The presence of large focal lesions is a strong independent prognostic factor in multiple myeloma

Leo Rasche,† Edgardo J. Angtuaco,‡ Terri L. Alpe,Grant H. Gershner,James E. McDonald,Rohan S. Samant,Manoj Kumar,Rudy Van Hemert,Joshua Epstein,Shayu Deshpande,Ruslyna Tytarenko,Shmuel Yaccoby,Jens Hillengass,Sharmanil Thanendrarajan,Carolina Schinke,Frits van Rhee,Maurizio Zangari,Brian A. Walker,Bart Barlogie,Gareth J. Morgan,Faith E. Davies,† and Niels Weinhold†‡

†Myeloma Institute and ‡Radiology Department, University of Arkansas for Medical Sciences, Little Rock, AR; and ‡Roswell Park Comprehensive Cancer Center, Buffalo, NY

Spatial intratumor heterogeneity is frequently seen in multiple myeloma (MM) and poses a significant challenge for risk classifiers, which rely on tumor samples from the iliac crest. Because biopsy-based assessment of multiple skeletal sites is difficult, alternative strategies for risk stratification are required. Recently, the size of focal lesions (FLs) was shown to be a surrogate marker for spatial heterogeneity, suggesting that data from medical imaging could be used to improve risk stratification approaches. Here, we investigated the prognostic value of FL size in 404 transplant-eligible, newly diagnosed MM patients. Using diffusion-weighted magnetic resonance imaging with background suppression, we identified the presence of multiple large FLs as a strong prognostic factor. Patients with at least 3 large FLs with a product of the perpendicular diameters >5 cm² were associated with poor progression-free survival (PFS) and overall survival (OS; median, 2.3 and 3.6 years, respectively). This pattern, seen in 13.8% of patients, was independent of the Revised International Staging System (R-ISS), gene expression profiling (GEP)–based risk score, gain(1q), or extramedullary disease (hazard ratio, 2.7 and 2.2 for PFS and OS in multivariate analysis, respectively). The number of FLs lost its negative impact on outcome after adjusting for FL size. In conclusion, the presence of at least 3 large FL is a feature of high risk, which can be used to refine the diagnosis of this type of disease behavior and as an entry criterion for risk-stratified trials. (Blood. 2018;132(1):59-66)

Introduction

Significant progress has been made in the therapy of multiple myeloma (MM) but a subset, considered to have high-risk (HiR) disease, have not benefitted to the same extent and still undergo early disease progression within 2 to 3 years. It is crucial to develop new therapies for these patients and a prerequisite for the design of dedicated trials is the use of effective tools to identify HiR at presentation. The past decade has seen considerable effort put into the development of risk classifiers. However, the latest clinical risk stratifier, the Revised International Staging System (R-ISS), and gene expression profiling (GEP)–based risk classifiers such as the GEP70, which include molecular data for tumor cells from the iliac crest, have limited sensitivity. Thus, there is considerable potential to improve risk stratification approaches.

Focal lesions (FLs) are discrete areas of plasma cell (PC) accumulations that are important contributors to MM progression. Recently, we showed that the genomic profiles of FLs could be very different compared with iliac crest–derived PCs. We also demonstrated that HiR subclones could be restricted to FLs in a subset of patients with poor outcome and not present at the iliac crest. Thus, spatial clonal heterogeneity provides an explanation for the lack of sensitivity of current risk stratifiers, which rely on information from a single site at the iliac crest. However, biopsy assessment of multiple skeletal sites is impractical and alternative strategies are required to detect this type of “hidden” HiR disease. In this respect, we have identified the diameter of FL using medical imaging as a surrogate marker for the extent of spatial genomic heterogeneity. Specifically, large FLs with a diameter >2.5 cm are associated with site-specific enrichment of HiR driver mutations, including bi-allelic deletions of tumor suppressor genes, consistent with them being key mediators of drug resistance and treatment failure. Although the total number of FLs has been shown to be a significant prognostic factor in several imaging studies either using magnetic resonance imaging (MRI) or ¹⁸Ffluoro-deoxyglucose positron emission tomography computed tomography (FDG PET-CT), the impact of the size of individual FLs on prognosis has not been described.
With the hypothesis that the size of FLs has prognostic implications, we assessed whole-body imaging scans derived from 404 newly diagnosed MM (NDMM) patients who underwent diffusion-weighted MRI with background suppression (DWIBS). In DWIBS, FLs are visualized in a clear and reproducible fashion, which makes it the optimal technique to address our hypothesis. Furthermore, it is independent of the metabolic activity of PCs; as a result, detects disease more frequently than PET-CT.11,12

For the first time, we show that a pattern characterized by multiple large FLs is associated with poor outcome even in patients who are classified as favorable according to the R-ISS and GEP70. Accounting for FL size, the number of FLs loses its significance, consistent with FL size being the crucial imaging variable that can improve current risk stratification approaches.

### Methods

#### Patients

We included 404 transplant-eligible NDMM patients who underwent DWIBS at baseline and were enrolled into Total Therapy trials between 2009 and 2015. Total Therapy encompasses novel agent-containing induction therapy, tandem autologous stem cell transplantation, and maintenance.13,14 With a median follow-up of 5.2 years, the median progression-free survival (PFS) and overall survival (OS) rates were 5.2 years and not reached, respectively. Patients’ characteristics are summarized in Table 1. The study was performed under University of Arkansas for Medical Sciences institutional review board approval (205415); all patients signed written consent in accordance with the Declaration of Helsinki.

#### FISH and global gene expression profiling

CD138 enrichment of PCs and fluorescence in situ hybridization (FISH) were performed as published previously.15,16 R-ISS stage was determined as described previously.2 GEP of CD138-enriched PCs using Affymetrix U133 2.0 plus arrays (Affymetrix, Santa Clara, CA), GEP70-based risk designation, and GEP-based prediction of t(4;14) and t(14;16) were done as described elsewhere.17-19 Baseline FISH for del(17p) and gain(1q), and GEP were available for 97% and 100% of patients, respectively. For the GEP70 paired test, we included 21 patients for whom GEP for PCs from the iliac crest and a paired FL were available.

#### Imaging analysis

DWIBS examinations were performed on a 1.5-Tesla Philips Achieva scanner (Phillips, Koninklijke, The Netherlands). The protocol included scanning from vertex to toes in 7 to 9 slabs depending on patient height. Each slab constituted 50 slices and was scanned from vertex to toes in 7 to 9 slabs. The protocol included scanning from vertex to toes in 7 to 9 slabs. An FL was defined as a “well delineated focal intensity” above the surrounding background on DWI maps that measured ≥0.5 cm in the largest diameter. For convenience, we counted up to 20 FLs (21 was assigned to patients with >20 FLs). The size of FL was measured on DWI maps in 2 dimensions at an angle of 90°. In analogy to the Lugano classification for lymphoma by Cheson et al,20 we determined the product of perpendicular diameters (PPDs) of the 3 largest lesions. Extramedullary disease (EMD) was documented, but its size was not considered to avoid EMD as confounder of the size analysis. All DWI scans were read by 3 experienced investigators who were blinded to diagnosis and clinical parameters in consensus read.

FDG PET-CT was performed on a Biograph, Reveal, or Discovery scanner, as previously described.12 An FL was defined as a discrete focus with increased FDG uptake compared with its surroundings identified on the maximum intensity projection image, localized to bone on the fused PET-CT image on the 3 reconstructed planes.6,12,21 EMD was defined as FDG-avid lesions not adjacent to bone. Paramedullary disease (PMD) was defined as soft-tissue component arising from bone. Two independent radiologists and 1 nuclear medicine physician reported the scans and were blinded to the clinical information.

#### Statistics

The Kaplan-Meier method was used for survival analyses. PFS time was measured from enrollment to relapse or death from any cause or censored at the date of last contact. OS time was defined as time from enrollment to death from any cause. Prognostic imaging patterns were tested in a multivariate Cox regression model including established risk factors. The impact of the variables measured on patient subgroups was analyzed using recursive partitioning as implemented in the R package party.22 The Bonferroni-Holm method was used to correct for multiple testing. Wilcoxon’s or Fisher’s exact test was used to compare the median of a continuous variable or the distribution

### Table 1. Patients’ characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No. of patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>404</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>&gt;65 y</td>
<td>140 (35)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>244 (60)</td>
</tr>
<tr>
<td><strong>Myeloma type</strong></td>
<td></td>
</tr>
<tr>
<td>IgG</td>
<td>246 (61)</td>
</tr>
<tr>
<td>Non-IgG</td>
<td>158 (39)</td>
</tr>
<tr>
<td><strong>R-ISS</strong></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>76 (19)</td>
</tr>
<tr>
<td>II</td>
<td>277 (69)</td>
</tr>
<tr>
<td>III</td>
<td>46 (11)</td>
</tr>
<tr>
<td>Missing data</td>
<td>5 (1)</td>
</tr>
<tr>
<td><strong>GEP70</strong></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>349 (86)</td>
</tr>
<tr>
<td>High</td>
<td>55 (14)</td>
</tr>
</tbody>
</table>

IgG, immunoglobulin G.
of discrete variables across groups, respectively. A linear regression model was fitted to investigate the relationship between the serum lactate dehydrogenase (LDH) level and the number of FLs. Correlation coefficients were determined using Spearman’s rank correlation. Main analyses were undertaken using R (v3.3.1) software.

Results
FL were detected in 340 (84%) NDMM patients. To address the prognostic impact of FL size, we used the largest diameter and PPD as predictors of outcome. Including data for the 3 largest FLs, we showed that the PPD was the best discriminator: the presence of at least 3 FLs with a PPD of $\geq 5 \text{ cm}^2$ was significantly associated with early disease progression and death (median PFS and OS, 2.3 and 3.6 years, $P < .0001$ and $P < .0001$, respectively; Figure 1). We named this pattern FL-HiR, which was seen in 13.8% of patients ($n = 56$). Characteristics of the largest FLs in these patients are presented in supplemental Table 1 (available on the Blood Web site). The negative prognostic impact of FL-HiR was observed in each of the investigated Total Therapy trials (supplemental Figure 1). Of note, our data indicated that multiple large FLs are required to define a subset of patients with poor outcome, because a pattern with $< 3$ large FLs was not associated with shortened survival (supplemental Figure 2).

Figure 1. The impact of focal lesion size on outcome. (A) The FL-HiR pattern as seen in whole-body diffusion-weighted MRI scan with background suppression. It consists of $\geq 3$ large FLs, each with a PPD $\geq 5 \text{ cm}^2$. (B) Outcome of 404 NDMM patients enrolled into Total Therapy trials stratified by this pattern. Log-rank test was used to perform the group comparison.
The number of FLs is an established risk variable,\textsuperscript{6,8,10} and we determined its value in the current dataset. Using recursive partitioning, we identified >8 FLs as a factor, which defines a subset of patients with unfavorable PFS and OS (hazard ratio [HR], 1.47, \(P = .006\); and HR, 1.62, \(P = .008\); supplemental Figure 3). However, taking the FL-HiR pattern into account, the FL number (as a dichotomous or continuous variable) lost its association with poor outcome (\(P > .05\) for PFS and OS; supplemental Figure 2). The same holds true for >3 FDG-avid FLs, the FL number prognostic for PET-CT (supplemental Figure 4).\textsuperscript{6,8}

Together, the presence of at least 3 large FLs on DWIBS is associated with poor outcome and the total number of FLs lost its negative prognostic effect when the data were adjusted for FL size.

**Correlation of the FL-HiR pattern with PMD and EMD**

EMD is a strong risk factor for adverse outcome,\textsuperscript{23,24} and a trend to adverse outcome was also shown for breakout FLs with paramedullary components.\textsuperscript{25} To address the question of whether the negative impact of large FLs merely reflects the presence of soft-tissue components, we incorporated data from PET-CT, which is considered the optimal method for detection of EMD and PMD.\textsuperscript{26}

We detected soft-tissue components in 52 (93%) FL-HiR patients (39 PMD, 2 EMD, 11 PMD and EMD). The impact of FL-HiR was independent of EMD (\(P < .001\); HR, 2.8 and 2.5 for PFS and OS, respectively, after adjustment for EMD; supplemental Figure 5). Although PMD was associated with adverse prognosis, the negative impact was not seen in patients without multiple large FL (n = 99 patients; supplemental Figure 6). Together, these results indicate that soft-tissue manifestations are enriched in patients with at least 3 large FLs, but alone do not explain the negative prognostic effect of this pattern.

For this analysis, we used the EMD status determined using PET-CT. Because EMD is an important prognostic factor, we also assessed the performance of DWIBS in detecting this type of disease distribution. Using DWIBS, only 11/23 patients with EMD according to PET-CT scans were identified. This discrepancy mainly resulted from solitary PET-avid lymph nodes suspicious for EMD, which were either too small (<1.5 cm) or considered to be nonspecific/nonreactive on DWIBS (6 cases in our series). EMD, which was detectable on PET-CT only, was still prognostically relevant (\(P = .005\); HR, 2.5 compared with cases without EMD), suggesting that the respective lymph nodes were not just reactive.

In summary, the prognostic impact of FL-HiR is independent of soft-tissue manifestations; these extramedullary components are best detected using PET-CT.

**FL-HiR is an independent prognostic marker**

It is important to take account of other established prognostic factors, such as R-ISS\textsuperscript{2} and GEP70.\textsuperscript{3} Therefore, we investigated whether the FL-HiR pattern was an independent predictor of outcome when these classifiers and other prognostic markers were included in the analysis. The FL-HiR pattern was not significantly associated with tumor-load parameters, such as ISS III or a bone marrow PC infiltration >75%. However, the negative prognostic molecular markers del(17p), gain(1q), and GEP70 HiR were more frequently seen in these patients (\(P < .05\); Figure 2). The FL-HiR pattern was not associated with LDH level (\(P > .05\)), in contrast to the number of FLs on DWIBS, which showed a significant positive association with these levels (\(P = .01\)).

Next, we performed a multivariate analysis including FL-HiR together with the established risk factors R-ISS, GEP70, gain (1q), and EMD according to PET-CT scans. Furthermore, we included the number of FL on DWIBS. Statistics for univariate tests are shown in supplemental Table 2. We did not include albumin, \(\beta\)-2 microglobulin, LDH, del(17p), t(4;14), and t(14;16), because these factors are used to determine the R-ISS. In this multivariate analysis, FL-HiR remained significant (HR, 2.7, \(P < .001\); and HR, 2.2, \(P = .001\) for PFS and OS, respectively), supporting its value as a complementary risk factor for early relapse and death in MM (Figure 2). Kaplan-Meier plots for the combination of FL-HiR with the R-ISS or the GEP70 are shown in Figure 3. These results highlight the power of DWIBS to identify patients with adverse outcome who would be classified as low risk (LR) according to these 2 established classifiers. Interestingly, gain (1q21) also independently affected PFS and OS, suggesting that this aberration may improve the R-ISS, which only takes the HiR molecular aberrations del(17p), t(4;14), and t(14;16) into account.

**Spatial risk profiles in FL-HiR patients**

Despite the enrichment of various established poor prognostic markers, the FL-HiR pattern was independently associated with PFS and OS in the multivariate analysis. One possible explanation for this observation could be that HiR disease is not homogeneously distributed in these patients and therefore not detectable in samples from the iliac crest. In an initial attempt to address this hypothesis, we studied GEP70 risk profiles of PCs from large FLs in comparison with PCs derived from the iliac crest. For 21 FL-HiR patients with GEP70 LR disease according to the iliac crest sample, GEP for at least 1 large FL was available; of those, 4 had GEP70 HiR disease in the FL. In another 12, we observed increased GEP70 scores in the FL compared with the paired iliac crest sample. Overall, GEP70 scores were significantly higher in large FL in comparison with the iliac crest (median GEP70 score 0.29 vs 0.02, \(P = .003\)). Although this observation is based on small numbers and needs to be confirmed in future studies, it suggests that, at least in a subset of patients, advanced disease located in FLs contributes to the independent prognostic impact of FL-HiR.

**Discussion**

Spatial genomic heterogeneity is a frequent feature of NDMM and poses a significant challenge for risk stratification and targeted treatment approaches.\textsuperscript{2,27} We recently showed that the size of FLs could be a surrogate marker for spatial genomic heterogeneity.\textsuperscript{7} Here, we have addressed the clinical relevance of FL size as a prognostic marker and show for the first time that it negatively affects the outcome of NDMM patients. A pattern, characterized by the presence of a minimum of 3 FL with a PPD of 5 cm\textsuperscript{2} (FL-HiR), was seen in \(\sim 14\)% of patients and was associated with early disease progression. Notably, this dismal outcome occurred despite treatment with multiagent induction therapy, 2 transplants, and intense maintenance. FL-HiR is a pattern that is independent of R-ISS and GEP70. It is a new and clinically useful risk factor in MM, which highlights the value of medical imaging as a complementary approach that can enhance risk stratification.
Figure 2. Association test and multivariate analysis including the FL-HiR pattern. (A) Association between the FL-HiR pattern and disease features. Statistical significance was determined by Fisher's exact test. (B) Forrest plot illustrating the results of a Cox regression model including the depicted variables. Hazard ratios and 95% confidence intervals are indicated by squares and crossing horizontal lines, respectively. *P < .05; **P < .01; ***P < .001. AIC, Akaike information criterion.

### Hazard ratio for progression free survival

<table>
<thead>
<tr>
<th>FL-HiR</th>
<th>FL Number</th>
<th>GEP70</th>
<th>R-ISS</th>
<th>EMD</th>
<th>gain(1q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FL-HiR</th>
<th>FL Number</th>
<th>GEP70</th>
<th>R-ISS</th>
<th>EMD</th>
<th>gain(1q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

### Hazard ratio for overall survival

<table>
<thead>
<tr>
<th>FL-HiR</th>
<th>FL Number</th>
<th>GEP70</th>
<th>R-ISS</th>
<th>EMD</th>
<th>gain(1q)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td>reference</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

# Events: 176; Global p-value (Log-Rank): 2.2204e-15
AIC: 1841.53; Concordance Index: 0.68

# Events: 100; Global p-value (Log-Rank): 2.2204e-16
AIC: 1039.97; Concordance Index: 0.74
This is not the first study to use medical imaging to risk stratify NDMM. Patients presenting with >7 FLs on axial MRI were shown to be associated with adverse outcomes, a number recently confirmed in an independent study.\(^8\) For PET-CT, a FL number >3 was identified as a marker for poor prognosis.\(^6,8\) In the current dataset, a FL number >8 was associated with shortened survival. However, the proportion of patients with >8 FLs was nearly 50% of the cases, suggesting that the overall group was enriched for patients at high risk of early relapse, rather than representing a distinct subgroup. Consistent with this proposal, the total number of FLs lost its significance after correction for FL size.

However, the number of large FLs remains important because a pattern with only 1 or 2 large lesions was not associated with poor outcome. The biology underlying this phenomenon remains elusive. One possible explanation is that multiple large FLs are linked to a higher probability of HiR drug-resistant disease that is restricted to FLs in patients who would otherwise be classified as LR. Indeed, such a pattern was seen in patients with higher GEP70 risk scores in FLs. However, because a molecular analysis of all large FLs per patient is not feasible, we have not sufficiently addressed this hypothesis and our data simply indicate that a nonhomogeneous distribution of HR disease contributes to the prognostic impact of FL-HiR. An alternative explanation is that the presence of multiple large FLs reflects an advanced stage of tumor evolution associated with an increased level of intrapatient genomic heterogeneity. In this latter concept, the clone resulting in relapse does not necessarily originate from large FLs, but these lesions are indicators of a critical level of intraclonal diversity, which is more likely to be associated with an increased risk of the emergence of drug-resistant subclones. To fully address this concept, a spatiotemporal study will be required that is based on multiregion sequencing, whole-body imaging, and probably an analysis of circulating cell-free DNA or PCs.\(^28-31\) Although in this work we have focused on FL size and its effect on outcome, future efforts investigating other FL features, such as anatomical location, shape, restriction (ADC values), and metabolic activity are warranted.

In this study, we used whole-body DWIBS to determine the size of FLs. DWIBS has emerged as a powerful technique for staging and follow-up of malignant diseases, which can be performed on state-of-the-art MRI systems supplied by all major vendors.\(^32\) Besides its wide availability, other advantages of this functional imaging technique include the short scan time (~30 minutes using the protocol described in this study) and, compared with PET-CT, its lack of prescan diet requirements, exposure to radioactive tracers, or dependence on the metabolic activity of tumor cells. The main advantage of this technique, which prompted us to use it, is that it delineates focal involvement in a clear and reproducible fashion making it the optimal approach for the detection of the FL-HiR pattern. However, we expect that this pattern can also be
detected using other imaging techniques such as conventional MRI or PET-CT. Furthermore, DWIBS also has some drawbacks, such as the limited sensitivity for detection of EMD. In our opinion, the functional techniques DWIBS and PET-CT are complementary and should be combined in future studies.

In conclusion, the presence of ≥3 large FLs on DWIBS adds an additional HiR criterion that can be used to identify patients for dedicated HiR trials. Because this pattern is essentially a simple visual diagnosis, we assume that it will be rapidly integrated into clinical risk stratification approaches.

Acknowledgments
The authors thank the patients and staff of the Myeloma Institute, University of Arkansas for Medical Sciences and the Department of Radiology, University of Arkansas for Medical Sciences.

This work was supported by grants from the National Institutes of Health, National Cancer Institute (P01 CA 55819), the Deutsche Forschungsgemeinschaft (L.R.), and the National Institutes of Health, National Institute of General Medical Sciences (P20GM125503) (N.W.).

Authorship
Contribution: L.R., N.W., F.E.D., and G.J.M. undertook conception approaches. L.R., N.W., and G.J.M. wrote the paper; and all authors reviewed and approved the paper.

Conflict-of-interest disclosure: B.B. is a coinventor on patents and patent applications related to use of gene expression profiling in cancer medicine that have been licensed to Quest Diagnostics. The remaining authors declare no competing financial interests.

Edgardo J. Angtuaco died on 28 May 2018.

Correspondence: Leo Rasche, Myeloma Institute, University of Arkansas for Medical Sciences, 4301 W. Markham, #816, Little Rock, AR 72205; e-mail: lrasche@uams.edu.

Footnotes

*L.R. and E.J.A. contributed equally to this study.

†F.E.D. and N.W. co-directed this study.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

REFERENCES
24. Moreau P, Attal M, Caillot D, et al. Prospective evaluation of magnetic resonance imaging...


