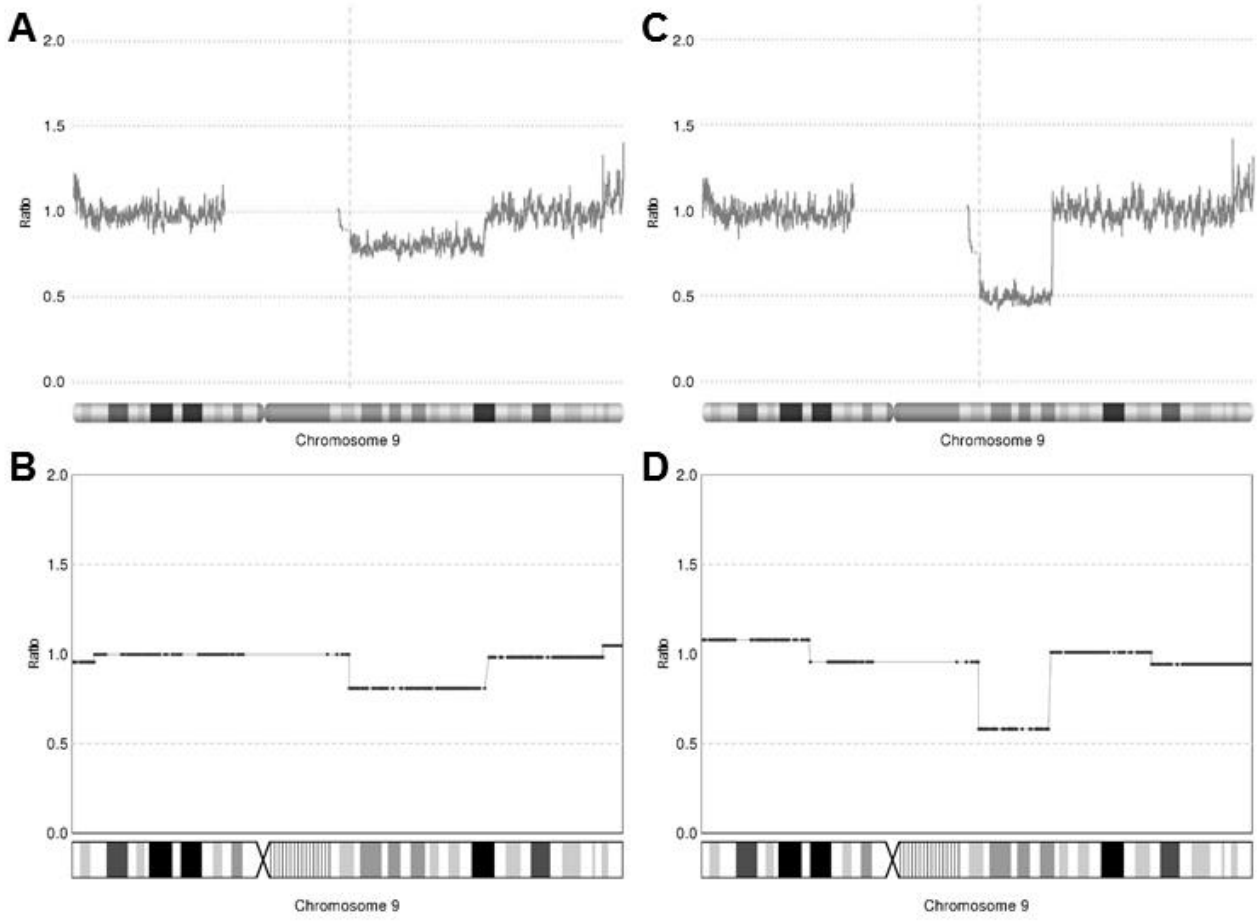

FIGURE 6: Detection of a commonly deleted region in AML del(9q) patients using gene panel sequencing. Upper part: black dots, mean coverage ratio values; grey area, CDR. Lower part: genes in CDR with detailed coverage ratio distribution.