Interlaboratory Analytical Validation of a Next-Generation Sequencing Strategy for Clonotypic Assessment and Minimal Residual Disease Monitoring in Multiple Myeloma

Alejandro Medina, PhD; Cristina Jiménez, PhD; Noemí Puig, MD, PhD; María Eugenia Sarasquete, PhD; Juan Flores-Montero, PhD; María García-Álvarez, PhD; Isabel Prieto-Conde, PhD; Carmen Chillón, PhD; Miguel Alcoceba, PhD; Verónica González-Calle, MD, PhD; Norma C. Gutiérrez, MD, PhD; Austin Jacobsen, PhD; Edgar Vigil, PhD; Kasey Hutt, PhD; Ying Huang, PhD; Alberto Orfao, PhD; Marcos González, MD, PhD; Jeffrey Miller, PhD; Ramón García-Sanz, MD, PhD

● *Context.*—Minimal residual disease (MRD) is a major prognostic factor in multiple myeloma, although validated technologies are limited.

Objective.—To standardize the performance of the LymphoTrack next-generation sequencing (NGS) assays (Invivoscribe), targeting clonal immunoglobulin rearrangements, in order to reproduce the detection of tumor clonotypes and MRD quantitation in myeloma.

Design.—The quantification ability of the assay was evaluated through serial dilution experiments. Paired samples from 101 patients were tested by LymphoTrack, using Sanger sequencing and EuroFlow's next-generation flow (NGF) assay as validated references for diagnostic and follow-up evaluation, respectively. MRD studies using LymphoTrack were performed in parallel at 2 laboratories to evaluate reproducibility.

Results.—Sensitivity was set as 1.3 tumor cells per total number of input cells. Clonality was confirmed in 99% and 100% of cases with Sanger and NGS, respectively, showing great concordance (97.9%), although several samples had minor discordances in the nucleotide sequence of rearrangements. Parallel NGS was performed in 82 follow-up cases, achieving a median sensitivity of 0.001%, while for NGF, median sensitivity was 0.0002%. Reproducibility of LymphoTrack-based MRD studies (85.4%) and correlation with NGF ($R^2 > 0.800$) were high. Bland-Altman tests showed highly significant levels of agreement between flow and sequencing.

Conclusions.—Taken together, we have shown that LymphoTrack is a suitable strategy for clonality detection and MRD evaluation, with results comparable to gold standard procedures.

(*Arch Pathol Lab Med.* 2022;146:862–871; doi: 10.5858/arpa.2021-0088-OA)

Multiple myeloma (MM) is a plasma-cell dyscrasia characterized by the accumulation of plasma cells in the bone marrow that produces an excess of clonal immunoglobulins (M-protein or monoclonal component).[1] New treatment approaches have increased the number of patients achieving complete response (CR),[2–5] progressively improving progression-free and overall survival rates in the last 10 years.[6–11] Nonetheless, the presence of low levels of drug-resistant cells (known as minimal residual disease, MRD)[12–14] that remain undetected by conventional serologic and morphologic methods explains frequent relapses with this disease, which is still considered an incurable illness.
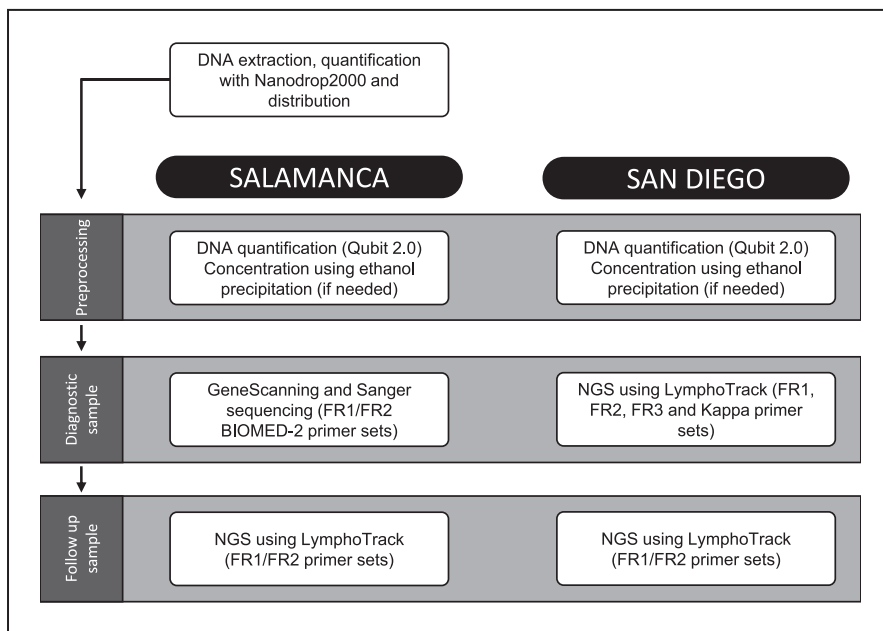
Figure 1. *Proposed workflow for Salamanca and San Diego laboratories. Abbreviations: FR, framework region; NGS, next-generation sequencing.*

Minimal residual disease is currently considered one of the most informative prognostic parameters, since those patients with undetectable disease have shown prolonged survival rates as compared with MRD-positive patients,[15–17] and this difference is still significant even when patients achieving only stringent complete response (sCR) are taken into account.[18] The International Myeloma Working Group (IMWG) defined MRD positivity as the persistence of clonal malignant plasma cells assessed with a sensitivity of at least 10^{-5} (1 malignant cell per hundred thousand normal cells)[19]; therefore, MRD should be monitored with only highly sensitive methods.

To date, 3 different approaches have been tested for MRD monitoring in hematologic malignancies: immunophenotypic (multiparametric flow cytometry [MFC]),[20] molecular (quantitative polymerase chain reaction [PCR], next-generation sequencing [NGS], digital PCR),[21–23] and imaging tools (positron emission tomography–computed tomography; magnetic resonance imaging).[24,25] However, in MM standardization has been achieved only for MFC[26] and NGS.[27,28] As a result, the IMWG recommended the use of highly sensitive, standardized flow and sequencing approaches,[19] including EuroFlow's next-generation flow (NGF)[29] and Adaptive Biotechnologies' ClonoSEQ solutions (Adaptive Biotechnologies, Seattle, Washington).

NGF is a 2-tube, 8-color flow assay that allows the simultaneous analysis of 10 million cells, providing a sensitivity of around $2 \cdot 10^{-6}$. Thus, it represents an improvement over previous flow protocols (sensitivity threshold: 10^{-4} to 10^{-5}), but in contrast it is highly dependent on the precise identification of the aberrant immunophenotype, which demands a high level of expertise[29]; in addition, samples have to be rapidly processed in order to maintain a high cell viability.[30]

Next-generation sequencing relies on the identification of clonal heavy- (*IGH*) and light-chain (*IGK*) immunoglobulin rearrangements. Currently, only 1 strategy (Adaptive Biotechnologies' ClonoSEQ) has been approved by the IMWG[19] and cleared by the US Food and Drug Administration (FDA),[31] although it is not commercially available in

Europe; turnaround time is long (~10 days) and the cost is high. Our group already explored other NGS alternatives such as LymphoTrack (Invivoscribe, Inc, San Diego, California) and found that the evaluation of MRD using this commercial solution is similar to NGF, with both providing similar prognostic information.[32] However, the role of the LymphoTrack strategy to detect clonal rearrangements at baseline was not explored and, more importantly, a technical validation to assess its analytical performance was lacking. Therefore, we have evaluated LymphoTrack, targeting clonal immunoglobulin *V(D)J* rearrangement sequences, to determine its analytical properties in serial dilution experiments. Additionally, we performed baseline clonotypic characterization and MRD monitoring in a cohort of 101 patients with myeloma who were tested in parallel at 2 institutions, comparing the results with standardized methods (BIOMED-2–based Sanger sequencing and NGF, respectively).

MATERIALS AND METHODS

Patient Samples and Study Design

Bone marrow samples from 101 patients with MM were collected at diagnosis and early follow-up time points (after the end of induction or 3 months after transplant), irrespective of their response, as part of the GEM2012MENOS65 clinical trial (NCT01916252).[11] Sixty-nine had been included in a previous publication from our group, exploring the clinical significance of MRD evaluation by next-generation techniques.[32] Patient selection was based on the following criteria: (1) immunophenotypic studies performed with conventional 8-color MFC (diagnosis) or Euro-Flow's NGF (follow-up), and (2) DNA availability for paired diagnostic and MRD studies. Responses to treatment were evaluated according to the IMWG response criteria.[19]

Samples were distributed and analyzed according to Figure 1. Thus, the Salamanca laboratory carried out the clonotypic characterization of diagnostic samples targeting *IGH* (FR1 and FR2 regions) using Sanger sequencing, while the San Diego laboratory used LymphoTrack (Invivoscribe Technologies, San Diego, California) targeting *IGH* (FR1, FR2, and FR3) and *IGK* rearrangements. For MRD studies both centers used NGS (FR1, or FR2 if necessary), allowing the measurement of interlaboratory

precision. In addition, baseline and follow-up MFC studies had been previously carried out in Salamanca. However, bone marrow samples for MFC and NGS were extracted in different pulls.

The Ethical Committee of the University Hospital of Salamanca (Salamanca, Spain) gave approval for this study, in accordance with the Spanish law and the Declaration of Helsinki principles. Written informed consent was obtained from every patient before their inclusion.

DNA Extraction and Quantification

Genomic DNA (gDNA) was isolated from bone marrow aspirates with the automated Maxwell DNA Purification Kit (Promega, Madison, Wisconsin). DNA quality and quantity were first assessed in the Salamanca laboratory using NanoDrop2000 (ThermoFisher, Waltham, Massachusetts) and then distributed. Samples were once more quantified by each laboratory, in this case using Qubit 2.0 and the dsDNA BR assay (ThermoFisher).

Follow-up samples with insufficient DNA concentration for MRD purposes (those with less than 100 ng/μL) were ethanol precipitated.[33] Sodium acetate 1/10 as well as 2-2, 5x volume ethanol (100% and stored at −20°C) of the total sample volume were added to the samples, then incubated overnight at −20°C. Afterwards, samples were centrifuged at 17 900g for 10 minutes at 4°C. The nucleic acid pellet was later washed with 500 μL ethanol (70%) and centrifuged again at 17 900g for 5 minutes at 4°C. Finally, the pellet was dried and rehydrated in ≈12 μL of water and quantified by using Qubit dsDNA BR assay.

Technical Validation

Limits of detection (LOD) and quantification (LOQ) of the LymphoTrack assay were assessed with the following experiment: diagnostic bone marrow biopsies from 20 patients with myeloma were sorted to select CD138[+] cells. Genomic DNA was obtained from all, and the LymphoTrack *IGH* panel (Invivoscribe Technologies) was used to determine the tumor-related clonotypic rearrangement. Afterwards, gDNA from all patients was pooled together at the same ratio (100 cells/patient/μL) based on the quantification made with Qubit 2.0.

Then, 5 serial dilutions at prespecified cell proportions (100, 30.3, 9.2, 2.8, and 0.85 cells/patient/μL) were made by using Tris-ethylenediaminetetraacetic acid buffer. Each was further diluted into buffy coat obtained from 7 healthy donors and tested with the LymphoTrack panel at 3 different total DNA inputs in 50 μL (DNA input: 500 ng) or 100 μL (DNA input: 1 μg or 2 μg) per reaction, so that 15 artificial MRD samples, each comprising 20 clonal rearrangements, were tested in the range of $1.3 \cdot 10^{-3}$ to $2.7 \cdot 10^{-6}$ (Supplemental Table 1, see Supplemental Digital Content 1, containing 7 tables and 1 figure at https://meridian.allenpress.com/aplm in the July 2022 table of contents). A constant proportion of spikein, consisting of a known quantity of DNA (equivalent to 100 cells) from a Waldenström macroglobulinemia cell line (MWCL1, kindly provided by Stephen Ansel, MD, PhD, Mayo Clinic, Rochester, Minnesota),[34] was added to each reaction. Sequencing was performed in the MiSeq platform (Illumina, San Diego, California) using v3 reagent kits and 2×251 cycles of sequencing.

Gene Amplification and Sequencing

For the amplification of complete *VDJH* rearrangements at diagnosis, the Salamanca laboratory used the BIOMED-2 (now Euroclonality) FR1 or FR2 primers in multiplexed PCR reactions.[35] All reactions were carried out in a 25-μL mixture containing 20 to 100 ng of baseline DNA and 10 μmol of forward and reverse primers. Monoclonal assessment of amplified products was carried out by GeneScanning using 1 μL of PCR reaction. PCR products were sequenced in an automated ABI3500 XL DNA sequencer using Big-Dye terminators v3.1 (Applied Biosystems, Foster City, California).

The LymphoTrack *IGH* panel was used for the analysis of diagnostic (San Diego laboratory) and MRD samples (both laboratories), aiming to detect tumor clonotypic rearrangements. Briefly, this commercial strategy uses primers targeting the immunoglobulin framework regions to amplify *V(D)J* rearrangements. In 1-step PCR (Figure 2, A through E) amplicons are generated and 1-side indexed, allowing the simultaneous sequencing of up to 24 samples in a single run.

All baseline samples were sequenced in San Diego, using 50 ng of DNA. Whenever it was possible, at least 650 ng of follow-up DNA was used in 1 or 2 PCR replicates to reach a sensitivity level of 10^{-5} (assuming 6.5 pg of DNA per cell, cell input for NGS would be at least 100 000). In addition, 1 μL of MWCL1, corresponding to 100 to 1000 cells, was added in each follow-up PCR reaction as a spike-in, to allow the absolute quantification of tumor plasma cells.

After a purification step using Agentcourt AMPure XP microbeads (Beckman Coulter Inc, Brea, California) and 70% ethanol, and once the purity and quantity of the amplicons were assessed with the TapeStation 4200 (Agilent, Santa Clara, California) and the KAPA Library Quantification Kit (KAPA Biosystems, Boston, Massachusetts) or Qubit 2.0, libraries of between 12 and 20 pM were prepared. These libraries were later sequenced in a MiSeq instrument using v3 reagent kits and 2×251 cycles of sequencing, aiming for 1 million reads per sample. To prevent cross-contamination during the sequencing step, samples from the same patient were always sequenced in different runs, using different indexes when possible. In addition, libraries were sequenced within 2 runs of each other, performing a sodium-hypochlorite post-run wash on the instrument to reduce carry-over contamination.

Sequence Analysis and MRD Evaluation

Sanger sequences were analyzed by using the IMGT/V-Quest database[36] taking into account the option for insertions/deletions. *IGHV*, *IGHD*, and *IGHJ* genes, as well as complementarity-determining region 3 (CDR3), were identified. Mutational status was assessed by using the closest germline gene as a reference.

FastQ files generated during NGS were processed with the LymphoTrack Software-MiSeq (Invivoscribe Technologies) to retrieve sequences from virtually every clonal B cell in the samples. Tumor clones at diagnosis were identified following 3 criteria: (1) 20 000 or more total reads; (2) at least 1 but not more than 2 merged top reads with 2.5% or greater of total reads; and (3) top first or second merged reads at least two times more abundant than the third most frequent read to be considered clonotypic. An NGS report, listing the top 5 rearrangements and sequencing parameters for all patients and primer sets, is included as Supplemental Digital Content 2.

Minimal residual disease was calculated with the LymphoTrack MRD Data Analysis tool 1.2.0 (Invivoscribe Technologies) considering the number of spike-in cells, the number of total cells that were used for each reaction, as well as tumor and spike-in read counts (Figure 2, E). Sequencing results were considered invalid when fewer than 20 000 total reads were retrieved. A sample was considered positive for MRD assessment when at least 2 identical tumor clonotypic reads were detected (allowing for up to 2-bp mismatches).

Phenotypic Analysis

Samples were analyzed within 48 hours after the extraction, using a conventional 8-color multiparametric flow cytometry panel (CD38/CD138/CD45/CD19/CD56/CD117/CD27/CD81) or the 2-tube, 8-color EuroFlow NGF panel (tube 1: CD38/CD138/CD27/CD45/CD19/CD56/CD81/CD117; tube 2: CD38/CD138/CD45/CD19/CD56/CyIgK/CyIgλ) for diagnosis and follow-up time points, respectively, following EuroFlow's guidelines for each assay.[26,29] Commercial antibody panels for diagnostic and MRD evaluation were used (Cytognos S.L., Salamanca, Spain). At least 500 000 (diagnosis) or 10 000 000 (follow-up) bone marrow cells were acquired in a BD FACSCanto II (BD Biosciences, Franklin Lakes, New Jersey). Data analysis was performed by using the calculation function of the INFINICYT v2.0 software (Cytognos S.L.).
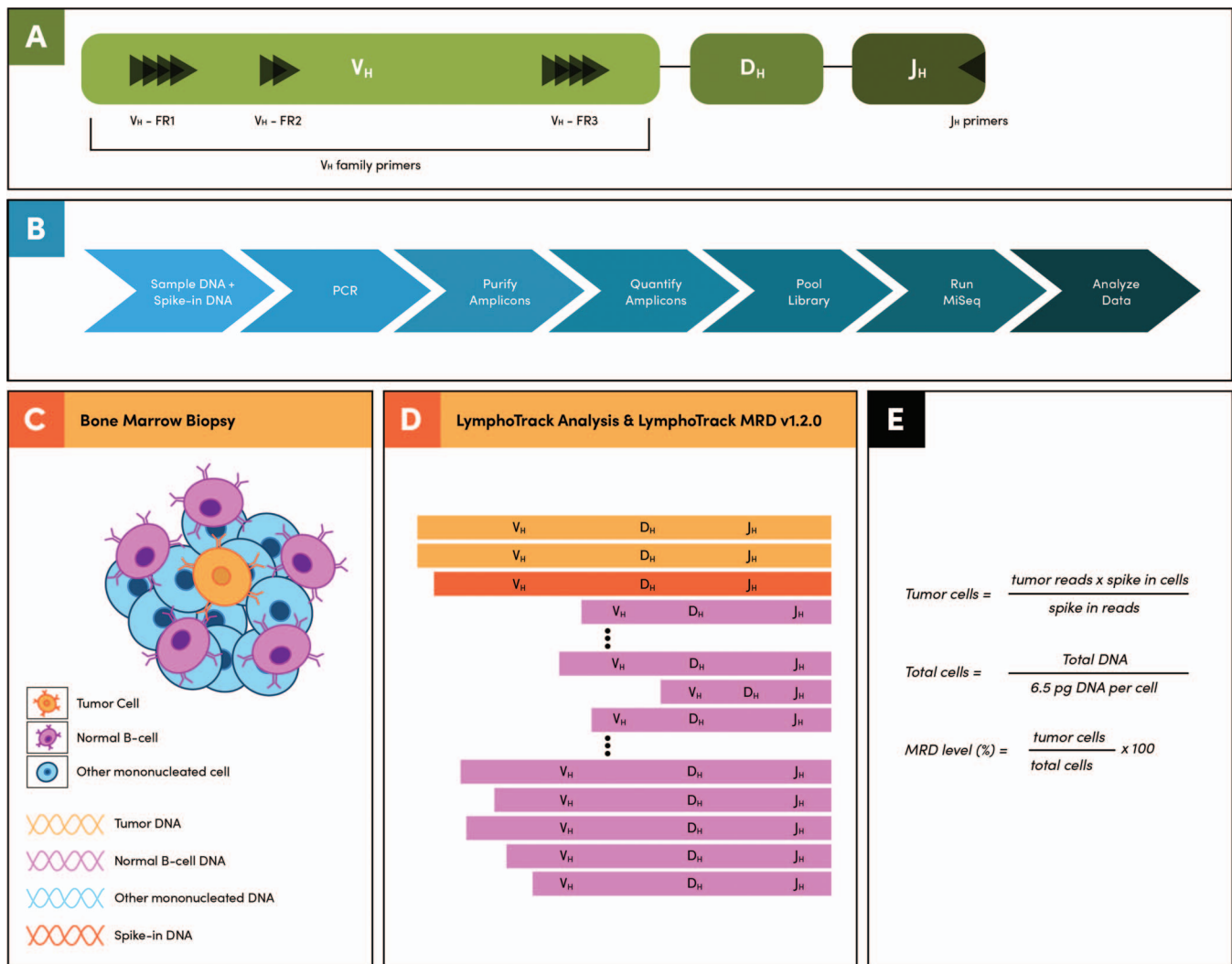
Figure 2. *Overview of the LymphoTrack method. A, Schematic distribution of consensus primer locations on the* IGH *locus. Forward primers target the 3 framework regions of* IGHV *genes. Reverse consensus primer targets the* IGHJ *genes. B, Workflow diagram of assay. C, Mixture of cells from biopsy results for multiple sources of DNA (color-coded based on source). gDNA from tumor and normal B cells, as well as other mononucleated cells, are mixed with the spike-in gDNA in the PCR reaction for normalization. D, Color-matched sources of amplicons are analyzed by software. E, Calculations used for tumor cells, total cells, and MRD level. Abbreviations: FR, framework region; gDNA, genomic DNA; MRD, minimal residual disease; PCR, polymerase chain reaction.*

Statistical Analysis

Patients' characteristics were tested with SPSS 20.0 software (IBM, Armonk, New York), using Fisher exact test for discrete variables and the Mann-Whitney test for continuous variables. Bland-Altman,[37] Cohen κ, and linear regression tests were used to assess potential agreement, concordance, and bias between methods.

LOD and LOQ were calculated as described by Ching et al.[38] The LOD, or sensitivity, was calculated by using a probit model and defined as the lowest number of input tumor cells at which the probit curve detected MRD with a type II error rate of 5%. Here, detection was considered positive when at least 2 identical clonal rearrangements were present, and negative otherwise. The LOQ was calculated as the lowest number of input tumor cells quantitated within a maximum total error of 70%. Total error was calculated as the root-mean-square error (RMSE) divided by the number of input tumor cells. RMSE was calculated as the square root of the squared bias plus the variance. Prespecified accepted values for the LOD and the lowest LOQ were both set at 2 ± 1 malignant cells per total number of input cells. All reported *P* values were obtained by a 2-sided exact method, at the

conventional 5% significance level ($P < .05$). To compare the overall performance of MRD methods, NGS results from the 2 laboratories were merged, considering "NGS positive" those cases for which MRD was detected in at least 1 laboratory, or "NGS negative" otherwise.

RESULTS

Patient Characteristics

One hundred one patients met the inclusion criteria. Follow-up samples were collected 3 months after autotransplant from 81 patients and following the end of induction from the remaining 20 patients. Median follow-up since MRD evaluation was 36.9 months (interquartile range [IQR], 27.27–44.32).

Clinical variables from our cohort are summarized in Supplemental Table 2. Male to female ratio was 62:38. Median age at diagnosis was 59 years. Patients with high-risk cytogenetics represented 21.7% (22 of 101) of the present series. The R-ISS[39] stages I, II, and III represented

Cell Input for NGS- and NGF-Based Minimal Residual Disease Studies[a]			
	NGS		NGF
Input Cells	Salamanca	San Diego	Salamanca
$< 10^5$	12.1%	39.6%[b]	0%
$10^5 - 2 \cdot 10^5$	67%	60.4%	0%
$> 2 \cdot 10^5$	20.9%	0%	100%
$> 10^6$	0%	0%	100%
Median cell input	147 224	107 800	10 001 037
Median sensitivity	$8.8 \cdot 10^{-6}$	$1.2 \cdot 10^{-5}$	$2 \cdot 10^{-6}$
Cell input range	13 035–618 464	3542–107 800	1 548 175–13 304 876

Abbreviations: NGF, next-generation flow; NGS, next-generation sequencing.

[a] Each cell shows the frequency of samples that were sequenced at different cell input levels for both methods. The number of cell equivalents used for NGS-based MRD detection was calculated from Qubit2.0 quantification, assuming 6.5 pg DNA/cell. Although NGF reached a higher sensitivity than NGS in our study (10^{-5} to 10^{-6} versus 10^{-4} to 10^{-6} ranges, respectively), MRD quantitation showed a good concordance.

[b] One sample was sequenced with a cell input lower than 10 000 cells.

21.8% (22 of 101), 62.1% (63 of 101), and 16.1% (16 of 101) of patients, respectively. At the corresponding MRD evaluation time point, the proportion of patients achieving sCR, CR, very good partial response, partial response, and stable/progressive disease was 32% (32 of 101), 20% (20 of 101), 30% (30 of 101), 14% (14 of 101), and 4% (4 of 101), respectively.

Technical Validation

Based on the data from the 5 serial dilutions and the 3 DNA input levels (2 μg, 1 μg, and 500 ng), the probit model estimated the LOD to be 1.30 myeloma cells (95% CI, 0.89–1.72) for any used total gDNA input (Supplemental Table 3). Herein, the lowest amount of MRD that could be effectively detected in 95% of measurements using 500 ng, 1 μg, and 2 μg of gDNA would be $1.7 \cdot 10^{-5}$, $8.2 \cdot 10^{-6}$, and $4.2 \cdot 10^{-6}$, respectively.

As expected, the number of input tumor cells greatly influenced precision in the quantification ability of the assay. The LOQ was set at 8 tumor cells, with total estimated errors ranging from 32.3% when using dilutions with 100 tumor cells/sample/μL to 1770% when using dilutions containing 0.85 tumor cells/sample/μL (Supplemental Table 3).

Analysis of Diagnostic Samples

Immunophenotypic Analysis.—All samples were successfully characterized by using flow cytometry. Phenotypic features and clonal plasma cell infiltration in the bone marrow met the IMWG criteria to confirm symptomatic MM diagnosis in the entire population. Median plasma cell infiltration in the bone marrow was 6.85% (IQR, 2.69%–17.16%); the median proportion of clonal, pathologic plasma cells was 99.85% (IQR, 99.33%–100%).

Molecular Analysis.—All samples but 1 (100 of 101, 99%) were monoclonal after PCR amplification, GeneScanning, and Sanger sequencing. Ninety-nine of 101 (98%) were successfully characterized by using FR1 primers. One could not be identified through FR1 Sanger sequencing but was later detected by FR2. Additionally, in another sample, the clonal rearrangement was not identified by FR1 or FR2 sequencing, most probably owing to the high polyclonal background observed by flow.

Regarding NGS by LymphoTrack, 83.2% (84 of 101), 79.2% (80 of 101), 62.4% (63 of 101), and 86.1% (87 of 101) of cases were considered monoclonal for *VDJH*-FR1, *VDJH*-FR2, *VDJH*-FR3, and *VJK* sequencing, respectively (see also

the supplemental files). Overall, when combining the 3 *IGH* regions and *IGK*, LymphoTrack showed 100% of clonality detection ability (meaning at least 1 valid result, as shown in Supplemental Table 4). Of note, LymphoTrack allowed the detection of 7 of 101 biallelic rearrangements (6.9% of the total), all undetected by Sanger. One patient with an unproductive rearrangement detected by Sanger was found to have a double-unproductive rearrangement when NGS by LymphoTrack was performed; this patient had nonsecretory myeloma. *VDJH*-FR3 sequencing using LymphoTrack showed poor performance results, with only 47 of 63 cases (74.6%) displaying the same CDR3 region that was detected by Sanger sequencing; in addition, *IGHV* genes of these sequences were misidentified in 29 of 47 cases (61.7%).

Therefore, we decided to consider only results obtained with FR1 and/or FR2: 95 of 101 cases (94%) fulfilled criteria to be considered clonal by LymphoTrack. Of these, 78 of 95 cases (82.1%) shared exactly the same CDR3 region when compared to Sanger sequencing. Fifteen of 95 (15.8%) had minor discordances in the nucleotide sequence, affecting the codification of 1 or 2 CDR3 amino acids; the last 2 of 95 cases (2.1%) were completely different in terms of clonal rearrangement and CDR3 amino acid composition (Supplemental Table 5). Thus, the LymphoTrack approach for FR1 plus FR2 was applicable in 95 of 101 cases (94%); of these, 93 of 95 (97.9%) were concordant with Sanger sequencing.

Minimal Residual Disease Analysis

After considering the results obtained with baseline samples, MRD studies were performed only in the 95 cases for which LymphoTrack was previously applicable. NGS results were compared in 82 of 95 cases (86.3%) for which parallel sequencing in Salamanca and San Diego was possible. Additionally, 11 cases were sequenced only in 1 laboratory, 3 in Salamanca (2 MRD-positive and 1 MRD-negative), and 8 in San Diego (4 MRD-positive and 4 MRD-negative). Another 2 cases were sequenced in only 1 laboratory (one in Salamanca, another in San Diego) but they could not be further analyzed owing to low total read counts (<20,000 total reads).

Ethanol precipitation often led to the loss of DNA instead of increase in its concentration, thus affecting the applicability of LymphoTrack for MRD analyses. Because of this, the minimum number of cell equivalents to reach a virtual sensitivity of 10^{-5} could only be used in 87.9% of cases (89 of
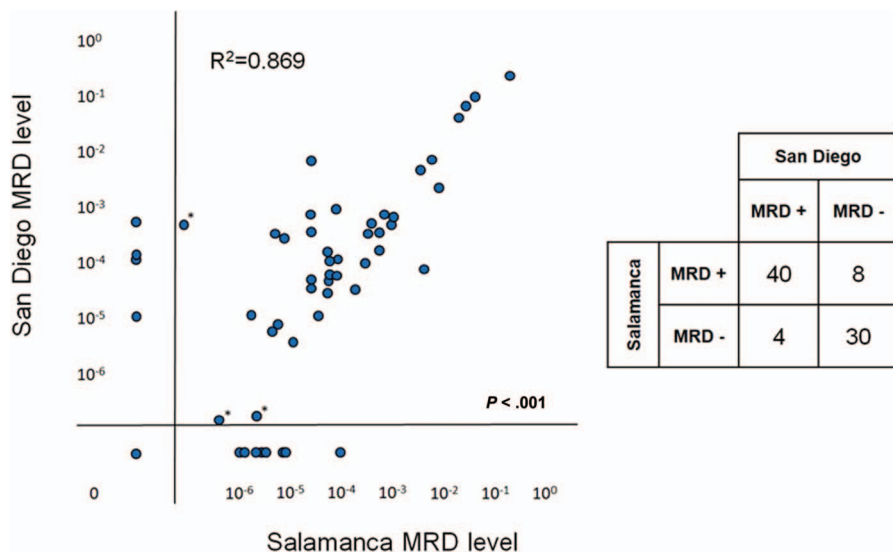
Figure 3. *Comparison of NGS-based MRD results. Seventy of 82 cases (85.3%) were concordant, while the remaining cases were positive only in 1 center (4 in San Diego and 8 in Salamanca). Cohen κ test also showed a substantial level of concordance between positive and negative cases (κ = 0.73; 95% CI, 0.58–0.89; P < .001). Of note, 3 cases (marked with an asterisk) were positive below the virtual sensitivity cutoff, probably due to variations in the DNA quantification steps or due to the variation introduced with the use of the spike-in for quantification. Abbreviation: MRD, minimal residual disease.*

101) by Salamanca and 60.4% of cases (61 of 101) by San Diego (Table). NGF was successfully used for the evaluation of our entire series, with a significantly higher number of cells required to perform the studies at the preestablished sensitivity threshold.

As shown in Figure 3, we found a good correlation for NGS between centers ($r = 0.932$; $R^2 = 0.869$, $P < .001$) with only 12 of 82 discordances (reproducibility: 85.4%) that could be due to several circumstances (Supplemental Table 6): most of the positive samples detected only in the Salamanca laboratory would be explained by a higher number of input cells than that used by the San Diego laboratory, while discordances in the other cases may be related to sample aliquoting, sample concentration, or suboptimal sequencing. High correlation was maintained between NGF and NGS performed in each center (NGS–San Diego versus NGF: $R^2 = 0.856$, $P < .001$; NGS–Salamanca versus NGF: $R^2 = 0.746$, $P < .001$) (Supplemental Figure 1, A and B). When NGS results from the 2 laboratories were merged, only 13 of 93 discordant cases (14%) between techniques were detected (4 NGF+/NGS−; 9 NGF−/NGS+). A total of 19 patients had relapse in our series, all preceded by a positive MRD result (mean, 11.6 months earlier) by NGS and NGF (Supplemental Table 7). When plotting progression-free survival curves considering only patients for whom a minimum sensitivity of 10^{-5} was achieved (Figure 4, A through C), no significant differences were observed between methods or laboratories. Thus, progression-free survival rates at 36 months for MRD-positive versus MRD-undetected were 62% versus 94% for NGS in Salamanca ($P = .003$), 59% versus 96% for NGS in San Diego ($P = .002$), and 59% versus 98% for NGF ($P < .001$), respectively.

Finally, Bland-Altman plots (Figure 5, A and B) showed a considerably high level of agreement between NGS and NGF. Overall, the mean difference in the quantification of MRD levels detected by NGF and NGS was lower than 2-fold (1.15-fold and 1.86-fold when comparing NGF with NGS in Salamanca and San Diego laboratories, respectively). This bias was significant only when comparing NGF versus NGS in San Diego (bias: −0.27; 95% CI, −0.55 to −6.2E-05; $P = .049$), showing a trend toward a slight overestimation made by NGS in this case, while it was not significant when comparing NGF versus NGS in Salamanca

(bias: 0.06; 95% CI, −1.5E-05 to 0.12; $P = .07$). Differences between methods were homogenously distributed across the range of MRD levels, with more than 70% of cases showing differences within the maximum range of ±1 log.

DISCUSSION

Minimal residual disease represents one of the most relevant prognostic factors for patients with MM, showing a significant advantage over previous, well-established prognostic factors such as cytogenetics and conventional responses.[15–17] From a revision of the results reported on clinical trials implementing this method, the IMWG encouraged the use of 2 different tools to detect MRD: EuroFlow's NGF and Adaptive's ClonoSEQ (NGS) approaches.[19] While equivalent validated methods are also accepted, there is scarce information about NGS alternatives; in fact, ClonoSEQ is the only sequencing-based method validated by the FDA to report MRD in myeloma to date.[31] To investigate new sequencing tools, we had previously evaluated the LymphoTrack assay.[32] Although our results showed that this strategy could be considered as a feasible alternative to NGF, given that the prognostic value was not altered, a more detailed characterization of its performance in both baseline and MRD settings was still needed. Here, we have evaluated LymphoTrack as a suitable and reproducible strategy for diagnostic and follow-up characterization of 101 patients with myeloma, while assessing its technical and analytical characteristics.

The LymphoTrack assay showed a high sensitivity, with a LOD of 1.3 malignant cells per reaction tube. However, the LOQ was outside the prespecified accepted values, allowing to precisely quantitate up to 8 tumor cells. Ideally, LOD and LOQ should be close to provide accurate and unbiased detection of an analyte. This level of discordance between parameters could be due to several reasons: (1) the statistical models were underpowered owing to a limited population size; (2) other factors contributing to total error (such as reagent lots, sequencer, operators, etc) were not taken into account; and (3) using a single cell line as a spike-in for normalization is a main source of error. From our perspective, the last point may be the most important factor, since amplification is not always proportionally adjusted to the initial concentration of DNA templates.[40]

Figure 4. *Kaplan-Meier plots comparing progression-free survival of patients according to their MRD status. Patients with a minimum sensitivity of 10^{-5} for MRD studies were included: N for NGS in Salamanca = 87 in (A); N for NGS in San Diego = 61 in (B); N for NGF = 99 in (C). Abbreviation: MRD, minimal residual disease.*
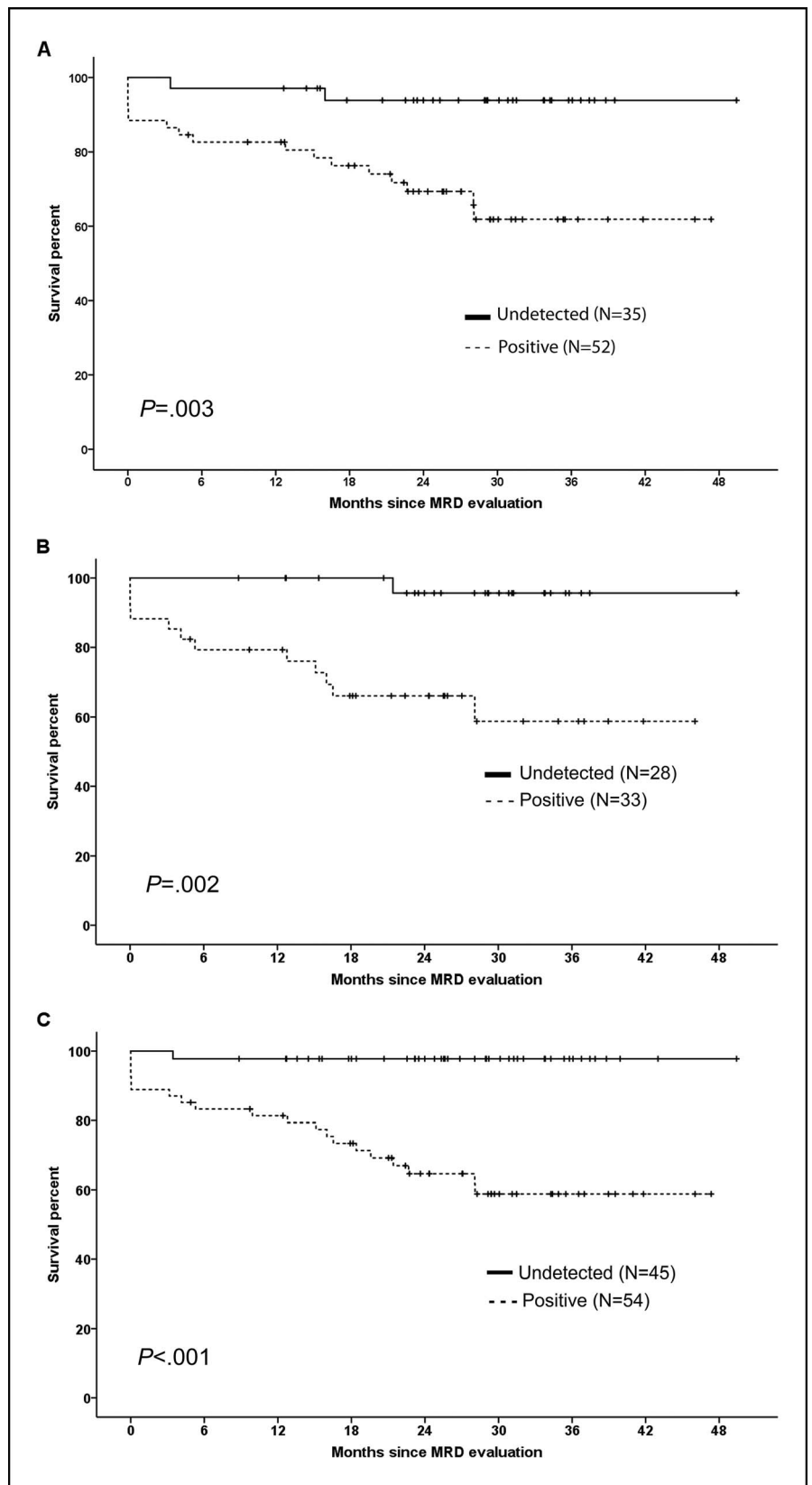
Further experiments are needed to verify the LOD and LOQ of the LymphoTrack assay, and a robust quantitation method should be developed to make this tool useful for clinical purposes.

When LymphoTrack was evaluated for clonality assessment and compared with conventional Sanger sequencing in baseline samples, the applicability of both techniques was similar, taking into account FR1 and FR2 sequencing: 95 of
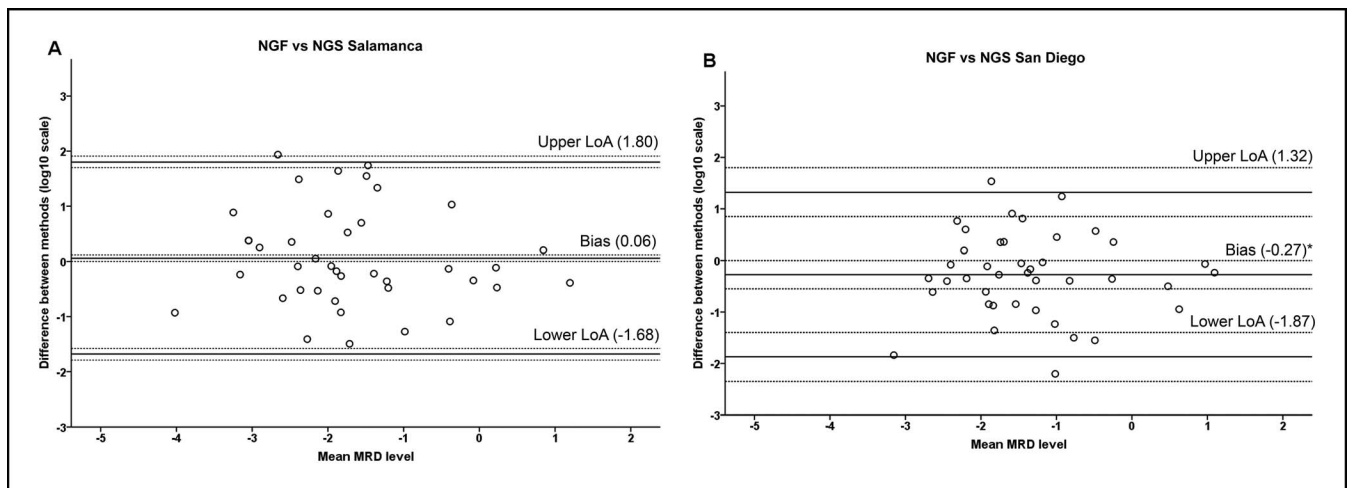
Figure 5. *Bland-Altman plots for MRD comparisons show similar performance of NGS and NGF. Mean MRD values of methods (shown in log scale on the x-axis) were calculated for 28 double-positive cases: (NGF + NGS-Salamanca) / 2 in (A); (NGF + NGS–San Diego) / 2 in (B). Normal distribution of these mean values was first assessed (Kolmogorov-Smirnov P =.09) for the 2 mean data sets. Differences between methods for each case (y-axis) were calculated as follows: log10 (higher MRD value / lower MRD value). Then, a negative value was assigned to those cases where NGS was greater than NGF, whereas a positive value was assigned to those cases where NGF was greater than NGS. The Student t test (39 degrees of freedom in plot A and 40 in plot B) was used to calculate the average of differences (bias), where a positive value indicates an overestimation made by NGF, and a negative value indicates an overestimation made by NGS. A statistically significant bias between methods is marked with an asterisk. Upper and lower limits of agreement were calculated as the bias ± 1.96 multiplied by the standard deviation of the differences. Ninety-five percent confidence interval limits for bias and limits of agreement are represented between dashed lines. Abbreviations: LoA, limit of agreement; MRD, minimal residual disease; NGF, next-generation flow; NGS, next-generation sequencing.*

101 (94%) for NGS and 100 of 101 (99%) for Sanger sequencing. Moreover, combining FR1 and *IGK* could be even more effective, as in our study 98 of 101 cases (97%) were monoclonal. Hence, our results using NGS would be in line with previous Sanger sequencing reports for MM, where up to 10% of cases can be missed owing to somatic hypermutation hampering primer annealing or to a high polyclonal background.[35] In fact, the use of longer primers for NGS would explain why some additional cases are lost. NGS primers are much longer than conventional Sanger sequencing primers (90 bp versus 25 bp) and this could be one reason to explain a poorer performance. A previous study that used LymphoTrack[41] showed even lower rates of clonality detection using only FR1 primers in plasma cell neoplasms (~70%), but these rates were improved after combining 2 or 3 targets for clonality detection, implying that this strategy should be followed to bypass the effect of hypermutation. Additionally, in this study the authors observed that applicability rates of FR1 primers greatly differ between B-cell lymphoproliferative disorders, ranging from ~60% in follicular lymphoma to more than 95% in mantle cell lymphoma, implying that other tumor characteristics play a role (eg, DNA integrity, tissue of origin, somatic hypermutation) and somehow have a greater impact on NGS.

Remarkably, the use of a lower cutoff to establish clonality by NGS (ie, 2.5% of total reads) allowed us to identify 5 tumor rearrangements that were concordant with Sanger and would not have met the threshold for clonal rearrangement if the 5% cutoff, instead of the 2.5% cutoff, had been used. In addition, 3 had detectable residual disease when baseline rearrangements detected by NGS were tracked in follow-up samples, which may suggest this threshold is optimal to confidently determine whether a rearrangement is truly tumor specific.

Nonetheless, NGS and Sanger sequencing results can be different in a small subset of cases: in our series, 17 of 95 cases (17.9%) had distinct *VDJH* usage and/or CDR3 sequences. When follow-up samples were evaluated, all the MRD-positive cases detected by NGS showed full clonotypic concordance with baseline NGS results, not with Sanger. This is crucial since it would indicate that Sanger sequencing is more error-prone and using NGS at baseline would be preferable when accurate detection of *V(D)J* rearrangements or somatic hypermutation is needed (eg, CDR3 identification as biomarker for MRD, or prognostic impact of somatic hypermutation in chronic lymphocytic leukemia). This is also a clear demonstration that *VDJH* rearrangements and CDR3 regions are stable over time[42] and are not subjected to further somatic hypermutation cycles once the myeloma clone leaves the germinal center. Compared to standard Sanger sequencing, NGS could also improve the analysis of samples with low plasma cell infiltration levels, or with a high polyclonal background, which are common findings in the bone marrow of patients with myeloma at diagnosis owing to the patchy nature of the disease and potential hemodilution.[1] In our series, 44 of 101 (43.6%) diagnostic samples had abnormal plasma cell infiltration rates below 5% of total cells by flow, and significant amplification of polyclonal B cells was observed in 31 of 101 cases (30.9%) by Sanger.

Despite being limited by the available gDNA amount in specific samples, NGS and NGF showed high concordance rates in terms of MRD detection, reproducibility of results, and quantification ability in a head-to-head comparison, even when samples were obtained from independent bone marrow pulls. The reported detection sensitivity at the level of 10[−5] in this article may not reflect what can potentially be achievable in other laboratories or clinical settings, as it was severely biased by the performance of ethanol precipitation or the maximum amount of available gDNA, which is

NGS for Clonality and MRD Detection—Medina et al **869**

sometimes low in samples extracted from the hypocellular marrow of posttreatment patients. Sequencing samples in duplicate or triplicate, increasing the total number of DNA templates, could further improve the limits of detection and the accuracy of MRD evaluation by NGS, which is a limitation of this study. It must be noted that, as several reports have stated before,[17,19] MRD evaluation in the clinical setting should be performed whenever patients achieve CR, and it gains significance with lower sensitivity, that is, when achieving the LOD of 0.0001% or 10^{-6}. While in clinical trials this helps in distinguishing truly negative, long-term surviving cases, this was not the scope of our study. Although we detected 13 discordant cases between NGS and NGF that could be explained by less restrictive experimental conditions on NGS, none of those patients have had relapse to date (median follow-up since MRD evaluation, 36.9 months; IQR, 27.27–44.32 months). By contrast, the 19 patients with relapse in our cohort (5 in CR/sCR at the time of MRD evaluation) had detectable MRD with both methodologies. Furthermore, survival plots showed an improved prognosis for those patients with undetectable MRD at a minimum sensitivity of 10^{-5}, implying that in our series, the prognostic value of MRD remained significant irrespective of the selected method. However, future implementation of stringent rules in the collection, processing, and analysis of MRD samples should be applied in order to harmonize NGS and NGF reports and reduce potential sources of error.[43]

Our results support the use of the LymphoTrack strategy to detect and evaluate clonality in patients with MM, with excellent applicability and comparable results to the current gold standard, namely Sanger sequencing. Although notable differences between protocols were present owing to suboptimal procedures and lack of standardization for this NGS strategy, the LymphoTrack-based NGS and NGF assays were also comparable in terms of MRD evaluation, indicating that both are suitable options for MRD studies in patients with myeloma, as our previous publication already pointed out.[32] A similar publication from the Memorial Sloan Kettering Cancer Center group using LymphoTrack has recently described similar findings, including a high clonality detection rate, and a good concordance with high-sensitivity flow cytometry techniques.[44] For any institution, the choice of methodology would depend on secondary factors, such as expertise or availability of equipment. Multiparametric flow cytometry has a shorter turnaround time and does not depend on the identification of clones at diagnosis, but samples must be rapidly processed to maintain cell viability, and a high level of expertise is required to deliver an accurate interpretation of the results. In contrast, NGS approaches can be successfully applied to formalin-fixed, paraffin-embedded or fresh tissues, as well as frozen DNA samples; both normal and tumor-related clones can be tracked over the course of the disease with NGS, allowing a dynamic analysis of the B-cell compartment.

Altogether, these findings support the usefulness of alternative NGS approaches in MM, demonstrating a statistically significant level of agreement with previously validated methods routinely used for clonality detection and MRD assessment, and underline the need for further standardization of quantitation procedures of the Lympho-Track assay for use as a suitable alternative to the ClonoSEQ assay.

References

1. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol.* 2014;15(12):e538–e548.
2. Durie BG, Harousseau JL, Miguel JS, et al. International uniform response criteria for multiple myeloma. *Leukemia.* 2006;20(9):1467–1473.
3. Ocio EM, Richardson PG, Rajkumar SV, et al. New drugs and novel mechanisms of action in multiple myeloma in 2013: a report from the International Myeloma Working Group (IMWG). *Leukemia.* 2014;28(3):525–542.
4. Goldschmidt H, Ashcroft J, Szabo Z, Garderet L. Navigating the treatment landscape in multiple myeloma: which combinations to use and when? *Ann Hematol.* 2019;98(1):1–18.
5. Mikhael J, Ismaila N, Cheung MC, Costello C, Dhodapkar MV, Kumar S. Treatment of Multiple Myeloma: ASCO and CCO Joint Clinical Practice Guideline. *J Clin Oncol.* 2019;37(14):1228–1263.
6. Lahuerta JJ, Mateos MV, Martínez-López J, et al. Busulfan 12 mg/kg plus melphalan 140 mg/m2 versus melphalan 200 mg/m2 as conditioning regimens for autologous transplantation in newly diagnosed multiple myeloma patients included in the PETHEMA/GEM2000 study. *Haematologica.* 2010;95(11):1913–1920.
7. Mateos MV, Oriol A, Martínez-López J, et al. Bortezomib, melphalan, and prednisone versus bortezomib, thalidomide, and prednisone as induction therapy followed by maintenance treatment with bortezomib and thalidomide versus bortezomib and prednisone in elderly patients with untreated multiple myeloma: a randomised trial. *Lancet Oncol.* 2010;11(10):934–941.
8. Rosiñol L, Oriol A, Teruel AI, et al. Superiority of bortezomib, thalidomide, and dexamethasone (VTD) as induction pretransplantation therapy in multiple myeloma: a randomized phase 3 PETHEMA/GEM study. *Blood.* 2012;120(8):1589–1596.
9. Durie BGM, Hoering A, Abidi MH, et al. Bortezomib with lenalidomide and dexamethasone versus lenalidomide and dexamethasone alone in patients with newly diagnosed myeloma without intent for immediate autologous stem-cell transplant (SWOG S0777): a randomised, open-label, phase 3 trial. *Lancet.* 2017;389(10068):519–527.
10. Dimopoulos M, Wang M, Maisnar V, et al. Response and progression-free survival according to planned treatment duration in patients with relapsed multiple myeloma treated with carfilzomib, lenalidomide, and dexamethasone (KRd) versus lenalidomide and dexamethasone (Rd) in the phase III ASPIRE study. *J Hematol Oncol.* 2018;11(1):49.
11. Rosiñol L, Oriol A, Rios R, et al. Bortezomib, lenalidomide, and dexamethasone as induction therapy prior to autologous transplant in multiple myeloma. *Blood.* 2019;134(16):1337–1345.
12. Mailankody S, Korde N, Lesokhin AM, et al. Minimal residual disease in multiple myeloma: bringing the bench to the bedside. *Nat Rev Clin Oncol.* 2015;12(5):286–295.
13. Landgren O, Owen RG. Better therapy requires better response evaluation: Paving the way for minimal residual disease testing for every myeloma patient. *Cytometry B Clin Cytom.* 2016;90(1):14–20.
14. Paiva B, García-Sanz R, San Miguel JF. Multiple myeloma minimal residual disease. *Cancer Treat Res.* 2016;169:103–122.
15. Munshi NC, Avet-Loiseau H, Rawstron AC, et al. Association of minimal residual disease with superior survival outcomes in patients with multiple myeloma: a meta-analysis. *JAMA Oncol.* 2017;3(1):28–35.
16. Lahuerta JJ, Paiva B, Vidriales MB, et al. Depth of response in multiple myeloma: a pooled analysis of three PETHEMA/GEM clinical trials. *J Clin Oncol.* 2017;35(25):2900–2910.
17. Perrot A, Lauwers-Cances V, Corre J, et al. Minimal residual disease negativity using deep sequencing is a major prognostic factor in multiple myeloma. *Blood.* 2018;132(23):2456–2464.
18. Martinez-Lopez J, Lahuerta JJ, Pepin F, et al. Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma. *Blood.* 2014;123(20):3073–3079.
19. Kumar S, Paiva B, Anderson KC, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncol.* 2016;17(8):e328–e346.
20. Paiva B, Merino J, San Miguel JF. Utility of flow cytometry studies in the management of patients with multiple myeloma. *Curr Opin Oncol.* 2016;28(6):511–517.
21. Pott C, Brüggemann M, Ritgen M, van der Velden VHJ, van Dongen JJM, Kneba M. MRD detection in B-cell non-Hodgkin lymphomas using Ig gene rearrangements and chromosomal translocations as targets for real-time quantitative PCR. *Methods Mol Biol.* 2019;1956:199–228.
22. Ladetto M, Brüggemann M, Monitillo L, et al. Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. *Leukemia.* 2014;28(6):1299–1307.

23. Drandi D, Ferrero S, Ladetto M. Droplet digital PCR for minimal residual disease detection in mature lymphoproliferative disorders. *Methods Mol Biol*. 2018;1768:229–256.

24. Cavo M, Terpos E, Nanni C, et al. Role of 18F-FDG PET/CT in the diagnosis and management of multiple myeloma and other plasma cell disorders: a consensus statement by the International Myeloma Working Group. *Lancet Oncol*. 2017;18(4):e206–e217.

25. Rasche L, Alapat D, Kumar M, et al. Combination of flow cytometry and functional imaging for monitoring of residual disease in myeloma. *Leukemia*. 2019;33(7):1713–1722.

26. Stetler-Stevenson M, Paiva B, Stoolman L, et al. Consensus guidelines for myeloma minimal residual disease sample staining and data acquisition. *Cytometry B Clin Cytom*. 2016;90(1):26–30.

27. *NCCN Clinical Practice Guidelines in Oncology: Multiple Myeloma*. Version 1.2018. NCCN; 2018.

28. Mateos MV, Dimopoulos MA, Cavo M, et al. Daratumumab plus bortezomib, melphalan, and prednisone for untreated myeloma. *N Engl J Med*. 2018;378(6):518–528.

29. Flores-Montero J, Sanoja-Flores L, Paiva B, et al. Next Generation Flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. *Leukemia*. 2017;31(10):2094–2103.

30. Bai Y, Orfao A, Chim CS. Molecular detection of minimal residual disease in multiple myeloma. *Br J Haematol*. 2018;181(1):11–26.

31. ClonoSEQ cleared for residual cancer testing. *Cancer Discov*. 2018;8(12):OF6.

32. Medina A, Puig N, Flores-Montero J, et al. Comparison of next-generation sequencing (NGS) and next-generation flow (NGF) for minimal residual disease (MRD) assessment in multiple myeloma. *Blood Cancer J*. 2020;10(10):108.

33. Green RM, Sambrook J. Precipitation of DNA with ethanol. *Cold Spring Harb Protoc*. 2016;2016(12).

34. Hodge LS, Novak AJ, Grote DM, et al. Establishment and characterization of a novel Waldenstrom macroglobulinemia cell line, MWCL-1. *Blood*. 2011;117(19):e190–e197.

35. van Dongen JJ, Langerak AW, Brüggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immuno-globulin and T-cell receptor gene recombinations in suspect lymphoprolifera-tions: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*. 2003;17(12):2257–2317.

36. Brochet X, Lefranc M-P, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res*. 2008;36:W503–W508.

37. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician*. 1983;32:307–317.

38. Ching T, Duncan ME, Newman-Eerkes T, et al. Analytical evaluation of the clonoSEQ assay for establishing measurable (minimal) residual disease in acute lymphoblastic leukemia, chronic lymphocytic leukemia, and multiple myeloma. *BMC Cancer*. 2020;20(1):612.

39. Palumbo A, Avet-Loiseau H, Oliva S, et al. Revised international staging system for multiple myeloma: a report from International Myeloma Working Group. *J Clin Oncol*. 2015;33(26):2863–2869.

40. Knecht H, Reigl T, Kotrová M, et al. EuroClonality-NGS Working Group: quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. *Leukemia*. 2019;33(9):2254–2265.

41. Arcila ME, Yu W, Syed M, et al. Establishment of immunoglobulin heavy (IGH) chain clonality testing by next-generation sequencing for routine characterization of B-cell and plasma cell neoplasms. *J Mol Diagn*. 2019;21(2):330–342.

42. Puig N, Conde I, Jiménez C, et al. The predominant myeloma clone at diagnosis, CDR3 defined, is constantly detectable across all stages of disease evolution. *Leukemia*. 2015;29(6):1435–1437.

43. Costa LJ, Derman BA, Bal S, et al. International harmonization in performing and reporting minimal residual disease assessment in multiple myeloma trials. *Leukemia*. 2021;35(1):18–30.

44. Ho C, Syed M, Roshal M, et al. Routine evaluation of minimal residual disease in myeloma using next-generation sequencing clonality testing: feasibility, challenges, and direct comparison with high-sensitivity flow cytom-etry. *J Mol Diagn*. 2021;23(2):181–199.