



OPEN ACCESS

ORIGINAL ARTICLE

Accurate quantification of chromosomal lesions via short tandem repeat analysis using minimal amounts of DNA

Johann-Christoph Jann,¹ Daniel Nowak,¹ Florian Nolte,¹ Stephanie Fey,¹ Verena Nowak,¹ Julia Obländer,¹ Jovita Pressler,¹ Iris Palme,¹ Christina Xanthopoulos,¹ Alice Fabarius,¹ Uwe Platzbecker,² Aristoteles Giagounidis,³ Katharina Götze,⁴ Anne Letsch,⁵ Detlef Haase,⁶ Richard Schlenk,⁷ Gesine Bug,⁸ Michael Lübbert,⁹ Arnold Ganser,¹⁰ Ulrich Germing,¹¹ Claudia Haferlach,¹² Wolf-Karsten Hofmann,¹ Maximilian Mossner¹

► Additional material are available. To view these files please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2017-104528>).

For numbered affiliations see end of article.

Correspondence to

Maximilian Mossner, Department of Hematology and Oncology, University Hospital Mannheim, Pettenkoferstraße 22, Mannheim 68169, Germany; maximilian.mossner@medma.uni-heidelberg.de; m.mossner@qmul.ac.uk

Received 10 January 2017
Revised 15 February 2017
Accepted 22 March 2017
Published Online First
9 June 2017

ABSTRACT

Background Cytogenetic aberrations such as deletion of chromosome 5q (del(5q)) represent key elements in routine clinical diagnostics of haematological malignancies. Currently established methods such as metaphase cytogenetics, FISH or array-based approaches have limitations due to their dependency on viable cells, high costs or semi-quantitative nature. Importantly, they cannot be used on low abundance DNA. We therefore aimed to establish a robust and quantitative technique that overcomes these shortcomings.

Methods For precise determination of del(5q) cell fractions, we developed an inexpensive multiplex-PCR assay requiring only nanograms of DNA that simultaneously measures allelic imbalances of 12 independent short tandem repeat markers.

Results Application of this method to n=1142 samples from n=260 individuals revealed strong intermarker concordance ($R^2=0.77-0.97$) and reproducibility (mean SD: 1.7%). Notably, the assay showed accurate quantification via standard curve assessment ($R^2>0.99$) and high concordance with paired FISH measurements ($R^2=0.92$) even with subnanogram amounts of DNA. Moreover, cytogenetic response was reliably confirmed in del(5q) patients with myelodysplastic syndromes treated with lenalidomide. While the assay demonstrated good diagnostic accuracy in receiver operating characteristic analysis (area under the curve: 0.97), we further observed robust correlation between bone marrow and peripheral blood samples ($R^2=0.79$), suggesting its potential suitability for less-invasive clonal monitoring.

Conclusions In conclusion, we present an adaptable tool for quantification of chromosomal aberrations, particularly in problematic samples, which should be easily applicable to further tumour entities.

5 (del(5q)) are the most commonly observed aberrations found in approximately 30% of affected patients with chromosomal abnormalities.²

The current gold standard for karyotypic stratification of patients with MDS is represented by chromosomal banding analysis (metaphase cytogenetics (MC)), which provides an overview of the whole karyotype without prior knowledge.³ However, major drawbacks of this method are the requirement for viable cells capable of cell division and a potential bias in the cytogenetic profile due to highly proliferative subclones. Moreover, despite using optimised protocols, MC analyses fail in approximately 5% of cases.^{2,4}

Fluorescence in situ hybridization (FISH) uses fluorescently labelled DNA probes for detection of chromosomal alterations in interphase cells. However, the method requires prospectively archived cells for analysis and is relatively expensive and labour-intensive.^{5,6} Alternatively, SNP microarrays solely require genomic DNA and are valuable tools for de novo genome-wide screening of (submicroscopic) copy number alterations,⁷⁻⁹ but are highly expensive and allow only semi-quantitative characterisation.

In order to overcome the limitations of the described methods, we developed a novel technique based on the assessment of allelic loss at heterozygous short tandem repeat (STR) markers. These represent repetitive DNA motifs consisting of 2–5 nucleotides that are densely distributed throughout the human genome. Consequently, we aimed to establish a quick, inexpensive and robust technique that allows accurate quantification of genomic aberrations from only minute amounts of input DNA.

MATERIALS AND METHODS

Patient and control subjects

Diagnostic BM and peripheral blood (PB) aspirates were collected from 135 patients suffering from MDS or secondary acute myeloid leukaemia (AML) in the Department of Hematology and Oncology of the Medical Faculty Mannheim, University of Heidelberg, Germany, during 2009 and 2015 after written informed consent. The use of human

INTRODUCTION

The acquisition of cytogenetic abnormalities is a frequent event among various clonal malignant disorders. In particular, such lesions can be detected in 30%–50% of the bone marrow (BM) of patients with myelodysplastic syndromes (MDS).¹ Interstitial deletions of the long arm of chromosome



CrossMark

To cite: Jann J-C, Nowak D, Nolte F, et al. *J Med Genet* 2017;**54**:640–650.

material within this study was approved by the local Institutional Review Board. Patients with MDS were subdivided into A) MDS with del(5q) either isolated or with additional cytogenetic aberrations (n=86, mean age 68 years (range 44–91)) and B) MDS with confirmed absence of del(5q) (n=49, mean age 70 years (range 20–90)) by means of MC following the guidelines of the International System for Human Cytogenetic Nomenclature (ISCN).¹⁰ As non-del(5q) controls (n=125), we isolated PB specimen from healthy donors and cord blood. In addition, serial chronological BM samples (n=95) following treatment with lenalidomide (LEN) were analysed from n=40 del(5q) patients, who were enrolled within the LEMON-5 clinical trial.¹¹

Sample preparation

BM or PB aspirates were subjected to Ficoll density gradient centrifugation (GE Healthcare, Munich, Germany) for isolation of mononuclear cells. For germline correction, mesenchymal stromal cells (MSCs) for n=47 patients were expanded in vitro as described previously.¹² Fluorescence activated cell sorting (FACS) analysis of MSCs demonstrated high purities and absence of residual haematopoietic cells. DNA was isolated using the Allprep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

Cytogenetics

For cytogenetic confirmation of del(5q), MC analyses were carried out as reported previously.¹³ For quantitative interrogation, 5×10^4 cells were fixed in methanol:acetic acid (3:1) and interphase FISH was carried out with probes targeting *EGR1* (5q31) and *RPS14* (5q33) (MetaSystems, Altlußheim, Germany). Subsequently, 200 cells were analysed on a MetaSystems scanning system (Metafer4, MetaCyte). Cytogenetic analyses were blinded to the results of the paired STR test.

Polymerase chain reaction

For multiplex-PCR amplification of 12 separate STR loci within the region of del(5q), the fluorochromes FAM, HEX and TAMRA were used as 5' forward primer labels and for each fluorochrome four non-overlapping STR-flanking PCR products with varying average sizes ranging from ~100 to ~400bp were generated (see online supplementary table S1).

In a volume of 25 µl PCR reactions were carried out using the Type-it Microsatellite PCR Kit (Qiagen) with 5 pmol of each primer and 10ng DNA unless stated otherwise. The following PCR conditions were used: 95°C 5 min, 29 cycles of 95°C for 30s, 57°C for 3 min, 72°C for 30s and a final elongation of 60 min at 60°C for complete extension of 'A'-overhangs. PCR products were subsequently diluted 1:100 with laboratory grade H₂O. A further 1:10 dilution was performed in Hi-Di formamide (Life Technologies, Carlsbad, California, USA) containing 1.5% GeneScan 500XL ROX Standard (Life Technologies) followed by 5 min denaturation at 95°C; 10 µl of the final dilution were loaded into an ABI XL3130 system and subjected to capillary electrophoresis.

Calculation of del(5q) cell fractions

Blinded peak calling was carried out on *fsa*-files using Genemapper V.4.0 software (Life Technologies) using bin sets derived from multiple healthy donors. This was followed by manual verification to ensure that peak patterns resembled the germline pattern derived from concomitantly analysed patient-matched MSC samples. If no germline control for profiling was available, only peaks were taken into account that could

confidently be differentiated from PCR-stutter. Markers for which the non-deleted allele peak exhibited an area under the curve (AUC) below 2500 (insufficient PCR product) or above 25 000 fluorescence units (signal saturation) were discarded from subsequent analysis. All data were assessed retrospectively for collected specimen.

For correction of PCR-stutter that might overlap with the shorter PCR fragment (peak1), we screened all control samples for homozygous STR loci. For homozygous markers with *R* repeats, the relative proportion of stutter peaks at positions *R-1*, *R-2*, *R-3* and *R-4* relative to the index peak (peak2) were determined. By averaging this proportion, a marker-specific correction factor (CF_{*i*}) was derived specifically for each combination of stutter position (from *R-1* to *R-4*) and corresponding absolute fragment length (total number of repeats '*R*').

If in a heterozygous sample the shorter allele overlapped with such a stutter peak at position *R-1* to *R-4*, the lower PCR peak was corrected as follows:

$$peak1.corr = peak1 - (CF_i \times peak2)$$

Subsequently, we calculated the degree of skewing for the superior allele, p_{sup} , with AUC peak values from corresponding normal (n) and tumour (t) samples:

$$p_{sup} = 1 - \left(\frac{\frac{\frac{peak2\ t}{peak1.CORR\ t + peak2\ t}}{peak2\ n}}{peak1.CORR\ n + peak2\ n} + \frac{\frac{peak1.CORR\ t}{peak1.CORR\ t + peak2\ t}}{peak1.CORR\ n + peak2\ n} \right)$$

The p_{sup} was subsequently translated into the proportion of cells carrying del(5q):

$$delcell[\%] = 1 - \frac{0.5 - |(0.5 - p_{sup})|}{0.5 + |(0.5 - p_{sup})|}$$

These calculations were performed for each marker individually and subsequently the mean value derived from all informative markers was used for further analysis.

The frequency of deleted cells by quantification of skewing of heterozygous SNPs derived from high throughput sequencing was determined as previously reported.¹⁴

Analyses were performed in Microsoft Excel (V.14.0.7153) as well as custom scripts using R, V.3.1.3.¹⁵ An excel file for calculation of del(5q) frequencies from 'Genemapper' software derived peak raw data is provided in online supplementary table S2.

Statistical analysis

Statistical analyses were carried out using Pearson's correlation for standard curve, quantitative correlation for FISH or method comparisons. For two group comparisons, Student's t-test was used. Analyses were performed using Graphpad Prism 6 (GraphPad Software, La Jolla, California, USA) or R. Diagnostic accuracy was determined using the R package pROC V.1.8¹⁶ with default settings.

RESULTS

Simultaneous analysis of 12 STR markers allows accurate estimation of del(5q) burden

In order to integrate the concept of STR-based analysis with quantitative copy number evaluation, we designed fluorochrome-labelled

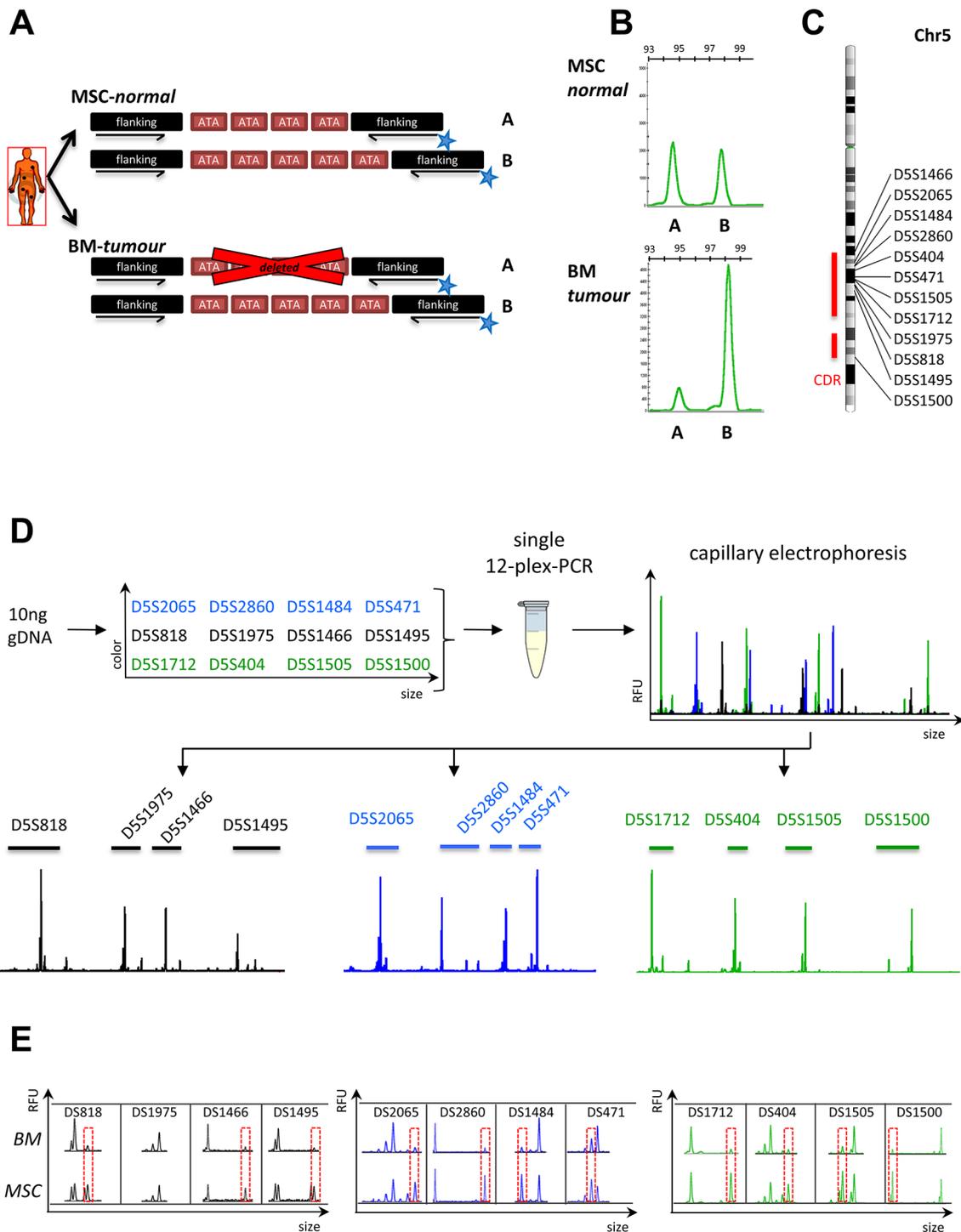


Figure 1 (A) Depiction of fluorochrome primer design for interrogation of allele-specific loss at short tandem repeat (STR) loci located in a deleted chromosomal region. (B) Exemplary peak profile of STR marker D5S1712 for mesenchymal stromal cells (MSCs) demonstrating equal amplification of both alleles and for the bone marrow (BM) sample from the same patient showing loss of allele 'A' due to genomic deletion of this particular locus. (C) Chromosomal distribution of selected STR markers among the long arm of chromosome 5 commonly deleted region (CDR) according to^{17 18} (D) experimental setup for multiplex-PCR amplification of 12 selected STR markers and subsequent separation of marker allele profiles for quantification of individual allelic ratios. (E) STR-marker profiles from an exemplary patient sample showed consistently reduced peak intensities of PCR fragments derived from the deleted allele for heterozygous loci (red boxes). Only D5S1975 appeared to be homozygous and therefore uninformative. RFU, relative fluorescence unit.

PCR primers flanking the repetitive regions of STR loci within the del(5q) region (see online supplementary table S1). Using these primers for PCR amplification of heterozygous STR loci should ideally result in equal amplification of two alleles in germline samples (eg, MSCs), but strongly reduced amplification from the

deleted allele in the tumour sample, for example, BM (figure 1A). Indeed, exemplary capillary electrophoresis reliably confirmed skewing of allele-specific PCR-amplicon intensities in the tumour but not corresponding germline sample (figure 1B), thus demonstrating the validity of this approach.

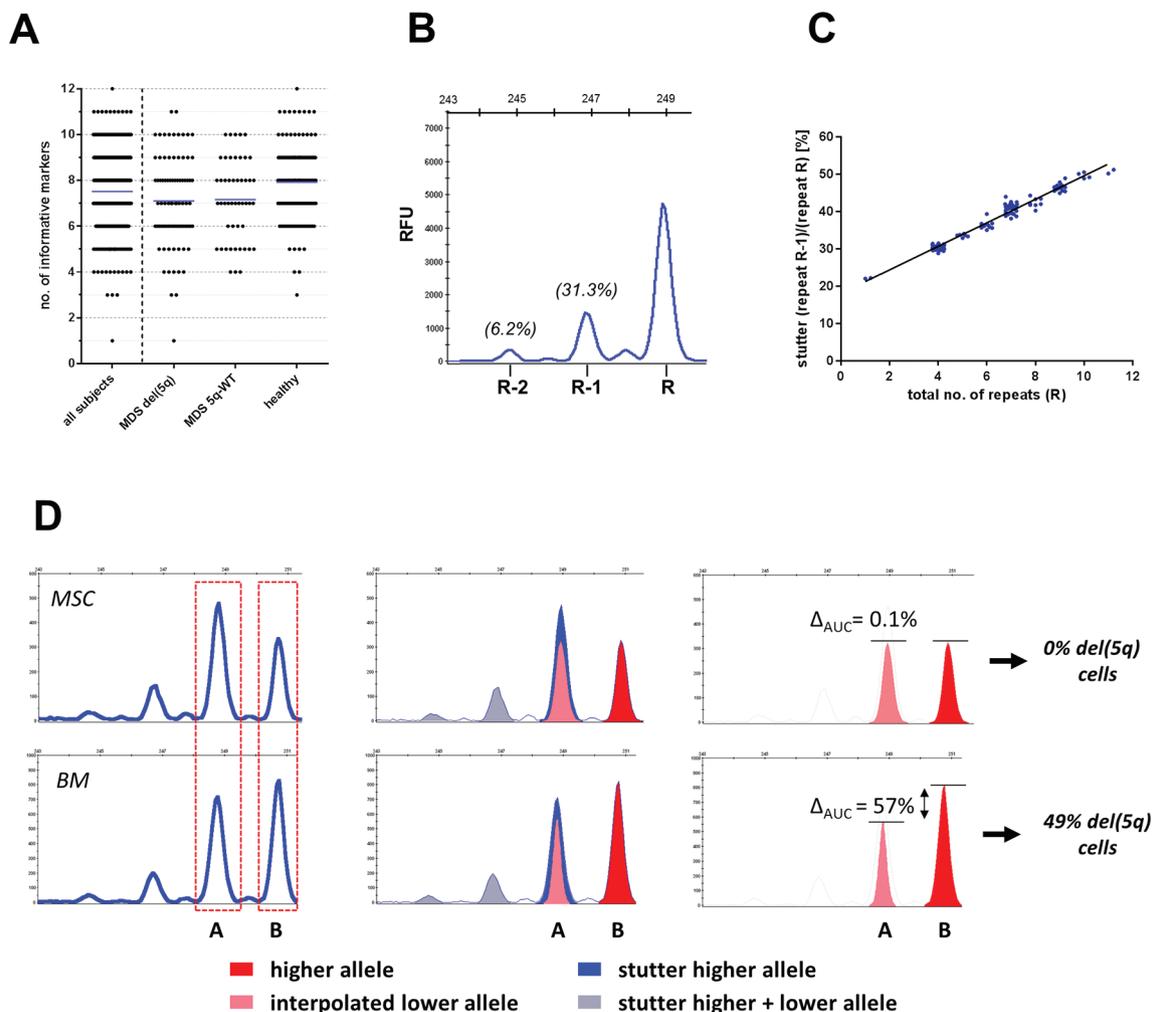


Figure 2 (A) Distribution of the number of informative (heterozygous) short tandem repeat (STR) markers for the entire study cohort and related subgroups. (B) Exemplary peak pattern for homozygous STR marker D5S417 showing its allelic peak at 248 bp and the corresponding PCR-stutter peaks at 246 and 244 bp. (C) Proportion of the 'R-1' stutter peak as a function of the total number of repeats exemplarily shown for D5S417. (D) Schematic depiction of correction for PCR-stutter contribution to a coinciding lower allele peak 'A' and subsequent translation into fractions of del(5q) cells from corrected allele ratios.

In order to generate robust estimates of del(5q) cell frequencies from DNA, we developed a multiplex-PCR assay that interrogates 12 independent STR markers, distributed along the commonly deleted region within del(5q),^{17,18} in a single reaction (figure 1C). By implementing three fluorochromes and non-overlapping PCR amplicon sizes in the PCR design (see online supplementary table S1), each individual marker could be reliably evaluated (figure 1D and E).

In total, $n=1142$ samples were analysed from $n=260$ individuals. By analysing 12 STR markers, we identified on average 7.5 markers (range 1–12) per subject that were heterozygous and therefore informative (figure 2A). Of note, none of the investigated samples exhibited microsatellite instability. For each marker separately, the degree of allelic skewing was translated into frequencies of cells carrying del(5q) using equations outlined in the 'Materials and methods' section. Subsequently, results from all informative markers were averaged to obtain the proportion of del(5q) cells for the respective sample.

A total of 259 out of 260 individuals (99.6%) harboured ≥ 3 informative markers for del(5q) estimation. In 93% (80/86) of del(5q) cases, we observed homogeneous skewing for all interrogated STR markers, suggesting that all markers were

located in the deleted region, which underlines the suitability of the selected STR marker based on a large cohort of del(5q) patients. In the remaining six cases, some markers were located outside the patient-specific deleted region. However, all six individuals harboured informative markers within the deleted region that could be used to quantify del(5q). Absolute marker intensities were homogeneous among all 12 loci indicating highly similar PCR amplification efficacies. Moreover, comparison of del(5q) cell frequencies obtained by any two marker combinations resulted in strong congruence (mean $R^2=0.92$, range: 0.77–0.97, see online supplementary figure S1).

PCR-stutter correction

For some STR markers, particularly those consisting of 2 bp repeats, allele peaks were accompanied by additional peaks that were exactly 1–4 repeat lengths shorter. This phenomenon of 'PCR-stutter' is well known¹⁹ (figure 2B). Because stutter peaks can potentially overlap with the shorter allele peak, we quantified their relative size in homozygous markers at position (R-1) to (R-4) for all observable alleles with R repeats in a cohort of $n=125$ healthy controls. On average, 10 individuals (range

1–61) were homozygous for any observed absolute marker fragment size. Strong correlations between relative stutter size and total number of repeats were found for several markers (figure 2C). Exemplarily, for a patient sample with heterozygous locus D5S471 the (R-1) stutter peak from the higher allele (251 bp) overlapped with the lower allele (249 bp) (figure 2D) resulting in a markedly increased lower allele peak in the MSC sample. Consequently, this pattern would theoretically indicate the presence of del(5q) cells in the germline sample. However, six healthy individuals showed an average relative stutter of 33.5% at the (R-1) position for the 251 bp allele of D5S471. After subtracting this proportion from the lower allele, the relative differences between lower and higher allele peak sizes changed to 0.1% for the germline control (<1% del(5q) cells) and 57% for the BM sample (49% del(5q) cells). This preprocessing strategy considerably improved the accuracy of del(5q) estimation from 1541 individual marker observations from $R^2=0.79$ to $R^2=0.87$ (figure 3A and B).

Determination of 'surrogate' germline control profiles

During our analysis of heterozygous markers in MSC germline controls, we observed slightly preferential amplification of the shorter allele resulting in a lower to higher peak ratio of >1. This phenomenon of 'allelic imbalance' was variable for each locus and dependent on the difference of fragment sizes between the alleles as well as the absolute length of each fragment. Fortunately, this bias could be effectively compensated by germline normalisation.

However, routine clinical acquisition of germline material is often impracticable and for archived patient DNA samples such controls are usually not available. In order to improve the accuracy of del(5q) quantification for samples without controls, we aimed to establish a comprehensive database of 'surrogate' germline profiles for the most frequently observed STR allele patterns. This was achieved by averaging the lower to higher allele peak ratio for every heterozygous marker allele combination observed in $n=125$ healthy donors. In total, a median number of 6 (range 1–69) observations for 190 unique combinations could be generated. The observed allele ratio imbalance for any particular allele combination was highly uniform among individuals (mean SD=8%).

Finally, for $n=295$ samples with available germline controls, we calculated fractions of del(5q) cells for every informative marker ($n=1397$) via normalisation 1) with our surrogate germline database and 2) using patient-matched MSC germline counterparts. Both approaches resulted in high concordance for del(5q) estimates from individual markers ($R^2=0.96$, figure 3C) and even more robust concordance for averaged *per sample* frequencies from all informative markers ($R^2=0.98$, figure 3D). Moreover, 'surrogate' germline correction considerably reduced the measurable offset for del(5q) frequency estimation in 5q-wild-type (5q-WT) cases from 16.9% to 6.2% ($p<0.0001$, figure 3E), reflecting the assay's low background noise level.

Validation of STR-based del(5q) quantification

In order to determine the accuracy of our PCR-based assay, we performed a serial dilution series with defined ratios of del(5q) and 5q-WT DNA obtained from the same individual. Correlation of expected and detected proportions of del(5q) cells revealed highly concordant results ($R^2>0.99$, figure 4A and B). Of note, paired analysis using our PCR-based assay and interphase-FISH from $n=34$ samples resulted in a strong correlation of measured del(5q) cell frequencies ($R^2=0.92$, figure 4C)

validating the suitability of STR-based quantification of del(5q) burdens. Moreover, this PCR-based method also showed high quantitative correlation with skewing of heterozygous SNPs for $n=37$ paired samples ($R^2=0.98$, figure 4D).

Assay reproducibility

To address the assay's reproducibility, we performed replicate analysis for $n=385$ samples and found an overall mean SD of 1.7% for del(5q) cell frequencies that were calculated from the average of all informative markers (see online supplementary table S2). Moreover, the reproducibility of individual markers was also highly comparable and ranged from 2.1% (D5S1495) to 4.6% (D5S1466).

PCR input

As this assay was designed to quantify del(5q) frequencies from samples with only limited available DNA, we aimed to determine its robustness with decreasing amounts of input DNA. Interphase-FISH analysis of a selected BM sample revealed 74% del(5q) positive cells (148/200 cells) (figure 4E). Consistently, STR-based quantification using 10 ng DNA isolated from the same specimen resulted in 75.3% (95% CI 73.7% to 76.9%) cells carrying del(5q). By serially decreasing the DNA input down to 0.3125 ng (equivalent to ~50 cells), the mean proportion of del(5q) positive cells for this DNA amount was 75.5% (95% CI 70.28% to 80.64%, figure 4E), which did not vary from 10 ng input ($p=0.94$) underlining the suitability of this assay even with ultra-low DNA input.

Monitoring of uniparental disomy and additional genomic lesions

Loss of heterozygosity at STR-loci can result from genomic deletions and from acquired uniparental disomy (UPD). As such, as an important advantage compared with FISH and MC, our established assay additionally provides quantitative information about UPDs. In a recent study, this approach allowed us to identify a patient with MDS with a clone carrying the del(5q) lesion that further evolved into telomeric UPD of 5q¹⁴ as validated by SNP-array analysis (figure 5A). For accurate distinction between del(5q) and 5q-UPD, we established a separate set of telomeric STR assays (figure 5B). This allowed us to reliably quantify the burden of cells carrying 5q-UPD (figure 5C), but also 7q-UPD (figure 5D).

Exemplarily, expanding STR-based analysis to additional chromosomal deletions allowed precise quantification of del(9q) and del(20q) aberrations (see online supplementary figure S2) indicating the assay's ability for interrogating a diverse set of genomic aberrations.

Suitability to monitor clonal burden under therapy

To evaluate the utility of STR-based del(5q) measurement as a diagnostic tool in a clinical setting, we first compared quantitative STR results of BM samples from patients with MDS who were cytogenetically confirmed del(5q) positive ($n=137$) or 5q-WT ($n=44$) as defined by ISCN criteria.¹⁰ Among del(5q) patients, we observed variable clonal burdens of del(5q) with clone sizes ranging from 2.7% to 94.4% del(5q) cells in BM (figure 6A). However, in the 5q-WT group we also found minor imbalances eventually translating into a mean proportion of 6.4% del(5q) positive cells which is considered as non-significant background. Thus, we aimed to determine our assay's diagnostic accuracy for classification of patients into del(5q) or 5q-WT cases and performed receiver operating characteristic analysis, which

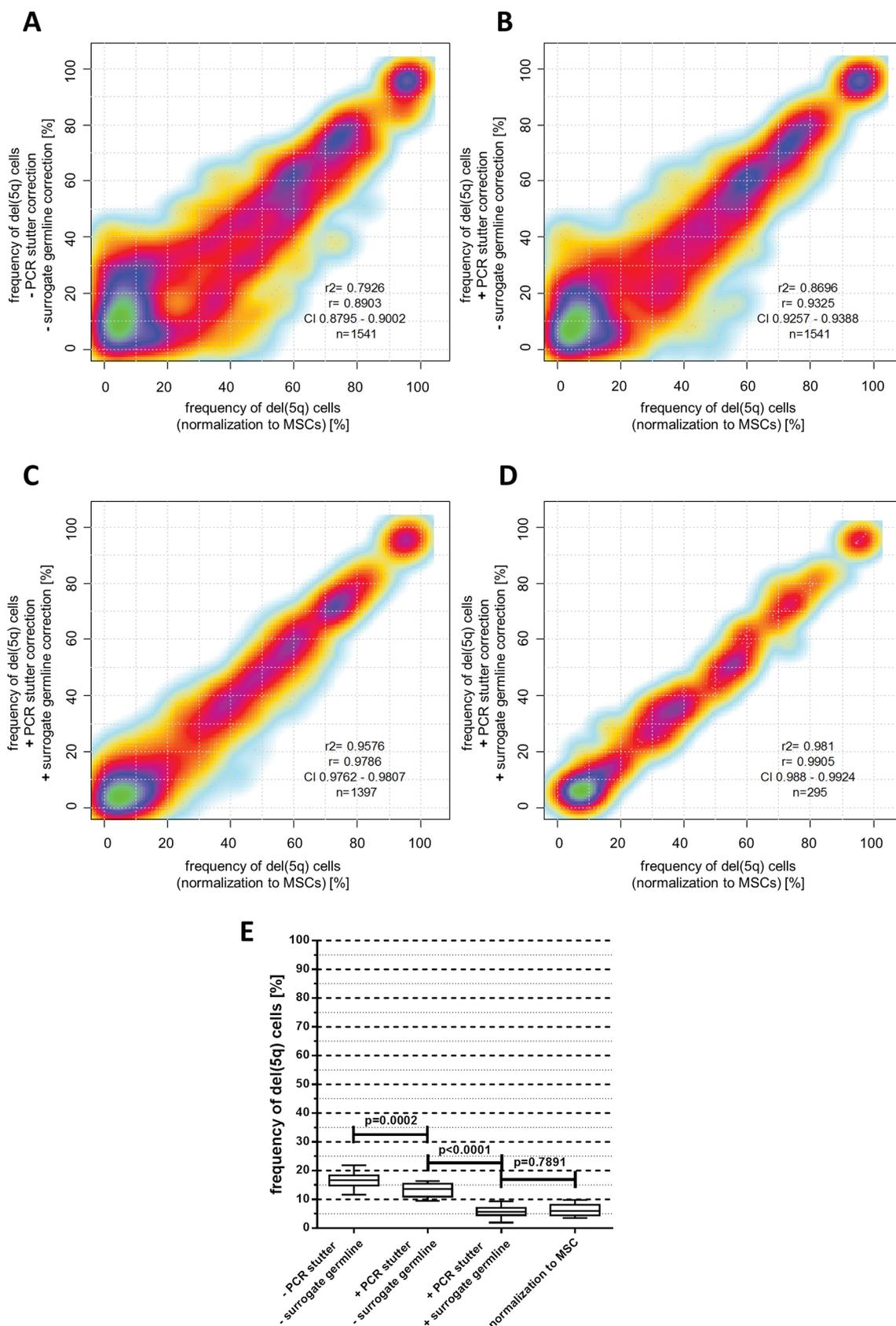


Figure 3 (A–C) Correlation of del(5q) frequencies calculated from individual short tandem repeat (STR) markers in haematopoietic cells normalised to matched mesenchymal stromal cell (MSC) germline samples (x-axis) and their corresponding del(5q) fractions (without MSC normalisation) calculated either with or without PCR-stutter and ‘surrogate’ germline correction (y-axis). (D) Comparison of average del(5q) fractions derived from all informative markers for $n = 295$ haematopoietic samples either normalised to a corresponding MSC sample (x-axis) or corrected for PCR-stutter and normalised to ‘surrogate’ germline controls (y-axis). (E) Reduction of offset del(5q) frequency measurements in 5q-wildtype (5q-WT) samples depending on PCR-stutter and surrogate germline correction (whiskers represent 10%–90% data range).

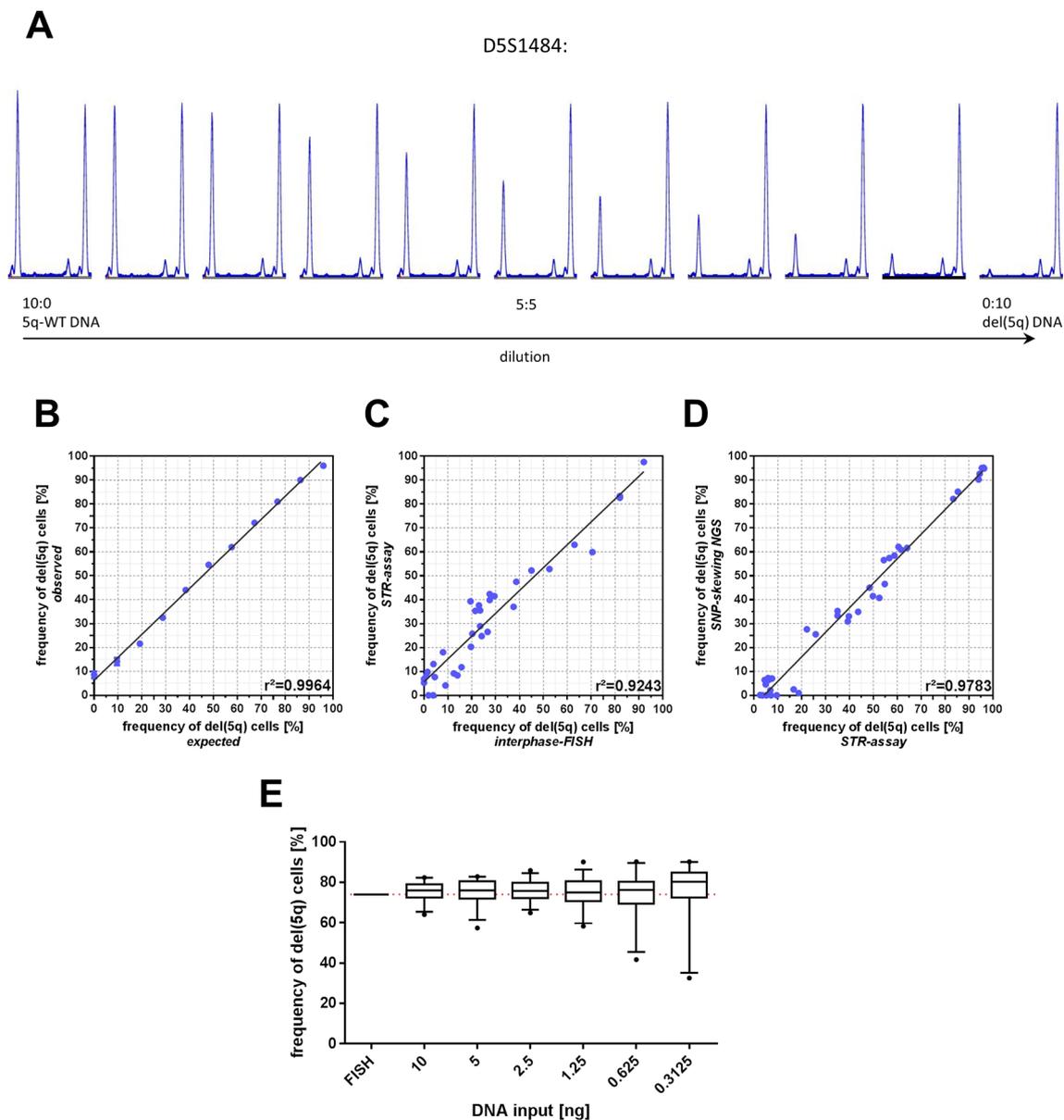


Figure 4 (A) Exemplary peak patterns for marker D5S1484 depicting the gradual increase of allele skewing in a serial dilution series. (B) Correlation of expected and observed del(5q) frequencies derived from defined mixed ratios of DNA carrying del(5q) (bone marrow) and 5q-wildtype (5q-WT) alleles (mesenchymal stromal cell (MSC)). Error bars depict SD from triplicate reactions. (C) Comparison of the frequency of del(5q) positive cells determined by concomitant interphase FISH and short tandem repeat (STR)-assay analysis. (D) Comparison of the frequency of del(5q) positive cells determined by next-generation sequencing-based SNP skewing or STR-assay analysis. (E) Impact of decreasing input amounts of DNA to the STR-PCR reaction. Each box plot reflects the results of individual markers obtained from three replicate reactions. Whiskers represent 5%–95% data range.

revealed an AUC of 0.97 (95% CI 0.94 to 0.99) (figure 6B). The optimal diagnostic cut-off, defined as the threshold providing the maximum distance to the diagonal identity line,²⁰ was found to be 11.4% del(5q) cells resulting in a sensitivity and specificity of 93.8% and 97.9% (positive predictive value: 98.9%, negative predictive value: 88.7%), respectively (see online Supplementary material).

Next, we sought to assess the utility of STR-based quantification for chronological patient monitoring. For this, serial samples from a cohort of $n=40$ patients, who were enrolled in the LEMON-5 trial from the German MDS study group were monitored for changes in clonal del(5q) burden upon treatment with LEN. To put these data into relation with established response criteria,²¹ centralised cytogenetic follow-up analysis was compared with STR-based del(5q) quantification.

While cytogenetic analyses failed (ie, metaphase failure) in 5/91 (5.5%) cases, our STR-PCR assay successfully generated del(5q) estimates in all interrogated samples. Among 12 patients, who achieved major cytogenetic response, determined by complete absence of aberrant metaphases, the mean proportion of cells carrying del(5q) was 6.7% (range 3%–10%) (figure 6C), which was below the previously defined detection threshold and therefore has to be classified as ‘negative’ in agreement with MC results. In addition, this proportion was significantly lower as compared with $n=15$ patients achieving only minor cytogenetic response, defined as $\geq 50\%$ reduction of aberrant metaphases (figure 6D and F, mean: 12.7% del(5q) cell frequency, range 5%–39%, $p=0.023$). Notably, none of the patients without cytogenetic response ($n=8$) showed a reduction below 38% del(5q) cells after treatment with LEN (figure 6E). In summary, these

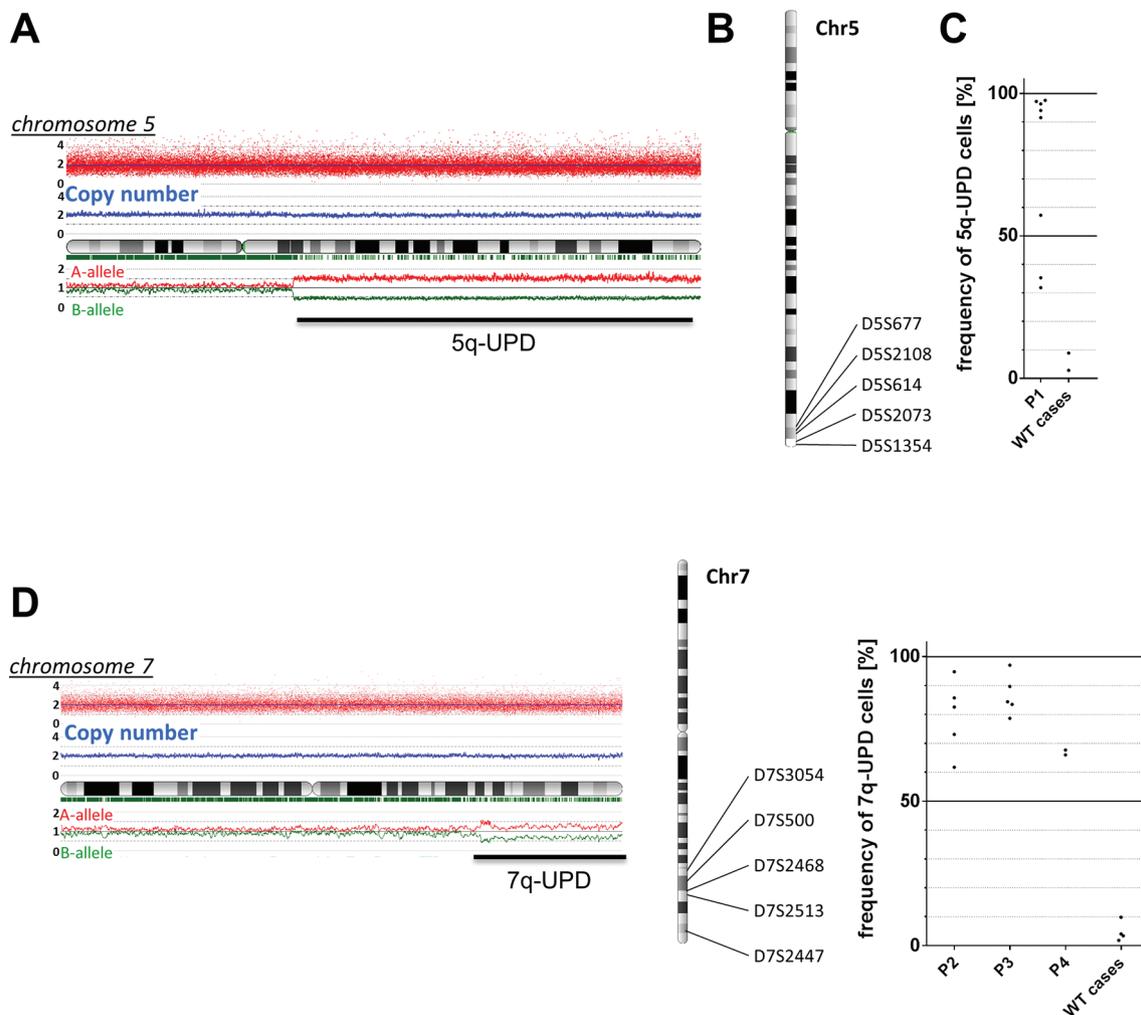


Figure 5 (A) Affymetrix SNP V.6.0 array analysed using CNAG V.3.0 software shows a telomeric uniparental disomy (UPD) at chromosome 5q. While the data indicate that a copy number of 2 is maintained throughout all chromosomal regions, telomeric loss of one allele can be detected on the long arm of the chromosome. (B) Chromosomal distribution of selected STR markers specific for the UPD region on chromosome 5q. (C) Quantification of the fraction of cells carrying 5qUPD in multiple samples from this patient (P1) and wildtype (WT) cases. (D) Left: SNP V.6.0 array shows a telomeric UPD at chromosome 7q. Middle: chromosomal distribution of selected STR markers for the UPD region on chromosome 7q surrounding the EZH2 locus. Right: quantification of the fraction of cells carrying the UPD for $n=3$ patients (P2–P4) and WT cases.

data suggest that STR-based quantification of $\text{del}(5q)$ represents a reliable approach for monitoring $\text{del}(5q)$ clonal burden during clinical follow-up.

Finally, we tested the applicability of the assay for $\text{del}(5q)$ monitoring in PB cells, which are generally easier to collect than BM aspirates. Indeed, frequencies of $\text{del}(5q)$ positive cells in $n=83$ matched PB and BM samples revealed a robust concordance ($R^2=0.79$, figure 6G) despite slightly lower burden in PB cells.

DISCUSSION

Using a multiplex-PCR assay for simultaneous measurement of 12 independent STR markers, we developed a highly adaptable tool for precise quantification of chromosomal lesions. By analysing a large collection of $n=1142$ samples, our DNA-based assay provided accurate assessment of cell frequencies carrying $\text{del}(5q)$, as an exemplary lesion, which was confirmed via replicate analysis and correlation with paired interphase-FISH results. Importantly, our assay does not require dividing or fixated cells and therefore is highly suitable for copy number quantification in samples for which only residual DNA is available. While microarray or qPCR-based methods could also be used for

targeted interrogation of chromosomal lesions from DNA,^{22–23} they are costly and provide only semi-quantitative results. Additionally, this assay resembles quantitative copy number quantification of current next-generation sequencing approaches that quantify allelic skewing of heterozygous SNPs, which are, however relatively expensive and require prior knowledge of the individual's SNP genotypes. With approximately US\$2 per reaction, multiplex-STR analysis is relatively inexpensive. With only 1 hour hands-on time for measurement of 96 DNA samples, the procedure generates results in <24 hours and is easily scalable for lower sample throughput.

In forensic analyses, STR markers represent a key element for genotyping of individuals due to their high rate of heterozygosity and therefore informativity.^{24–26} In line with this, all of our 260 investigated individuals harboured informative markers within the investigated region of $\text{del}(5q)$. Usually, analysis of STR markers suffers from PCR-related side effects such as PCR-stutter and small variations in allele-specific PCR efficiencies (allelic imbalances), which even occur in more stable trinucleotide to pentanucleotide repeats. Most of these artefacts can be corrected for by using patient-matched germline samples, for example, MSCs. However, because such controls are often unavailable for

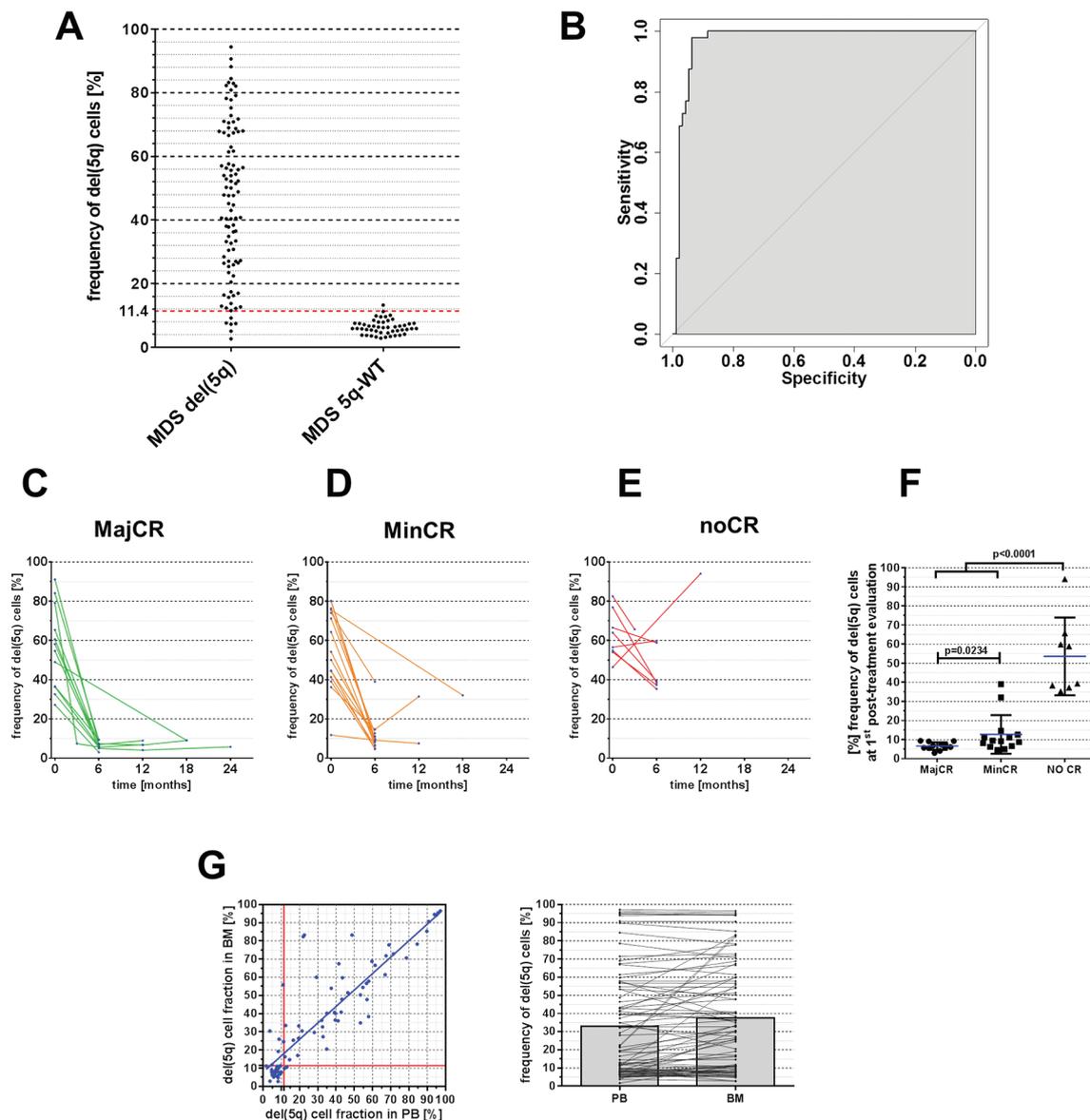


Figure 6 (A) Distribution of observed del(5q) fractions as determined via short tandem repeat (STR)-based quantification for patients with del(5q) and 5q-WT. Red dotted line at 11.4% indicates chosen diagnostic threshold from receiver operating characteristic (ROC) analysis. (B) ROC analysis illustrating the sensitivity and specificity of STR-based del(5q) stratification in patients with MDS with del(5q) and 5q-WT. (C–E) STR-based quantification of del(5q) burden in bone marrow (BM) following treatment with lenalidomide (LEN) subdivided to cytogenetic response groups. MajCR, major cytogenetic response; MinCR, minor cytogenetic response; noCR, no cytogenetic response. (F) Summary of del(5q) frequencies on patients’ first post-LEN BM examination according to observed clinical response. (G) Correlation of del(5q) cell fractions using STR-based assessment in matched peripheral blood (PB) and BM samples.

archival specimen or in a routine clinical setting, we developed an effective compensation matrix by integrating data from a large control cohort. With this, ‘surrogate’ germline profiles and PCR-stutter corrections could be established for almost every observed marker allele combination. Notably, del(5q) quantification using either ‘surrogate’ or patient-matched MSC germline correction demonstrated a strong correlation ($R^2 > 0.98$) across $n = 295$ individual samples. For convenience, we provide an excel sheet (see online Supplementary table S2) that can be easily employed for calculation of del(5q) frequencies from raw data. By implementing our comprehensive correction database in this file, all raw values are fully automatically corrected without need for manual intervention and subsequently translated into the fraction of del(5q) cells.

From specimens with low cell yield and quality, such as rare FACS-sorted fractions, BM smears or colony-forming units, it is

particularly challenging to obtain accurate and quantitative copy number data. Importantly, due to the small amplicon sizes our assay should also be applicable for investigation of fragmented DNA from formalin-fixed archived specimen. Moreover, our finding of high accuracy even with ultra-low DNA input equivalent to 50 cells suggests that STR-based analysis is highly applicable for copy number interrogation from such problematic samples. In this context, our assay has recently been successfully used to decipher the chronological relationship between chromosomal and point or small insertion/deletion mutations by investigating a large number of samples of low cellular abundance.¹⁴

Diagnosis of del(5q) has important implications for stratification of patients with MDS. Compared with MC, STR-based del(5q) quantification provided high diagnostic accuracy for del(5q) assessment. In addition, the method reliably monitored

del(5q) clonal burden during treatment with LEN suggesting its applicability as an alternative DNA-based technique for cytogenetic clone-size evaluation. Importantly, copy number neutral lesions such as acquired UPD can also be identified using this assay, representing a significant advantage over currently used standard diagnostic procedures.

Previous studies already indicated that copy number aberrations in BM cells can often also be detected in corresponding PB samples.^{27–31} In agreement with this, cross-comparison of STR-based del(5q) quantification in paired BM and PB samples showed reliable correlation, possibly allowing the use of such easily obtainable specimen for clinical clone size monitoring.

In summary, our newly developed DNA-based PCR assay provides an inexpensive tool to obtain quantitative data for a diverse set of chromosomal aberrations, which contain STRs and should be easily applicable to other clonal diseases.

Author affiliations

¹III Medizinische Klinik, Hämatologie und Onkologie, Universitätsmedizin Mannheim, Mannheim, Germany

²Medizinische Klinik und Poliklinik I, Universitätsklinikum Carl Gustav Carus, Dresden, Germany

³Klinik für Hämatologie, Onkologie und Palliativmedizin, Marienhospital, Düsseldorf, Germany

⁴III. Medizinischen Klinik des Klinikums rechts der Isar, Technische Universität München, München, Germany

⁵Medizinische Klinik für Hämatologie, Onkologie, Campus Benjamin Franklin, Charité Universitätsmedizin Berlin, Berlin, Germany

⁶Klinik für Hämatologie und Medizinische Onkologie, Georg-August-Universität Göttingen Universitätsmedizin, Göttingen, Germany

⁷NCT Trial Center, Nationales Centrum für Tumorerkrankungen (NCT), Heidelberg, Germany

⁸Medizinische Klinik II, Abteilung für Hämatologie/Onkologie, Klinikum der Johann Wolfgang Goethe-Universität Frankfurt, Frankfurt am Main, Germany

⁹Abteilung für Innere Medizin I, Hämatologie und Onkologie, Universitätsklinikum Freiburg, Freiburg, Germany

¹⁰Abteilung für Hämatologie, Hämostaseologie, Onkologie und Stammzelltransplantation, Medizinische Hochschule Hannover, Hannover, Germany

¹¹Abteilung für Hämatologie, Onkologie und klinische Immunologie, Heinrich-Heine-Universität Dusseldorf Medizinische Fakultät, Dusseldorf, Germany

¹²Münchener Leukämie Labor, München, Germany

Correction notice This paper has been amended since it was published Online First. Owing to a scripting error, some of the publisher names in the references were replaced with 'BMJ Publishing Group'. This only affected the full text version, not the PDF. We have since corrected these errors and the correct publishers have been inserted into the references.

Acknowledgements This work was supported by funds of the National Center for Tumor Diseases (NCT, Heidelberg, Germany), funds from the "Deutsche Krebshilfe" within a German wide MDS collaborative initiative ("Deutschland weites MDS Verbundprojekt, Teilprojekt I"), funds from the Gutermuth Foundation, funds from the ZOBEL initiative within the Innovation fund for Medicine Baden Württemberg and funds from the H.W. & J. Hector fund, Baden Württemberg. J.-C.J. is supported by a fellowship from the "Studienstiftung des deutschen Volkes". M.M. is supported by a fellowship from the "Deutsche José Carreras Leukämie Stiftung" (DJCLS F 11/05).

Contributors J-CJ, DN, W-KH and MM designed the study, analysed data and wrote the manuscript; J-CJ and MM conducted experimental design and most of the experiments including assessment of short tandem repeat patterns for lesion quantification; W-KH supervised the whole study and provided research infrastructure; AF, DH and CH provided metaphase cytogenetics and interphase-FISH analyses; SF, VN, JO, JP, IP and CX provided assistance for molecular analyses and clinical sample workup; FN, UP, AG, KG, AL, DH, RS, GB, ML, AG, UG and DN provided patient material and clinical data within the LEMON-5 trial.

Competing interests CH declares part ownership of the MLL Munich Leukemia Laboratory GmbH.

Ethics approval Institutional Ethics Review Board II of the Medical Faculty Mannheim, University of Heidelberg.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use,

provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- 1 Bejar R, Steensma DP. Recent developments in myelodysplastic syndromes. *Blood* 2014;124:2793–803.
- 2 Haase D, Germing U, Schanz J, Pfeilstöcker M, Nösslinger T, Hildebrandt B, Kundgen A, Lübbert M, Kunzmann R, Giagounidis AA, Aul C, Trümper L, Krieger O, Stauder R, Müller TH, Wimazal F, Valent P, Fonatsch C, Steidl C. New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* 2007;110:4385–95.
- 3 Haferlach C, Bacher U, Tiu R, et al. Myelodysplastic syndromes with del(5q): indications and strategies for cytogenetic testing. *Cancer Genet Cytogenet* 1872;2008:101–11.
- 4 Haferlach C, Rieder H, Lillington DM, Dastugue N, Hagemeijer A, Harbort J, Stilgenbauer S, Knuutila S, Johansson B, Fonatsch C. 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* 2007;46:494–9.
- 5 Halling KC, Kipp BR. Fluorescence in situ hybridization in diagnostic cytology. *Hum Pathol* 2007;38:1137–44.
- 6 Bigoni R, Cuneo A, Milani R, Cavazzini F, Bardi A, Roberti MG, Agostini P, della Porta M, Specchia G, Rigolin GM, Castoldi G. Multilineage involvement in the 5q- syndrome: a fluorescent in situ hybridization study on bone marrow smears. *Haematologica* 2001;86:375–81.
- 7 O'Keefe CL, Tiu R, Gondek LP, et al. High-Resolution genomic arrays facilitate detection of novel cryptic chromosomal lesions in myelodysplastic syndromes. *Exp Hematol* 2007;35:240–51.
- 8 Gondek LP, Tiu R, O'Keefe CL, Sekeres MA, Theil KS, Maciejewski JP. Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood* 2008;111:1534–42.
- 9 Nowak D, Nolte F, Mossner M, Nowak V, Baldus CD, Hopfer O, Noll S, Thiel E, Wagner F, Hofmann WK. Genome-wide DNA-mapping of CD34+ cells from patients with myelodysplastic syndrome using 500K SNP arrays identifies significant regions of deletion and uniparental disomy. *Exp Hematol* 2009;37:215–24.
- 10 International Standing Committee on Human Cytogenetic Nomenclature. In: Shaffer LG, McGowan-Jordan J, Schmid M, eds. *ISCN 2013: an international system for human cytogenetic nomenclature* (2013). Basel: Karger, 2013.
- 11 Schuler E, Giagounidis A, Haase D, Shirmeshan K, Büsche G, Platzbecker U, Nolte F, Götze K, Schlenk RF, Ganser A, Letsch A, Bräulke F, Lübbert M, Bug G, Schafhausen P, Bacher U, Gattermann N, Wulfert M, Haas R, Germing U. Results of a multicenter prospective phase II trial investigating the safety and efficacy of lenalidomide in patients with myelodysplastic syndromes with isolated del(5q) (LE-MON 5). *Leukemia* 2016;30:1580–2.
- 12 Medyouf H, Mossner M, Jann JC, Nolte F, Raffel S, Herrmann C, Lier A, Eisen C, Nowak V, Zens B, Müdder K, Klein C, Obländer J, Fey S, Vogler J, Fabarius A, Riedel E, Roehl H, Kohlmann A, Staller M, Haferlach C, Müller N, John T, Platzbecker U, Metzgeroth G, Hofmann WK, Trumpp A, Nowak D. Myelodysplastic cells in patients reprogram mesenchymal stromal cells to establish a transplantable stem cell niche disease unit. *Cell Stem Cell* 2014;14:824–37.
- 13 Fabarius A, Kalmanti L, Dietz CT, Lauseker M, Rinaldetti S, Haferlach C, Göhring G, Schlegelberger B, Jotterand M, Hanfstein B, Seifarth W, Hänel M, Köhne CH, Lindemann HW, Berdel WE, Staib P, Müller MC, Proetel U, Balleisen L, Goebeler ME, Dengler J, Falge C, Kanz L, Burchert A, Kneba M, Stegelmann F, Pfreundschuh M, Waller CF, Spiekermann K, Brümmendorf TH, Edinger M, Hofmann WK, Pfirrmann M, Hasford J, Krause S, Hochhaus A, Sauße S, Hehlmann R; SAKK and the German CML Study Group. Impact of unbalanced minor route versus major route karyotypes at diagnosis on prognosis of CML. *Ann Hematol* 20152015;94:94–24.
- 14 Mossner M, Jann JC, Wittig J, Nolte F, Fey S, Nowak V, Obländer J, Pressler J, Palme I, Xanthopoulos C, Boch T, Metzgeroth G, Röhl H, Witt SH, Dukal H, Klein C, Schmitt S, Gelß P, Platzbecker U, Balaian E, Fabarius A, Blum H, Schulze TJ, Meggendorfer M, Haferlach C, Trumpp A, Hofmann WK, Medyouf H, Nowak D. Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. *Blood* 2016;128:1246–59.
- 15 Team RC. *R: a language and environment for statistical computing*. Vienna, Austria: the R Foundation for Statistical Computing, 2015. ISBN: 3-900051-07-0. <http://www.R-project.org/>.
- 16 Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
- 17 Douet-Guilbert N, De Braekeleer E, Basinko A, Herry A, Gueganic N, Bovo C, Trillet K, Dos Santos A, Le Bris MJ, Morel F, Eveillard JR, Berthou C, De Braekeleer M. Molecular characterization of deletions of the long arm of chromosome 5 (del(5q)) in 94 MDS/AML patients. *Leukemia* 2012;26:1695–7.

- 18 Boultonwood J, Fidler C, Strickson AJ, Watkins F, Gama S, Kearney L, Tosi S, Kasprzyk A, Cheng JF, Jaju RJ, Wainscoat JS. Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. *Blood* 2002;99:4638–41.
- 19 Brookes C, Bright JA, Harbison S, Buckleton J. Characterising stutter in forensic STR multiplexes. *Forensic Sci Int Genet* 2012;6:58–63.
- 20 Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
- 21 Cheson BD, Bennett JM, Kantarjian H, Pinto A, Schiffer CA, Nimer SD, Lowenberg B, Beran M, de Witte TM, Stone RM, Mittelman M, Sanz GF, Wijermans PW, Gore S, Greenberg PL; World Health Organization international working group. Report of an international working group to standardize response criteria for myelodysplastic syndromes. *Blood* 2000;96:3671–4.
- 22 Weksberg R, Hughes S, Moldovan L, Bassett AS, Chow EW, Squire JA. A method for accurate detection of genomic microdeletions using real-time quantitative PCR. *BMC Genomics* 2005;6:180.
- 23 Gouas L, Goumy C, Véronèse L, Tchirkov A, Vago P. Gene dosage methods as diagnostic tools for the identification of chromosome abnormalities. *Pathol Biol* 2008;56:345–53.
- 24 Gill P, Haned H, Bleka O, Hansson O, Dørum G, Egeland T. Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches—Twenty years of research and development. *Forensic Sci Int Genet* 2015;18:100–17.
- 25 Lounsbury JA, Bienvenue JM, Landers JP. Sample-to-Result STR genotyping systems: potential and status. *Forensic Sci Rev* 2012;24:123–42.
- 26 Kristt D, Stein J, Yaniv I, Klein T. Assessing quantitative chimerism longitudinally: technical considerations, clinical applications and routine feasibility. *Bone Marrow Transplant* 2007;39:255–68.
- 27 Mohamedali AM, Gäken J, Ahmed M, Malik F, Smith AE, Best S, Mian S, Gaymes T, Ireland R, Kulasekararaj AG, Mufti GJ. High concordance of genomic and cytogenetic aberrations between peripheral blood and bone marrow in myelodysplastic syndrome (MDS). *Leukemia* 2015;29:1928–38. doi:10.1186/1471-2164-6-180
- 28 Coleman JF, Theil KS, Tubbs RR, Cook JR. Diagnostic yield of bone marrow and peripheral blood FISH panel testing in clinically suspected myelodysplastic syndromes and/or acute myeloid leukemia: a prospective analysis of 433 cases. *Am J Clin Pathol* 2011;135:915–20.
- 29 McDevitt MA, Condon M, Stamberg J, Karp JE, McDiarmid M. Fluorescent in situ hybridization (FISH) in bone marrow and peripheral blood of leukemia patients: implications for occupational surveillance. *Mutat Res* 2007;629:24–31.
- 30 Brulke F, Schanz J, Jung K, Shirneshan K, Schulte K, Schuetze C, Steffens R, Trümper L, Haase D. FISH analysis of circulating CD34+ cells as a new tool for genetic monitoring in MDS: verification of the method and application to 27 MDS patients. *Leuk Res* 2010;34:1296–301.
- 31 Cherry AM, Slovak ML, Campbell LJ, Chun K, Eclache V, Haase D, Haferlach C, Hildebrandt B, Iqbal AM, Jhanwar SC, Ohyashiki K, Sole F, Vandenberghe P, VanDyke DL, Zhang Y, Dewald GW. Will a peripheral blood (PB) sample yield the same diagnostic and prognostic cytogenetic data as the concomitant bone marrow (BM) in myelodysplasia? *Leuk Res* 2012;36:832–40.