

Standardization of flow cytometric minimal residual disease assessment in international clinical trials. A feasibility study from the European Myeloma Network

For many decades, international collaborative efforts have driven therapeutic advances in multiple myeloma (MM). The establishment of uniform response criteria by the International Myeloma Working Group (IMWG) has been pivotal for this progress, as adherence to strict definitions ensures data comparability between trials. An essential prerequisite for the use of uniform criteria is the application of standardized methods. Of particular interest herein is the assessment of minimal residual disease (MRD) by multiparametric flow cytometry (MFC). This has been incorporated into the IMWG response criteria since 2011 to enable better risk stratification of a growing number of patients reaching a complete remission and has the promise to be used both as a surrogate marker for overall and progression-free survival and to inform treatment decisions.^{1,2} However, in contrast to most routine diagnostic tests for response assessment in MM, this assay has until recently suffered from large interlaboratory variations in terms of sample processing and data acquisition, resulting in highly heterogeneous sensitivities.³ To enable uniform and sensitive MFC MRD assessment between laboratories, EuroFlow has developed standardized operating procedures.⁴⁻⁶ Their next-generation flow method has been incorporated as the gold stan-

dard for MFC MRD measurements in the latest IMWG response criteria, which is expected to greatly improve data validity and comparability.⁷

Even though the establishment of standardized protocols has been an important step towards achieving uniform MFC MRD assessment in MM, the usefulness of these protocols depends heavily on their successful implementation in a wide range of laboratories. This is of particular relevance for international clinical trials that depend on a collaborative effort of multiple reference laboratories for timely MRD assessment, irrespective of geographical location of sampling. Nevertheless, it remains largely unknown whether fully standardized multi-laboratory MM MFC MRD assessment can be achieved in such a setting.⁸ To investigate this, a novel quality assurance (QA) program was established in 2016 within the framework of the European Myeloma Network (EMN): the EMN MRD QA program. This program aims to assess the validity and comparability of MFC MRD measurements within and between EMN trials by distributing fresh MM bone marrow and peripheral blood samples and complements existing QA programs led by EuroFlow using peripheral blood samples from healthy donors⁹ or raw data files from MM MFC MRD measurements. Data obtained within the EMN MRD QA program show that it is feasible to fully standardize MFC MRD assessment between laboratories, resulting in a high concordance over the entire range of detectable MRD levels. Participation in QA programs is essential to ensure complete interlaboratory standardization without compro-

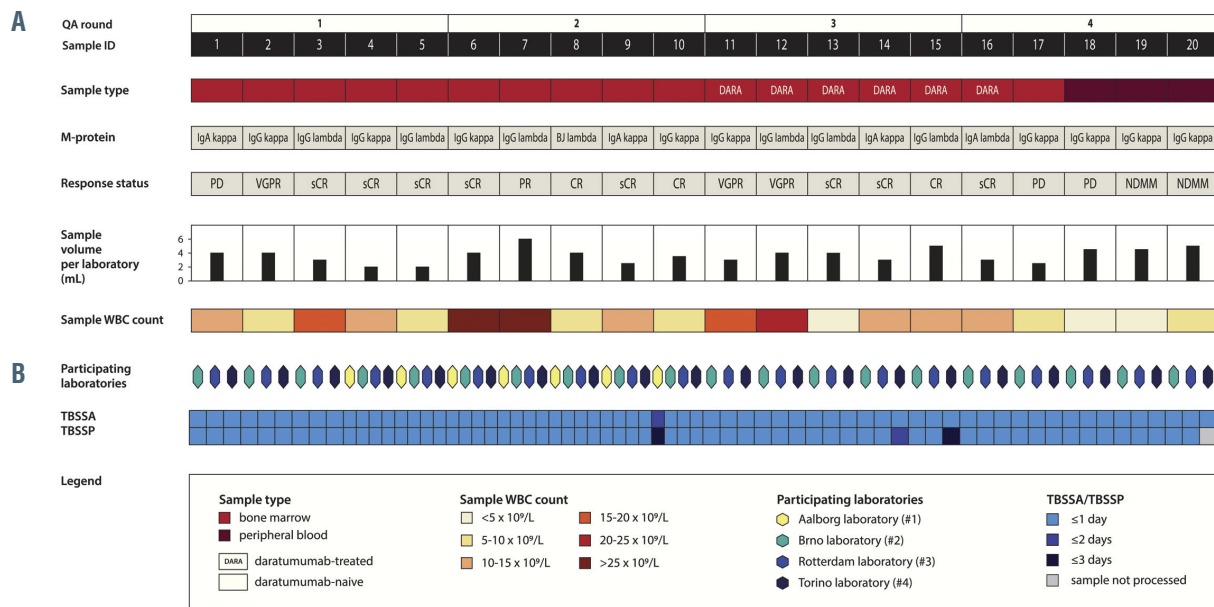


Figure 1. Sample characteristics and logistics for quality assurance rounds 1-4 of the European Myeloma Network minimal residual disease quality assurance program. A total of 20 fresh samples from patients with multiple myeloma (MM) were used in quality assurance (QA) rounds 1-4, resulting in the performance of 67 multiparameter flow cytometry (MFC) minimal residual disease (MRD) assessments in four participating laboratories in Europe. (A) In total, 17 bone marrow and three peripheral blood samples were collected from MM patients with variable disease burden, with six patients receiving daratumumab treatment at the time of sampling. Sample volumes ranged from 2-6 mL, whereas sample white blood cell counts ranged from $<5 \times 10^9/L$ to $>25 \times 10^9/L$. Response status was determined according to the International Myeloma Working Group (IMWG) 2016 criteria.⁷ (B) Between March 15, 2016 and December 17, 2019, MM samples were collected and distributed from hospitals in Rotterdam, the Netherlands and Turin, Italy. In QA rounds 1-2, second-generation flow protocols (EuroFlow) were used by all laboratories, whereas next-generation flow protocols (EuroFlow) were used in QA rounds 3-4. All laboratories participated in the full European Myeloma Network (EMN) MRD QA program, except for laboratory 1. In 2018 and 2019, this laboratory did not serve as a reference laboratory for any EMN trials requiring the use of next-generation flow protocols and therefore decided to not join QA rounds 3-4. In general, laboratories were able to process 86-100% of received samples within the IMWG recommended timeframe of 24-48 hours after sampling. CR: complete response; ID: identifier; NDMM: newly diagnosed multiple myeloma; PD: progressive disease; PR: partial response; QA: quality assurance; sCR: stringent complete response; TBSSA: time between sampling and sample arrival; TBSSP: time between sampling and sample processing; VGPR: very good partial response; WBC: white blood cell.

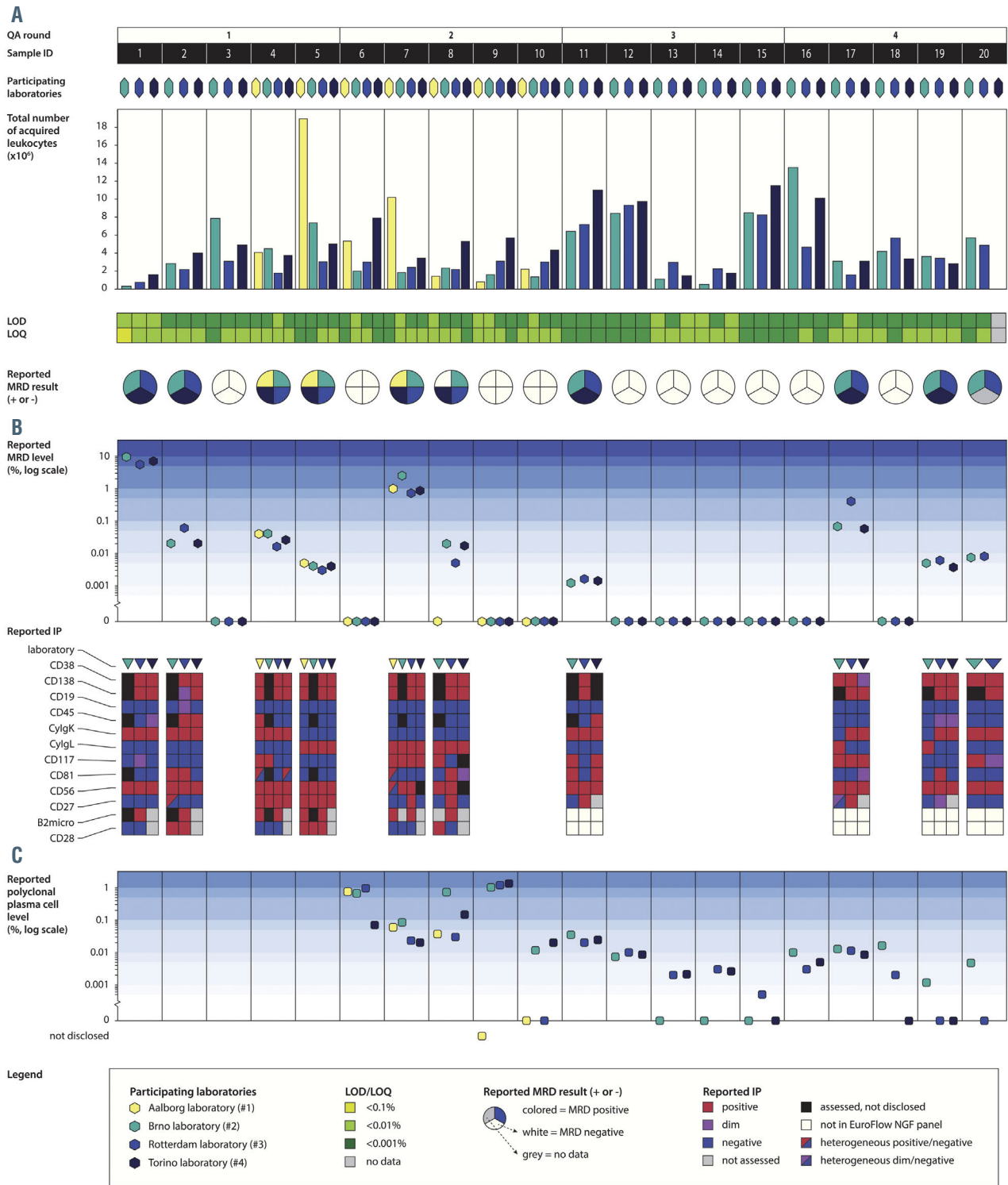


Figure 2. Concordance of minimal residual disease levels, monoclonal plasma cell immunophenotype, and polyclonal plasma cell levels between European Myeloma Network multiparameter flow cytometry laboratories. Minimal residual disease (MRD) levels were highly concordant between laboratories, irrespective of disease burden, disease stage, treatment status, sample type and quality assurance (QA) round. (A) A total of 10/20 (50%) samples were MRD-positive, which could be confirmed by all laboratories in 9/10 (90%) cases, using a cutoff of ≥ 20 monoclonal plasma cells (mPC) for MRD positivity. MRD-negative results were concordant between laboratories in 10/10 (100%) cases. In MRD-negative assays, a limit of detection $< 0.001\%$ was reached in 11/16 (69%) samples in QA rounds 1-2, versus 14/18 (78%) of samples in QA rounds 3-4. In contrast, a limit of quantitation $< 0.001\%$ was reached in 4/16 (25%) of MRD-negative assays in QA rounds 1-2, versus 9/18 (50%) of MRD negative assays in QA rounds 3-4. MRD-positive samples showed a high degree of concordance between laboratories at every level of (residual) disease, ranging from 0.001-0.01% to 1-10%. (B) Qualitative expression of essential markers for mPC gating (i.e., CD38, CD138, CD19, CD45, CD56, CyIgK, CyIgL) showed a high degree of concordance between laboratories, whereas other informative markers (i.e., CD27, CD117, CD81) showed greater variability, indicating that strict uniformization of protocols is essential to ensure reproducibility of immunophenotype data. (C) The level of polyclonal plasma cells (pPC) is commonly used as a surrogate marker for bone marrow sample quality and generally showed a good concordance between laboratories. Nevertheless, the concordance of pPC levels was inferior to that of mPC levels, suggesting that pPC levels are more susceptible than mPC levels to interlaboratory variations in sample processing and data analysis strategies. ID: identifier; IP: immunophenotype; LOD: limit of detection = 20/total number of leukocytes; LOQ: limit of quantitation = 50/total number of leukocytes;¹⁵ MRD: minimal residual disease. NGF: next-generation flow; QA: quality assurance.

mising data quality, as minor protocol deviations were commonly observed at initial implementation.

Between 2016 and 2019, MFC MRD results from 20 MM patients were compared within the EMN MRD QA program among four EMN reference laboratories willing to commit to EuroFlow protocols in the context of the EMN02/HO95 MM trial: Aalborg University Hospital, Denmark (laboratory 1), University Hospital Brno, Czech Republic (laboratory 2), Erasmus MC Rotterdam, the Netherlands (laboratory 3, EuroFlow member) and University of Turin, Italy (laboratory 4).^{10,11} In total, four QA rounds were organized, each comprising five different fresh samples from MM patients with variable levels of disease burden and variable treatment histories (Figure 1A). Samples were collected at Erasmus MC Rotterdam, the Netherlands and Ospedale Molinette di Torino, Italy, on random days throughout the year. This study was approved by the Medical Ethical Committees of Erasmus MC Rotterdam, the Netherlands and A.O.U. Città della Salute e della Scienza di Torino, Italy. Written informed consent was obtained from all participating patients, in accordance with the Declaration of Helsinki. Immediately after collection, samples were divided equally and shipped by overnight express courier to the participating laboratories. Samples from distributing hospitals were kept at room temperature for 24 h to ensure similar sample processing dates between laboratories. Using standardized forms, MRD results were collected centrally by one person, who kept these confidential until the end of each QA round, after which results were shared and discussed.

Timely sample processing is an essential prerequisite for high validity of MFC MRD results, as MM cells have a limited capacity to survive outside of the bone marrow. Hence, the IMWG recommends processing MFC MRD samples within 24–48 h. Considering all 67 samples from QA rounds 1–4, our data show that two laboratories were able to process 20/20 (100%) received samples within this recommended timeframe. Laboratory 1 processed 6/7 (86%) and laboratory 4 18/20 (90%) samples within 48 h after sampling (Figure 1B).

Throughout QA rounds 1–4, laboratory 3 adhered strictly to EuroFlow standardized operating procedures, which was considered the reference for all other participating laboratories. In QA rounds 1–2, second-generation flow protocols from EuroFlow were applied. QA round 1 was followed by a workshop to further standardize protocols and gating strategies, which resulted in the use of significantly more comparable standardized operating procedures between laboratories in QA round 2 (*Online Supplementary Tables S1 and S2*). A minimal number of 20 monoclonal plasma cells (mPC) was required for MRD positivity.¹² Despite complete standardization of protocols not being possible in laboratories 2 and 4 because of ongoing consumable contracts and local unavailability of certain reagents, MFC MRD results were highly concordant in QA rounds 1–2 at every level of residual disease. All participating laboratories reported the same MRD result for 9/10 (90%) samples (Figure 2A).

The ability to uniformly quantify MRD irrespective of daratumumab treatment status was tested in seven bone marrow samples that were distributed in QA rounds 3–4. Here, the EuroFlow next-generation flow pipeline was implemented. This pipeline contains a multi-epitope antibody against CD38 in its staining panel, which circumvents epitope blocking by daratumumab.⁵ Of note, at this stage all participating reference laboratories had committed to fully standardized protocols in terms of data collection, instrument setup, performance checks, sample

preparation, sample staining, data acquisition and data analysis (*Online Supplementary Tables S1 and S2*), resulting in a second series of highly concordant MFC MRD results and 10/10 (100%) samples with a uniformly classified MRD result (Figure 2A).

To compare interlaboratory test sensitivities in MRD-negative samples, the formula for limit of detection (LOD) was used: 20/number of acquired leukocytes. This showed a median LOD of 5.4×10^{-6} in the 34 MRD-negative samples from QA rounds 1–4. Laboratory 3 reached a LOD <0.001% in 10/10 (100%) MRD-negative assays, whereas the other laboratories achieved a LOD <0.001% in 50–80% of MRD-negative assays. Overall, in all except one assay a LOD <0.01% was reached.

Recent reports indicate that the majority of newly diagnosed MM patients have detectable mPC in their peripheral blood (i.e., circulating tumor cells) when the highly sensitive next-generation flow protocols are used.¹³ Recent reports indicate that the majority of newly diagnosed MM patients have detectable mPCs in their peripheral blood (i.e., circulating tumor cells) when the highly sensitive next-generation flow protocols are used.¹³ As mPC infiltration is typically low in both newly diagnosed MM peripheral blood and MRD bone marrow samples and the collection of peripheral blood is substantially less invasive than that of bone marrow, it has been questioned whether newly diagnosed MM peripheral blood samples could also be used for MM MRD QA purposes. The feasibility of doing so was assessed in QA round 4. Circulating tumor cells were uniformly detected in 2/2 (100%) peripheral blood samples from newly diagnosed MM patients, both at highly comparable levels between 0.001% and 0.01%. This indicates that peripheral blood samples from newly diagnosed MM patients may indeed be used as an alternative to MRD bone marrow samples to assess interlaboratory standardization of MM MRD protocols.

To test the interlaboratory concordance of the detected mPC immunophenotypes, laboratories were asked to report staining intensities as positive, dim or negative. 10/20 (50%) samples were classified as MRD-positive and generally showed strong similarity between laboratories for markers that are essential for mPC gating: CD38, CD138, CD45, CD19, CD56, CylgK and CylgL (Figure 2B). The reported expression of other informative markers (i.e., CD27, CD81 and CD117) showed more variability. Even though this did not affect mPC quantification, it underscores the importance of using strict definitions in terms of data analysis to ensure reproducibility of mPC immunophenotype data.

Finally, MFC MRD assessment has the advantage over molecular MRD techniques that it also generates information on the cellular composition of non-MM populations, which could be used to infer the quality of bone marrow samples. To this end, the EMN suggested in its consensus from 2008 that the polyclonal plasma cell (pPC) levels should always be stated in the final MRD report.¹⁴ To test the concordance of this reference population between laboratories, information on pPC levels was collected from all samples in QA rounds 2–4 (Figure 2C). As expected, peripheral blood samples had a lower median pPC level than bone marrow samples. The interlaboratory concordance of reported pPC levels was generally good, although it was inferior to that of reported MRD levels, suggesting that pPC levels are more susceptible than mPC levels to interlaboratory variations in sample processing and data analysis.

In conclusion, our data indicate that full standardization of interlaboratory MM MFC MRD assessment is fea-

sible, resulting in highly concordant MRD data. Moreover, QA programs using fresh material from MM patients are a straightforward and effective way to monitor and improve MFC MRD data quality within clinical trials. This is of particular relevance for studies that depend on reference laboratories with no or limited prior experience with the EuroFlow protocols, as these are not always fully adhered to from the beginning. We, therefore, strongly recommend the incorporation of both the annual EuroFlow QA rounds and trial group-specific QA rounds with fresh MM samples in future clinical trial designs to ensure further advancement of the field in terms of standardized MFC MRD response assessment.

Davine Hofste op Bruinink,^{1,2} Stefania Oliva,^{3*} Lucie Rihova,⁴ Alexander Schmitz,⁵ Milena Gilestro,³ Jeroen te Marvelde,² Romana Kralova,⁴ Helle Høholt,⁵ Annemiek Broijl,¹ Hans Erik Johnsen,^{5†} Roman Hajek,⁶ Mario Boccardo,³ Pieter Sonneveld,¹ Paola Omedè^{3#} and Vincent H.J. van der Velden^{2#} on behalf of the European Myeloma Network⁷*

¹Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands; ²Laboratory Medical Immunology, Department of Immunology, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands; ³Myeloma Unit, Division of Hematology, University of Turin, Turin, Italy; ⁴Department of Hematology, University Hospital Brno, Brno, Czech Republic; ⁵Department of Hematology, Clinical Cancer Research Center, Aalborg University Hospital, Aalborg, Denmark; ⁶Department of Hemato-oncology, University Hospital Ostrava and Faculty of Medicine, Ostrava University, Ostrava, Czech Republic and ⁷<https://www.myeloma-europe.org>

*DHoB and SO contributed equally as co-first authors.

#PO and VvdV contributed equally as co-senior authors.

†Deceased

Correspondence:

DAVINE HOFSTE OP BRUININK -

d.hofsteopbruinink@erasmusmc.nl

VINCENT H.J. VAN DER VELDEN -

v.h.j.vandervelden@erasmusmc.nl

doi:10.3324/haematol.2020.267831

Received: August 7, 2020.

Accepted: September 23, 2020.

Pre-published: October 5, 2020.

Disclosures: SO has received honoraria from Celgene, Janssen, Amgen, and Adaptive Biotechnologies. AB has received honoraria or advisory board remuneration from Celgene, Janssen, Amgen, and Takeda. PS has received research support and advisory board remuneration from Janssen, Celgene, and Amgen. DHoB, LR, AS, MG, JtM, RK, HH, HEJ, RH, MB, PO, and VvdV have no potential conflicts of interest.

Contributions: DHoB and VvdV conceived the study. DHoB, SO, LR, AS, MG, JtM, RK, HH, PO, and VvdV performed the investigations. DHoB, SO, and AB were responsible for resources. DHoB, SO, LR, AS, PO, and VvdV managed the data. DHoB wrote the original draft of the work, SO, LR, AS, MG, JtM, RK, HH, AB, RH, MB, PS, PO, and VvdV reviewed and edited it. DHoB: preparation of

figures. HEJ, RH, MB, PS, PO, and VvdV supervised the study. DHoB was the project administrator. RH and PS acquired funding.

Acknowledgments: the authors would like to thank all patients for providing material for this study and members from the clinical teams and laboratories at Aalborg University Hospital, Denmark, University Hospital Brno, Czech Republic, Erasmus MC Rotterdam, the Netherlands and University of Torino, Italy for their assistance.

Funding: this work was supported by grants from the European Myeloma Network and the Ministry of Health of the Czech Republic (grant number 17-30089A).

References

- Rajkumar SV, Harousseau JL, Durie B, et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. *Blood*. 2011; 117(18):4691-4695.
- 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.
- Flanders A, Stetler-Stevenson M, Landgren O. Minimal residual disease testing in multiple myeloma by flow cytometry: major heterogeneity. *Blood*. 2013;122(6):1088-1089.
- van Dongen JJ, Lhermitte L, Bottcher S, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia*. 2012;26(9):1908-1975.
- Kalina T, Flores-Montero J, van der Velden VH, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia*. 2012;26(9):1986-2010.
- 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.
- 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.
- Scott SD, Fletcher M, Whitehouse H, et al. Assessment of plasma cell myeloma minimal residual disease testing by flow cytometry in an international inter-laboratory study: Is it ready for primetime use? *Cytometry B Clin Cytom*. 2019;96(3):201-208.
- Kalina T, Brdickova N, Glier H, et al. Frequent issues and lessons learned from EuroFlow QA. *J Immunol Methods*. 2019;475:112520.
- Oliva S, Hofste op Bruinink D, Rihova L, et al. Minimal residual disease (MRD) monitoring by multiparameter flow cytometry (MFC) in newly diagnosed transplant eligible multiple myeloma (MM) patients: results from the EMN02/HO95 phase 3 trial. *J Clin Oncol*. 2017;35(15_suppl):8011.
- Cavo M, Gay F, Beksac M, et al. Autologous haematopoietic stem-cell transplantation versus bortezomib-melphalan-prednisone, with or without bortezomib-lenalidomide-dexamethasone consolidation therapy, and lenalidomide maintenance for newly diagnosed multiple myeloma (EMN02/HO95): a multicentre, randomised, open-label, phase 3 study. *Lancet Haematol*. 2020;7(6):e456-e468.
- Paiva B, Puig N, Cedena MT, et al. Measurable residual disease by next-generation flow cytometry in multiple myeloma. *J Clin Oncol*. 2020;38(8):784-792.
- Sanoja-Flores L, Flores-Montero J, Garces JJ, et al. Next generation flow for minimally-invasive blood characterization of MGUS and multiple myeloma at diagnosis based on circulating tumor plasma cells (CTPC). *Blood Cancer J*. 2018;8(12):117.
- Rawstron AC, Orfao A, Beksac M, et al. Report of the European Myeloma Network on multiparametric flow cytometry in multiple myeloma and related disorders. *Haematologica*. 2008;93(3):431-438.
- Arroz M, Came N, Lin P, et al. Consensus guidelines on plasma cell myeloma minimal residual disease analysis and reporting. *Cytometry B Clin Cytom*. 2016;90(1):31-39.