Article:
Kuiper, R, van Duin, M, van Vliet, MH et al. (11 more authors) (2015) Prediction of high- and low-risk multiple myeloma based on gene expression and the International Staging System. Blood, 126 (17). pp. 1996-2004. ISSN 0006-4971

https://doi.org/10.1182/blood-2015-05-644039

Prediction of high- and low-risk multiple myeloma based on gene expression and the International Staging System

Running head: EMC92-ISS RISK-STRATIFICATION IN MULTIPLE MYELOMA

Rowan Kuiper[1], Mark van Duin[1], Martin H. van Vliet[2], Annemiek Broijl[1], Bronno van der Holt[3], Laila el Jarari[3], Erik van Beers[2], George Mulligan[4], Hervé Avet-Loiseau[5], Walter M. Gregory[6], Gareth Morgan[7], Hartmut Goldschmidt[8], Henk M. Lokhorst[9] and Pieter Sonneveld[1]

[1]Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands; [2]SkylineDx, Erasmus MC Cancer Institute, Rotterdam, The Netherlands; [3]HOVON Data Center, Erasmus MC Cancer Institute-Clinical Trial Center, Rotterdam, The Netherlands; [4]Millennium Pharmaceuticals, Cambridge, MA, USA; [5]Unité de Génomique du Myélome, CHU Rangueil, Toulouse, France; [6]Clinical Trials Research Unit, University of Leeds, Leeds, UK; [7]Royal Marsden Hospital, London, UK; [8]University of Heidelberg, Heidelberg, Germany; [9]Department of Hematology, University Medical Center Utrecht, Utrecht, The Netherlands.

Abstract word count: 187

Article word count: 2729

Number of figures: 4 sets

Number of tables: 3 sets

Number of References: 40

Correspondence:

Prof. P. Sonneveld, MD, PhD

Erasmus MC Cancer Institute

Department of Hematology, Na-822

P.O. box 2040

3000 CA Rotterdam

The Netherlands

E-mail: p.sonneveld@erasmusmc.nl

Phone: +31·10·70 33740

Fax: +31·10·70 35814

Cell: +31 6 11 60 82 30

Scientific section designations: LYMPHOID NEOPLASIA

Key Points:

- Combination of ISS and the EMC92 gene expression classifier is a novel clinically applicable risk classification for survival in multiple myeloma.
- ISS has clear independent additive prognostic value in combination with GEP classifiers or FISH markers.

Abstract

Patients with multiple myeloma have variable survival, and require reliable prognostic and predictive scoring systems. Currently, clinical and biological risk markers are used independently. Here, ISS, FISH markers and gene expression (GEP) classifiers were combined to identify novel risk classifications in a discovery/validation setting. We used the datasets of HOVON-65/GMMG-HD4, UAMS-TT2, UAMS-TT3, MRC-IX, APEX and IFM-G (total number of patients: 4750). A total of 20 risk markers were evaluated including t(4;14) and deletion of 17p (FISH), EMC92 and UAMS70 (GEP classifiers) and ISS. The novel risk classifications demonstrated that ISS is a valuable partner to GEP classifiers and FISH. Ranking all novel as well as existing risk classifications showed that the EMC92-ISS combination is the strongest predictor for overall survival, resulting in a four group risk classification. The median survival was 24 months for the highest risk group, 47 and 61 months for the intermediate risk groups and median not reached after 96 months for the lowest risk group. The EMC92-ISS classification is a novel prognostic tool, based on biological and clinical parameters, which is superior to current markers and offers a robust clinically relevant 4-group model.

Introduction

In multiple myeloma (MM) patients, malignant plasma cells accumulate in the bone marrow, leading to a wide range of clinical symptoms which include bone disease, hypercalcemia, renal impairment and anemia.[1] The prognosis is variable, with survival for newly diagnosed patients ranging from less than two to more than twenty years.[2] Adequate prognostication of disease outcome is important in order to make treatment choices and to allocate high-risk patients to alternative treatment options. Clinical trials that address specific treatment of high-risk patients include TT4, TT5 and MUK9 (TT4: Total Therapy 4, NCT00734877; TT5: Total Therapy 5, NCT02128230; MUK9, OPTIMUM trial, Myeloma UK Clinical Trial Network).

Heterogeneous treatment outcome can in part be explained by different biological subgroups in MM, which are characterized by primary translocations involving genes such as MMSET (t(4;14)), and c-MAF (t(14;16)).[3,4] These subgroups can be identified using gene expression profiling.[5,6] In addition, gene expression profiling has been utilized to establish classifiers for prognostication. The EMC92 is a robust risk marker for the identification of high-risk MM, and was validated in independent clinical trials showing a solid and independent performance in comparison to other MM GEP classifiers such as UAMS70.[7-13] Clinical prognostic systems for MM, are primarily based on beta$_2$-microglobulin (β_2M), albumin, lactate dehydrogenase, C-reactive protein, calcium and creatinine.[14,15] The International Staging System (ISS) is based on β_2M and albumin, with stage I representing limited disease, stage II intermediate and stage III the most unfavorable disease.[16] Today it is used as the standard clinical risk classification for MM.

FISH based cytogenetics and gene expression profiling are biology based prognostic markers.[17] ISS was combined with high-risk cytogenetic markers t(4;14) and deletion of 17p (del17p) to establish novel prognostic risk classifications as proposed by Neben[18] and Avet-Loiseau[19]. Recently, serum lactate dehydrogenase (LDH) was added as a component to this

marker combination.[20] Other prognostic systems include combinations of cytogenetic markers, such as the combination of del17p, translocation t(4;14) and gain of 1q (gain1q).[21]

The goal of this study was to evaluate all published risk markers used in MM and to compare combinations of FISH, ISS and GEP based prognostic systems. By applying a study design with independent discovery and validation sets, we demonstrated that ISS can be combined with gene expression signatures into powerful classifiers for MM.

Methods

Clinical data

The clinical data from the HOVON-65/GMMG-HD4 (HO65/HD4), MRC-IX, UAMS-TT2, UAMS-TT3, IFM-G (all newly diagnosed patients) and APEX (relapse patients) trials were used.[7-9,19,22,23] The IFM-G cohort is a clinical database of patients not separately published and was included in the ISS development.[16] Treatment regimens of the trials from which these datasets were derived are summarized in Table 1. Overall survival (OS) or progression-free survival (PFS) and at least one prognostic marker were available for all patients (Table 1; Figure S1). All patients signed an informed consent in accordance with the Declaration of Helsinki and all protocols were approved by institutional review boards.

Gene expression profiling (GEP)

All GEP data are Affymetrix HG U133 Plus 2.0 platform based, except for the APEX study (Affymetrix U133 A/B platform). HO65/HD4 GEP was performed in our lab as described previously (n=327; GSE19784).[6,7,21] Other GEP sets were: TT2 (n=345; GSE24080)[8], TT3 (n=238; E-TABM-1138 and GSE24080)[24], MRC-IX (n=247; GSE15695)[22] and APEX (n=264; GSE9782)[23]. Due to unavailable survival data, the HM dataset (n=206; E-MTAB-362), was

used only to determine the probe set means and variances for the training set of the HM19 classifier.[12]

Standard prognostic markers

Availability of risk markers and patients per dataset is shown in Table 1 and Figure S1. The International staging system (ISS) was determined by combining serum levels of β_2M and albumin.[16] Cytogenetics by Fluorescence in situ hybridization (FISH) was used with a 10% cut-off level except for a 20% cut-off used for numerical abnormalities in the MRC-IX trial.[19,25-27] Gain of chromosome 9 (gain9), one of the hyperdiploid chromosomes and most frequently available marker for this purpose, was used as a proxy for hyperdiploidy.[28] FISH probes used in MRC-IX and HO65/HD4 were described before.[25,29] Cytogenetic data obtained by methods other than FISH were excluded. High-risk FISH was defined as having either del17p or t(4;14) or gain1q, denoted here as HR.FISH.A.[21] The risk classification described by Avet-Loiseau *et al.*[19], is denoted here as HR.FISH.B/ISS. This risk classification distinguishes grade-I (ISS=1 or 2 with FISH markers t(4;14) and del17p both negative), grade-II (not grade-I or III) and grade-III (ISS=2 or 3 with FISH markers t(4;14) or del17p positive). In case of an arbitrary situation due to missing data for one of the markers, the observation was excluded.

Gene expression classifiers

The following MM gene expression classifiers were used: EMC92[7], UAMS17[8], UAMS70[8], UAMS80[9], IFM15[10], MRCIX6[13] (all two risk group classifiers) and HM19[12], GPI50[11] (both three risk group classifiers). Normalization and cut-offs were calculated as described previously (see supplemental methods for a brief description).

Statistical analyses

In Figure 1, a flowchart of the analyses is given. The association of risk markers with survival was assessed using a Cox survival model (R '*survival'* package, version 2.38-1).[30-32] To account for heterogeneous survival between studies, models were stratified per trial cohort. **The trial cohorts were HO65/HD4, MRC-IX intensive, MRC-IX non-intensive, UAMS-TT2, UAMS-TT3, IFM-G and APEX. Datasets used for generating risk markers were systematically excluded in validation analyses in order to avoid training bias. For instance, HO65/HD4 patients were excluded in analyses involving the EMC92 classifier (Table 1).**

The method for finding novel combination markers (compound markers) is illustrated in Figure S2B and extensively described in the supplemental methods. Briefly, since missing data may confound the analyses, combinations with increased risk for confounding were excluded (Table S1; supplemental methods). Subsequently, the data were randomly split into a discovery and validation set. The discovery set was used for finding meaningful combinations of markers as well as the most optimal way to split patients into subgroups, using these combinations. Stringent validation was performed in the designated validation set to confirm their prognostic strength. Finally, all new combinations and existing markers were ranked, with a low rank score indicating a high performing risk marker.

Results

Confirmation of existing risk markers

The value of 20 existing risk markers was evaluated in a data set of 4750 patients. The markers and used cohorts are given in Table 1. The prognostic value was evaluated correcting for the differences in survival between cohorts (Figure 2; Figures S3-5; Table S2). For all markers at least 2 cohorts were available. All gene expression (GEP) classifiers demonstrated a highly significant performance for OS. Hazard ratios for GEP classifiers ranged from 2.0 (95%CI = 1.6 - 2.4; IFM15) up to 3.3 (2.6 - 4.3; UAMS70). Furthermore, hazard ratios for GEP classifiers were consistently higher than any of the other risk markers, including all FISH markers and ISS. This suggests better risk separation for GEP classifiers compared to FISH markers. GEP classifiers generally performed better for OS than for PFS (Figures S3A-B, S4 and S5; Table S2) with PFS hazard ratios between 1.8 (95%CI: 1.5 - 2.1; IFM15) up to 2.3 (1.9 - 2.7; EMC92). The percentage of high-risk patients varied between classifiers: 18% (EMC92), 12% (UAMS17), 10% (GPI50), 9% (UAMS70), 8% (UAMS80 and HM19; Table 1).

FISH markers with prognostic strength can be distinguished from markers with no or disputable value. For OS, markers t(4;14), del17p, gain1q and del13q performed well with hazard ratios ranging between 1.7 (95%CI: 1.5 - 1.8; del13q) up to 2.3 (2.0 - 2.6; del17p). The markers gain9, t(11;14), t(14;16) and t(14;20) were clearly not significant or had high variance due to lack of predictive value or small number of positive cases. These markers were excluded from further analyses. A similar pattern was found for PFS, but the strength of the markers was generally lower with PFS hazard ratios ranging from 1.4 (95%CI: 1.3 - 1.5; del13q) up to 1.8 (1.6 - 2.0; t(4;14)).

ISS was confirmed as a valuable and highly significant prognostic marker. A hazard ratio of 1.6 (95%CI: 1.4 - 1.8; ISS=2) and 2.3 (2.1 - 2.6; ISS=3) was found for OS and 1.4 (1.3 - 1.6; ISS=2) and 1.7 (1.6 - 1.9; ISS=3) for PFS.

Other previously published compound risk markers, denoted here as HR.FISH.A[21] (either t(4;14) or del17p or gain1q) and a combined FISH/ISS marker (HR.FISH.B/ISS[19]) showed good performance. The hazard ratio was 2.3 (2.0 - 2. 5; HR.FISH.A). For the three group HR.FISH.B/ISS risk classification, hazard ratios of 1.8 (1.4 - 2.4; intermediate risk) and 3.6 (2.7 - 4.7; high-risk) were found.

To correct for heterogeneity between studies, all analyses were corrected for the survival differences between trials as a result of differences in treatment, disease stage and patient populations. To evaluate the effect of this correction, all analyses were repeated per cohort and highly similar results were obtained, suggesting that these risk markers perform similarly across different cohorts (supplemental results).

Pair-wise combinations of risk markers

The next analysis was performed to explore combinations of risk markers. As indicated above, 16 of 20 evaluated markers had significant associations with OS and/or PFS. Based on these 16, all possible pair-wise combinations were generated. 20 combinations were significant in the discovery set of which 16 remained significant in the independent validation set (Figure 2 and Figure S8A-B; Table S3). In 10 of 16 combinations, ISS was combined with either GEP classifiers (n=5) or FISH markers (n=5), illustrating the strong additive power of ISS to these markers. Combinations of GEP (n=3) and FISH markers were observed (n=3), but no combinations of FISH with GEP. Two combinations divided patients in 3 groups, ten in 4 groups and four into 5 groups.

Ranking of existing and novel markers

The markers described above, i.e. 16 existing plus 16 validated new risk markers, were ranked on the basis of performance, as described in the Supplemental methods.

ISS-GEP combinations consistently ranked at the top with the EMC92-ISS compound risk marker having the best median rank score (RS) (Figure 3; RS = 0.05). Other high scoring markers included ISS-UAMS17 (RS = 0.11), ISS-HM19 (RS = 0.13) and ISS-UAMS70 (RS = 0.19). The HR.FISH.B/ISS compound marker ranked in 5th place (RS = 0.20) and ISS ranked in 23rd place (out of 32; RS = 0.61). In general, compound markers tended to score better than single markers. The best single marker was EMC92 in 7th position (RS = 0.26). EMC92-ISS classifies patients into four groups with proportions of 38%, 24%, 22% and 17% for the lowest to the highest-risk group, respectively (Figure 4A-B). The hazard ratios relative to the lowest-risk group were 2.6 (1.6 - 4.5; intermediate-low), 3.2 (1.9 - 5.4; intermediate-high) and 6.9 (4.1 - 11.7; high). Median survival times were 24 (high), 47 (intermediate-high) and 61 months (intermediate-low) for the three highest-risk groups, with median survival not reached after 96 months for the lowest-risk group. **To gain insight into the performance of this marker over time, we determined the proportions of surviving patients in each risk group and analyzed the EMC92-ISS at different time points. This marker is clearly applicable to younger as well as older and relapsed patients, and holds its value during follow up (Table 2, Figure S10).**

The composition of the four groups in terms of ISS, EMC92 and FISH markers is shown in Table 3. Interestingly, within the EMC92-ISS lowest-risk group, 75% of patients – with truly favorable prognosis (Table S4) – were positive for either t(4;14), del(17p) or gain1q. In the other risk categories 32%, 42% and 86% of patients were positive (intermediate-low,

intermediate-high and high-risk, respectively) indicating that EMC92-ISS and FISH only partly represent overlapping patient sets.

Biological relevance of GEP classifiers

Genes within GEP classifiers are selected based on association with survival, rather than a direct link to biology. Still, a gene ontology enrichment analysis[33] can highlight biological processes important for a poor outcome (Tables S5A-H). All GEP classifiers had enrichment of cell-cycle related genes. When all probe-sets in all classifiers were pooled 191 biological processes were found to be enriched (FDR <0.05). Top processes included 'nuclear division', ‘mitosis’ and ‘cell division’, processes sharing the genes BIRC5, BUB1 and UBE2C. Other prominent processes included ‘DNA metabolic process’, ‘DNA packaging’ and ‘DNA replication’ (genes such as TOP2A and MCM2).

Discussion

Important prognostic markers in MM are based on ISS, FISH markers and GEP classifiers.[7-13,16,17] Previously, we showed that combining various GEP classifiers resulted in a stronger prediction of the high-risk population.[7] Here we systematically evaluated additional, new combinations of prognostic markers. We limited the search for new compound risk markers to pair-wise combinations of existing markers. This choice is mainly driven by the lack of complete data sets which contain all risk markers (as shown in Figure S1), which hinders the analyses of more complex risk-models The number of patients positive for specific markers was remarkably stable between cohorts, irrespective of the type of marker. This adds strength to the belief that these markers, and thus decisions based on them can be reliably replicated. Three findings are of particular interest: first, ISS has a clear and independent value in combination with either GEP classifiers or FISH markers. GEP classifiers combined with ISS

are the strongest risk classifications found here. By combining the EMC92 gene classifier with ISS, patients are effectively stratified into four risk groups including a distinctive low-risk group of 38% and a high-risk group of 17%. This strong additive strength of ISS to GEP has been recognized before in a previous smaller study.[34] Also ISS was integrated with GEP and other factors, but this risk-score did not take into account correlations between markers, and was generated without using a solid discovery/validation design.[35] In contrast, we have opted for a study design in which part of the data was reserved for validation.

Secondly, our study confirmed that FISH markers can be divided into those consistently associated with shorter OS as opposed to inconsistent markers. Consistent FISH markers included t(4;14), gain1q, del17p and del13q. Combinations of any of these markers with ISS constituted solid prognostic predictors as reported previously, with t(4;14) and del17p currently regarded as the most important high-risk FISH markers.[17] Thirdly, by combining these FISH markers into the previously defined risk classifications HR.FISH.A and HR.FISH.B/ISS, a major improvement of prognostic strength is achieved. Interestingly, patients classified as high-risk according to the HR.FISH.A marker but that actually had favorable survival, were correctly identified as low-risk patients by the EMC92-ISS compound marker.

In addition to validating EMC92-ISS, we have now also validated the HR.FISH.B/ISS risk classification for the first time in independent data by excluding training data from the analyses. Combining FISH and ISS is thus a valid choice for routine clinical practice, including the existing HR-FISH.B/ISS, as proposed by Avet-Loiseau *et al*.[19] Incorporating LDH and bone imaging was outside the scope of this study because these markers were not consistently available.[20]

Combining GEP with ISS may become an attractive option for prognostication. The EMC92-ISS classification is independent from therapy choice: the EMC92 was shown to function in

bortezomib clinical trials as well as in thalidomide and more conventional regimens.[7] In contrast, bortezomib and other novel agents may abrogate the unfavourable impact of some FISH markers on PFS.[29] EMC92-ISS is useful since it can identify both high-risk and low-risk MM. This is an advantage over FISH markers which only seem to identify high risk patients. Moreover, the technical applicability of GEP and its costs are thought to be comparable to FISH.[36]

The agreement between GEP classifiers in terms of pathways is of interest. Although the primary force for classifier discovery is association with survival, the genes within classifiers appear to converge on the cell cycle pathways. Indeed, proliferative capacity, assessed as the plasma cell labeling index or by Ki-67 staining, has long been recognized to be an important prognostic factor.[37,38]

The clinical applicability of stratification into four risk groups will be increasingly relevant in the era of novel treatment modalities being available. First, increased accuracy of prognosis can improve patient counseling.[17] Secondly, and more important, risk-stratification may lead to adaptation of treatment according to risk status. This composite risk marker opens the way to better risk-stratification in clinical trials and explore novel drugs in different risk groups.[39,40] This could effectively be a first step towards a more individual treatment, using patient specific markers as a directional key.

Based on the current study we conclude that the combination of EMC92 with ISS is a strong disease based prognosticator for survival in MM. This risk classification is a good candidate to stratify patients for treatment options in a clinical trial.

Acknowledgments

We would like to acknowledge Mathijs Sanders for his contribution. This work was supported by the Center for Translational Molecular Medicine (BioCHIP), a clinical research grant from

the European Hematology Association, an EMCR Translational Research Grant, a BMBF grant from CLIOMMICS (01ZX1309A), SkylineDx, Janssen Pharmaceuticals and FP7 grant MSCNET (LSHC-Ct-2006-037602).

Authorship Contributions

Contribution: R.K., M.v.D., M.H.v.V. and P.S. designed research; A.B. collected data; R.K. and M.v.D. analyzed and interpreted the data and wrote the manuscript; M.H.v.V. and E.v.B. critically reviewed the paper; B.vd.H. provided the clinical data of the HO65/HD4 and critically reviewed the paper; L.e.J. performed central data management for the HO65/HD4; G.Mulligan provided the CEL files from the APEX dataset and critically reviewed the paper; G.Morgan and W.M.G. provided CEL files and clinical data from the MRC-IX dataset and critically reviewed the paper; H.A-L. provided the clinical data of the IFM-G and critically reviewed the paper; H.G. is principal investigator research of the German part of the HO65/HD4 and critically reviewed the paper; H.M.L organized trial and critically reviewed the paper. P.S. organized trial, is principal investigator of the performed research and HO65/HD4 and critically reviewed the paper.

Disclosure of Conflict of Interest

Conflict-of-interest disclosure: R.K., A.B., B.vd.H., L.e.J. and M.v.D.: no disclosures; M.H.v.V. and E.v.B. are employees of Skyline Dx; W.M.G. received unrestricted educational grants from Novartis, Schering Health Care Ltd, Chugai, Pharmion, Celgene and Ortho Biotech. G. Morgan, HAL, G. Mulligan, H.G., H.M.L.: member of advisory board of pharmaceutical companies; P.S.:, advisory board of Skyline DX, received honoraria and research funding from Janssen-Cilag, Celgene, Onyx and research funding from Millennium.

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):e538-548.
2. Brenner H, Gondos A, Pulte D. Recent major improvement in long-term survival of younger patients with multiple myeloma. *Blood*. 2008;111(5):2521-2526.
3. Bergsagel PL, Chesi M, Nardini E, Brents LA, Kirby SL, Kuehl WM. Promiscuous translocations into immunoglobulin heavy chain switch regions in multiple myeloma. *Proc Natl Acad Sci U S A*. 1996;93(24):13931-13936.
4. Fonseca R, Blood E, Rue M, et al. Clinical and biologic implications of recurrent genomic aberrations in myeloma. *Blood*. 2003;101(11):4569-4575.
5. Zhan F, Huang Y, Colla S, et al. The molecular classification of multiple myeloma. *Blood*. 2006;108(6):2020-2028.
6. Broyl A, Hose D, Lokhorst H, et al. Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. *Blood*. 2010;116(14):2543-2553.
7. Kuiper R, Broyl A, de Knegt Y, et al. A gene expression signature for high-risk multiple myeloma. *Leukemia*. 2012;26(11):2406-2413.
8. Shaughnessy JD, Jr., Zhan F, Burington BE, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007;109(6):2276-2284.
9. Shaughnessy JD, Jr., Qu P, Usmani S, et al. Pharmacogenomics of bortezomib test-dosing identifies hyperexpression of proteasome genes, especially PSMD4, as novel high-risk feature in myeloma treated with Total Therapy 3. *Blood*. 2011;118(13):3512-3524.
10. Decaux O, Lode L, Magrangeas F, et al. Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the Intergroupe Francophone du Myelome. *J Clin Oncol*. 2008;26(29):4798-4805.

11. Hose D, Reme T, Hielscher T, et al. Proliferation is a central independent prognostic factor and target for personalized and risk-adapted treatment in multiple myeloma. *Haematologica*. 2011;96(1):87-95.

12. Reme T, Hose D, Theillet C, Klein B. Modeling risk stratification in human cancer. *Bioinformatics*. 2013;29(9):1149-1157.

13. Dickens NJ, Walker BA, Leone PE, et al. Homozygous deletion mapping in myeloma samples identifies genes and an expression signature relevant to pathogenesis and outcome. *Clin Cancer Res*. 2010;16(6):1856-1864.

14. Greipp PR, Lust JA, O'Fallon WM, Katzmann JA, Witzig TE, Kyle RA. Plasma cell labeling index and beta 2-microglobulin predict survival independent of thymidine kinase and C-reactive protein in multiple myeloma. *Blood*. 1993;81(12):3382-3387.

15. Kyle RA, Gertz MA, Witzig TE, et al. Review of 1027 patients with newly diagnosed multiple myeloma. *Mayo Clin Proc*. 2003;78(1):21-33.

16. Greipp PR, San Miguel J, Durie BG, et al. International staging system for multiple myeloma. *J Clin Oncol*. 2005;23(15):3412-3420.

17. Chng WJ, Dispenzieri A, Chim CS, et al. IMWG consensus on risk stratification in multiple myeloma. *Leukemia*. 2014;28(2):269-277.

18. Neben K, Jauch A, Bertsch U, et al. Combining information regarding chromosomal aberrations t(4;14) and del(17p13) with the International Staging System classification allows stratification of myeloma patients undergoing autologous stem cell transplantation. *Haematologica*. 2010;95(7):1150-1157.

19. Avet-Loiseau H, Durie BG, Cavo M, et al. Combining fluorescent in situ hybridization data with ISS staging improves risk assessment in myeloma: an International Myeloma Working Group collaborative project. *Leukemia*. 2013;27(3):711-717.

20. Moreau P, Cavo M, Sonneveld P, et al. Combination of International Scoring System 3, High Lactate Dehydrogenase, and t(4;14) and/or del(17p) Identifies Patients With Multiple Myeloma (MM) Treated With Front-Line Autologous Stem-Cell Transplantation at High Risk of Early MM Progression-Related Death. *J Clin Oncol*. 2014;32(20):2173-2180.

21. Broyl A, Kuiper R, van Duin M, et al. High cereblon expression is associated with better survival in patients with newly diagnosed multiple myeloma treated with thalidomide maintenance. *Blood*. 2013;121(4):624-627.

22. Morgan GJ, Davies FE, Gregory WM, et al. Long-term follow-up of MRC Myeloma IX trial: Survival outcomes with bisphosphonate and thalidomide treatment. *Clin Cancer Res*. 2013;19(21):6030-6038.

23. Mulligan G, Mitsiades C, Bryant B, et al. Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. *Blood*. 2007;109(8):3177-3188.

24. Nair B, van Rhee F, Shaughnessy JD, Jr., et al. Superior results of Total Therapy 3 (2003-33) in gene expression profiling-defined low-risk multiple myeloma confirmed in subsequent trial 2006-66 with VRD maintenance. *Blood*. 2010;115(21):4168-4173.

25. Ross FM, Ibrahim AH, Vilain-Holmes A, et al. Age has a profound effect on the incidence and significance of chromosome abnormalities in myeloma. *Leukemia*. 2005;19(9):1634-1642.

26. Walker BA, Leone PE, Chiecchio L, et al. A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood*. 2010;116(15):e56-65.

27. Ross FM, Avet-Loiseau H, Ameye G, et al. Report from the European Myeloma Network on interphase FISH in multiple myeloma and related disorders. *Haematologica*. 2012;97(8):1272-1277.

28. Wuilleme S, Robillard N, Lode L, et al. Ploidy, as detected by fluorescence in situ hybridization, defines different subgroups in multiple myeloma. *Leukemia*. 2005;19(2):275-278.

29. Sonneveld P, Schmidt-Wolf IG, van der Holt B, et al. Bortezomib induction and maintenance treatment in patients with newly diagnosed multiple myeloma: results of the randomized phase III HOVON-65/ GMMG-HD4 trial. *J Clin Oncol*. 2012;30(24):2946-2955.

30. Cox DR, Oakes D. Analysis of Survival Data: Taylor & Francis; 1984.

31. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995;310(6973):170.

32. Therneau TM, Grambsch PM. Modeling survival data : extending the Cox model. New York: Springer; 2000.

33. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-29.

34. Waheed S, Shaughnessy JD, van Rhee F, et al. International staging system and metaphase cytogenetic abnormalities in the era of gene expression profiling data in multiple myeloma treated with total therapy 2 and 3 protocols. *Cancer*. 2011;117(5):1001-1009.

35. Meissner T, Seckinger A, Reme T, et al. Gene expression profiling in multiple myeloma--reporting of entities, risk, and targets in clinical routine. *Clin Cancer Res*. 2011;17(23):7240-7247.

36. Hose D, Seckinger A, Jauch A, et al. The role of fluorescence in situ hybridization and gene expression profiling in myeloma risk stratification. *Srp Arh Celok Lek*. 2011;139 Suppl 2:84-89.

37. Steensma DP, Gertz MA, Greipp PR, et al. A high bone marrow plasma cell labeling index in stable plateau-phase multiple myeloma is a marker for early disease progression and death. *Blood*. 2001;97(8):2522-2523.

38. Gastinne T, Leleu X, Duhamel A, et al. Plasma cell growth fraction using Ki-67 antigen expression identifies a subgroup of multiple myeloma patients displaying short survival within the ISS stage I. *Eur J Haematol*. 2007;79(4):297-304.

39. Ludwig H, Sonneveld P, Davies F, et al. European Perspective on Multiple Myeloma Treatment Strategies in 2014. *Oncologist*. 2014.

40. Crawley C, Iacobelli S, Bjorkstrand B, Apperley JF, Niederwieser D, Gahrton G. Reduced-intensity conditioning for myeloma: lower nonrelapse mortality but higher relapse rates compared with myeloablative conditioning. *Blood*. 2007;109(8):3588-3594.

TABLES

Table 1. Distribution of risk markers and treatments per dataset. The numbers of patients per data set are given with in brackets the number or percentage of positive patients according to the markers' risk classification.

	HO65/HD4	MRC-IX		TT2	TT3	APEX	IFM-G	POOLED
		Intensive	non-intensive					
N	827	701	491	351	238	264	1878	4750 **
age [median [IQR]][yrs]	57 (51 - 61)	58 (54 - 63)	74 (70 - 77)	57 (49 - 64)	60 (53 - 66)	61 (54 - 67)	57 (51 - 61)	57 (51 - 62)
Treatment [n]	PAD (413)	CTD (351)	CTDa (257)	VTD (175)	VTD (238)	BOR (188)	VD (740)	BOR (1579) / THAL (783)
Control [n]	VAD (414)	CVAD (350)	MP (234)	VMD (176)	No controls	DEX (76)	VAD (1138)	BOR (1628) / THAL (760)
High-dose alkylator	YES	YES	NO	YES	YES	YES	YES	
EMC92 [n (% positive)]	*	138 (17%)	109 (24%)	345 (19%)	238 (15%)	264 (16%)		1094 (18%)
UAMS17 [n (%positive)]	327 (12%)	138 (9%)	109 (16%)	*	238 (14%)	264 (12%)		1076 (12%)
UAMS70 [n (%positive)]	327 (9%)	138 (7%)	109 (10%)	*	238 (12%)	264 (8%)		1076 (9%)
UAMS80 [n (%positive)]	327 (8%)	138 (8%)	109 (9%)	345 (9%)	*	264 (7%)		1183 (8%)
MRCIX6 [n (%positive)]	327 (5%)	*	*	345 (7%)	238 (5%)	264 (3%)		1174 (5%)
IFM15 [n (%positive)]	327 (25%)	138 (25%)	109 (28%)	345 (24%)	238 (24%)			1157 (25%)
HM19 [n (low /medium /high %)]	327 (43 /49 /8%)	138 (45 /48 /7%)	109 (39 /53 /8%)	345 (50 /47 /8%)	238 (47 /47 /7%)	264 (41 /50 /8%)		1420 (44 /48 /8%)
GPI50 [n (low /medium /high %)]	327 (34 /51 /15 %)	138 (52 /41 /7%)	109 (52 /38 /10%)	345 (63 /31 /7%)	238 (58 /34 /8%)			1159 (51 /39 /10%)
ISS [n (1/2/3 %)]	756 (38 /37 /25%)	636 (25 /39 /36%)	449 (13 /41 /45%)	351 (54 /25 /21%)	208 (50 /28 21%)	202 (34 /33 /33%)	1475 (34 /39 /28%)	4074 (34 /37 /30%)
t(4;14) [n (%positive)]	492 (12%)	619 (12%)	434 (10%)				1635 (14%)	3180 (13%)
t(11;14) [n (%positive)]	437 (16%)	617 (15%)	434 (12%)					1488 (15%)
t(14;16) [n (%positive)]	360 (2%)	612 (3%)	434 (3%)				456 (4%)	1862 (3%)
t(14;20) [n (%positive)]	255 (0%)	612 (2%)	429 (1%)					1296 (1%)
IgH split [n (%positive)]	372 (48%)	609 (44%)	429 (40%)					1410 (44%)
gain1q [n (%positive)]	344 (32%)	531 (37%)	371 (41%)	248 (47%)			891 (37%)	2385 (38%)
del13q [n (%positive)]	686 (41%)	612 (46%)	428 (43%)				1807 (48%)	3533 (46%)
del17p [n (%positive)]	351 (11%)	591 (8%)	423 (9%)				1651 (15%)	3016 (12%)
gain9 [n (%positive)]	454 (57%)	480 (60%)	351 (66%)					1285 (60%)
HR.FISH.A [n (%)]	354 (46%)	535 (48%)	368 (48%)	116 (100%)***			1022 (64%)	2395 (57%)
HR.FISH.B/ISS [n (1 /2 /3 %)]	334 (60 /22 /18%)	*	*				516 (55 /29 /17%) *	850 (57 /26 / 17%)

*, training set for these markers. Only the proportion and number that are not used for building the marker, if any, are shown.
**, intersection of patients with available data between datasets is shown in Figure S1
***, the HR.FISH.A compound risk classification is based on a patient having either del17p, t(4;14) or gain of 1q. If only gain of 1q is known (in TT2 patients), these are the only patients classified with certainty as high-risk. The remaining patients cannot be classified, since the status of t(4;14) and del17p are unknown. If the missing bias is strong enough (see methods), that marker is excluded from the combination analyses.
PAD: bortezomib, doxorubicin, dexamethasone**; VAD:** vincristine, doxorubicin, dexamethasone**; CVAD**: cyclophosphamide, vincristine, doxorubicin, dexamethasone; **CTD**: cyclophosphamide, thalidomide, dexamethasone; **MP**: melphalan, prednisone; **CTDa**: attenuated CTD; **VTD:** bortezomib, thalidomide, dexamethasone**; VMD:** bortezomib, melphalan, dexamethasone**; VD:** vincristine, dexamethasone**; BOR:** bortezomib**; THAL:** thalidomide**.**

Table 2. Proportion of surviving patients at multiple time points per EMC92-ISS risk group in a Kaplan Meier analysis on the validation data (from top to bottom: 6, 12, 24 and 72 months respectively). In the left column patient groups are pooled (n = 328). Subsequent columns show percentages for newly diagnosed patients younger than 65 years (n = 174), newly diagnosed older than 65 years (n = 90) and relapsed patients (n = 64) respectively. For the last category the 72 months' time point is not available.

6 months	**Pooled**	**<65**	**≥65**	**Relapse**
Low-risk	98 %	97 %	96 %	95 %
Intermediate low-risk	96 %	95 %	95 %	85 %
Intermediate high-risk	86 %	93 %	77 %	79 %
High-risk	84 %	88 %	75 %	57 %
Total survival	**92 %**	**94 %**	**87 %**	**83 %**

12 months	**Pooled**	**<65**	**≥65**	**Relapse**
Low-risk	97 %	97 %	96 %	89 %
Intermediate low-risk	87 %	93 %	91 %	54 %
Intermediate high-risk	74 %	93 %	73 %	42 %
High-risk	67 %	72 %	56 %	57 %
Total survival	**84 %**	**91 %**	**81 %**	**60 %**

24 months	**Pooled**	**<65**	**≥65**	**Relapse**
Low-risk	92 %	97 %	92 %	55 %
Intermediate low-risk	76 %	88 %	73 %	23 %
Intermediate high-risk	57 %	77 %	58 %	24 %
High-risk	46 %	56 %	31 %	0 %
Total survival	**72 %**	**84 %**	**67 %**	**30 %**

72 months	**Pooled**	**<65**	**≥65**	**Relapse**
Low-risk	77 %	86 %	69 %	-
Intermediate low-risk	43 %	59 %	32 %	-
Intermediate high-risk	27 %	39 %	28 %	-
High-risk	22 %	33 %	0 %	-
Total survival	**48 %**	**62 %**	**36 %**	

Table 3. Distribution of markers in each of the four EMC92-ISS based risk groups. Shown are the numbers in the data for which the EMC92-ISS risk classification could be determined**.** n, number of patients in the EMC92-ISS based risk group for which the specified marker was available. Positive, the percentage of patients positive for the specified marker; HR, the percentage of patients indicated as high-risk according to the specified marker. For the classifications based on del13q, 1q gain and HR.FISH.A, a clear correlation was found to the EMC92-ISS classifications. For instance, 93% of EMC92-ISS high-risk patients are positive for HR.FISH.A compared to 44% - 55% of the intermediates and 75% of low group.

	EMC92		**ISS**				**del17p**		**del13q**		**1q gain**		**HR.FISH.A**	
EMC92-ISS	HR	n	1	2	3	n	Positive	n	Positive	n	Positive	n	HR	n
Low	0%	365	100%	0%	0%	365	8%	39	44%	39	34%	154	75%	76
Intermediate-Low	0%	231	0%	100%	0%	231	5%	60	37%	60	34%	92	44%	70
Intermediate-High	0%	211	0%	0%	100%	211	8%	66	44%	66	41%	101	55%	84
High	100%	166	30%	32%	39%	166	16%	38	74%	39	76%	90	93%	76

Figure Legends

Figure 1. Flowchart of analyses. The analyses are organized as follows: 1) confirmation of existing risk markers, 2) systematically finding novel risk markers with improved prognostic strength by combining existing risk markers and 3) validating them; 4) ranking of confirmed existing- and validated novel risk markers. See Figure S2A-C for more details.

Figure 2. Risk markers in relation to overall survival. Both existing markers and validated novel combinations are shown. For novel combinations, the results shown represent the validation. For confirmation of existing markers no discovery/validation split is required and results shown are based on all available data. In the left panel, existing markers and novel combinations (denoted by an asterisk) are listed. For each marker, the number of risk groups (n. groups) and number of available patients is given (n. patients). Markers are sorted by the number of risk groups. In the center panel, the hazard ratios are shown (open circle), with Bonferroni adjusted 95% confidence intervals (indicated by two lines and closed circles). For coherent notation, hazard ratios are expressed relative to the lowest-risk group. Every additional risk group results in an extra hazard ratio. For instance, for the novel combination EMC92 – ISS, 4 risk groups result in 3 hazard ratios, as indicated in the text and Table S2A (intermediate-low risk relative to low risk: hazard ratio (HR) 2.6 (confidence interval (CI): 1.6 - 4.5), intermediate-high risk relative to low risk: HR: 3.2 (CI: 1.9 - 5.4) and high risk relative to low risk: HR 6.9 (4.1 - 12)). In the right panel, a plus sign indicates whether a data set could be used for the analysis of a specific marker or combination (for details of available data, see Table 1 and Figure S1). For the EMC92-ISS combination, the following datasets could be used: APEX, MRC-IX, TT2 and TT3.

Figure 3. Ranking of confirmed existing risk markers and validated novel risk markers, in relation to overall survival on the validation data. The markers are vertically ordered by rank score, which reflects the observed proportion of risk markers with a better performance. Each box shows the interquartile range of the rank score per marker.

Figure 4. Survival analysis of EMC92-ISS, FISH and ISS. Kaplan-Meier plots and Cox regression model data are given. Kaplan-Meier plots are not stratified; Cox regression results are stratified, i.e. corrected for differences in survival in different cohorts. A) EMC92-ISS in the discovery set; B) EMC92-ISS in the validation set; C) EMC92-ISS in all data; D) ISS in all data; E) HR.FISH.A in all data; F) HR.FISH.B/ISS in all data. In order of increasing risk: low-risk (blue); intermediate-low (purple); intermediate-high (orange); high (red); SR = standard-risk; HR = high-risk. Below the Kaplan-Meier curves, results of the stratified Cox model are found. Hazard = hazard ratio relative to the lowest risk group; 95% CI = 95% confidence interval; P = *p*-value relative to the lowest risk group; % positive = percentage of patients within the specified risk group. The bottom line shows the result of the likelihood ratio goodness of fit test.

Figure 1

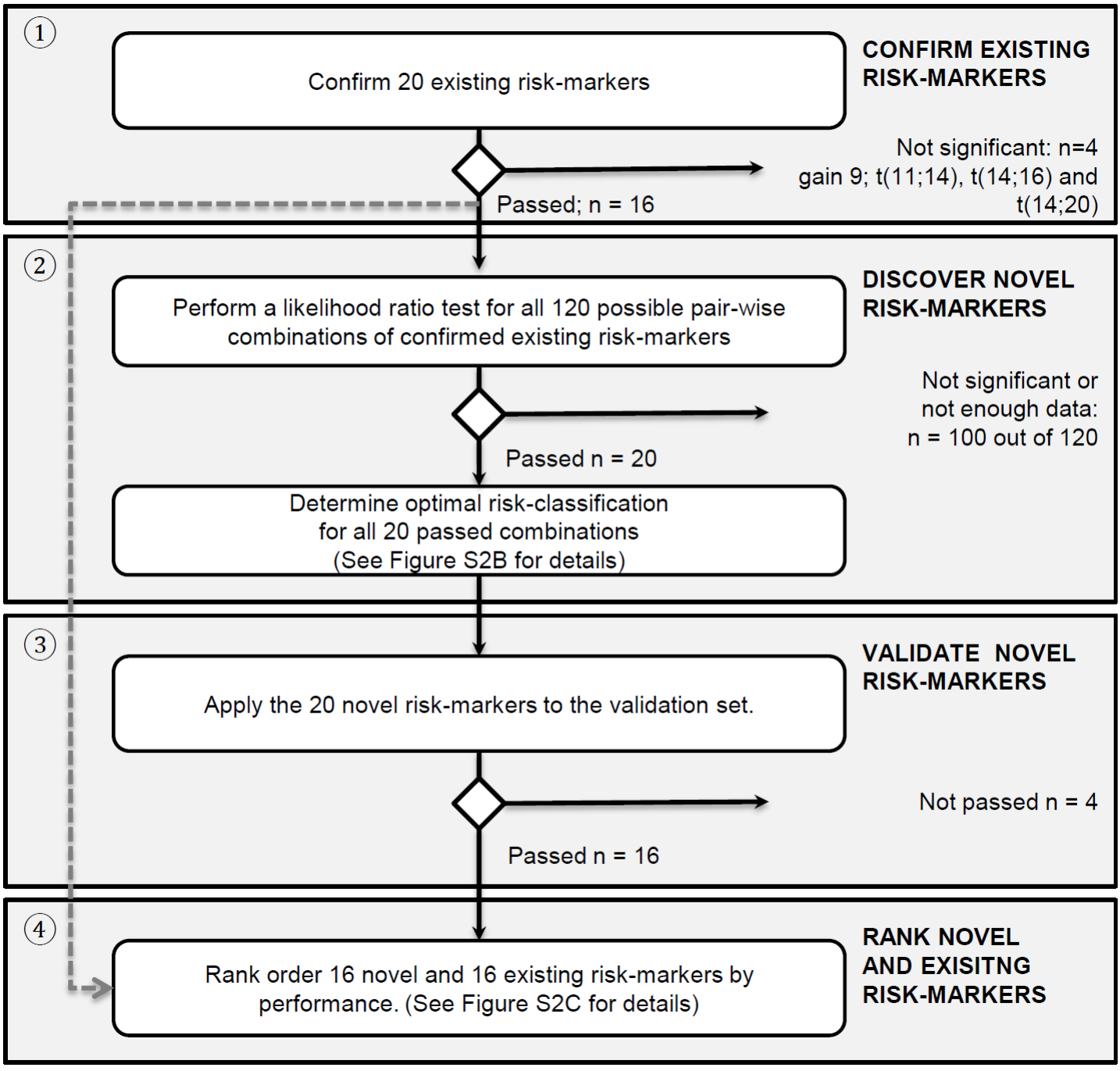

1
Confirm 20 existing risk-markers
CONFIRM EXISTING RISK-MARKERS
Not significant: n=4 gain 9; t(11;14), t(14;16) and t(14;20)
Passed; n = 16
2
Perform a likelihood ratio test for all 120 possible pair-wise combinations of confirmed existing risk-markers
DISCOVER NOVEL RISK-MARKERS
Not significant or not enough data: n = 100 out of 120
Passed n = 20
Determine optimal risk-classification for all 20 passed combinations (See Figure S2B for details)
3
Apply the 20 novel risk-markers to the validation set.
VALIDATE NOVEL RISK-MARKERS
Not passed n = 4
Passed n = 16
4
Rank order 16 novel and 16 existing risk-markers by performance. (See Figure S2C for details)
RANK NOVEL AND EXISITNG RISK-MARKERS

Figure 2

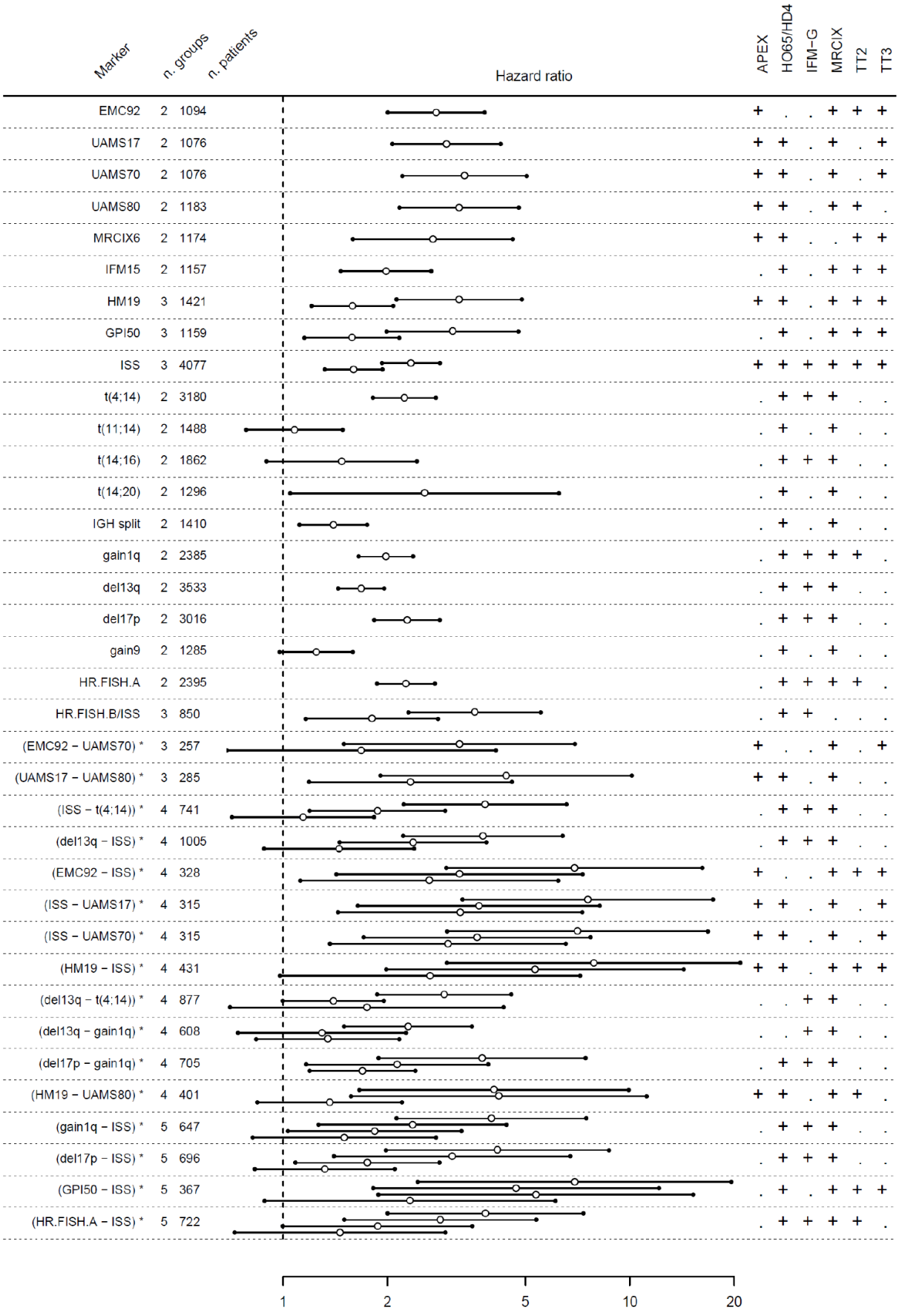

Marker
n. groups
n. patients
Hazard ratio
APEX
HO65/HD4
IFM−G
MRCIX
TT2
TT3
EMC92 2 1094 + . . + + +
UAMS17 2 1076 + + . + . +
UAMS70 2 1076 + + . + . +
UAMS80 2 1183 + + . + + .
MRCIX6 2 1174 + + . . + +
IFM15 2 1157 . + . + + +
HM19 3 1421 + + . + + +
GPI50 3 1159 . + . + + +
ISS 3 4077 + + + + + +
t(4;14) 2 3180 . + + + . .
t(11;14) 2 1488 . + . + . .
t(14;16) 2 1862 . + + + . .
t(14;20) 2 1296 . + . + . .
IGH split 2 1410 . + . + . .
gain1q 2 2385 . + + + + .
del13q 2 3533 . + + + . .
del17p 2 3016 . + + + . .
gain9 2 1285 . + . + . .
HR.FISH.A 2 2395 . + + + + .
HR.FISH.B/ISS 3 850 . + + . . .
(EMC92 − UAMS70) * 3 257 + . . + . +
(UAMS17 − UAMS80) * 3 285 + + . + . .
(ISS − t(4;14)) * 4 741 . + + + . .
(del13q − ISS) * 4 1005 . + + + . .
(EMC92 − ISS) * 4 328 + . . + + +
(ISS − UAMS17) * 4 315 + + . + . +
(ISS − UAMS70) * 4 315 + + . + . +
(HM19 − ISS) * 4 431 + + . + + +
(del13q − t(4;14)) * 4 877 . . + + . .
(del13q − gain1q) * 4 608 . . + + . .
(del17p − gain1q) * 4 705 . + + + . .
(HM19 − UAMS80) * 4 401 + + . + + .
(gain1q − ISS) * 5 647 . + + + . .
(del17p − ISS) * 5 696 . + + + . .
(GPI50 − ISS) * 5 367 . + . + + +
(HR.FISH.A − ISS) * 5 722 . + + + + .
1
2
5
10
20

Figure 3

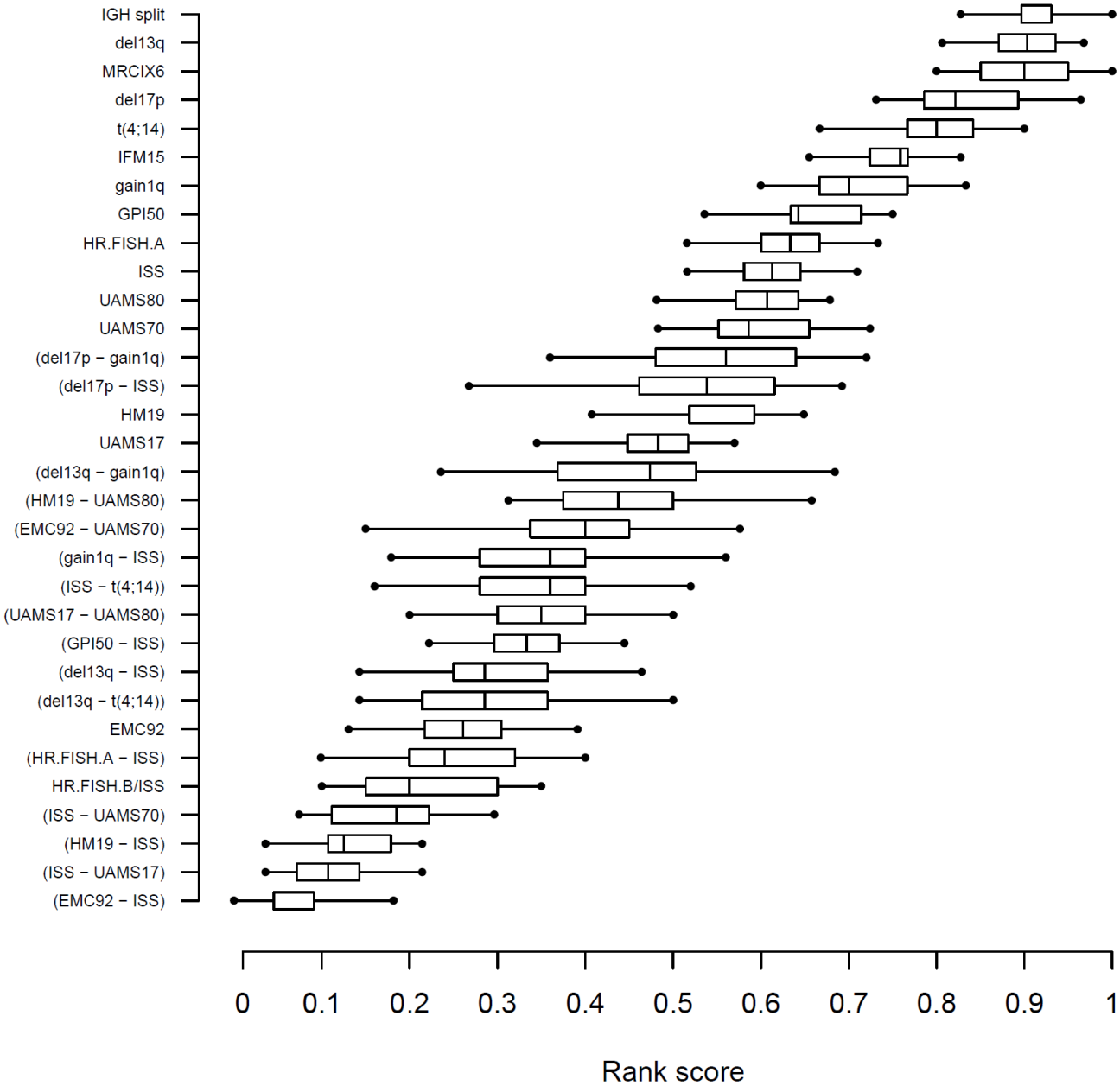
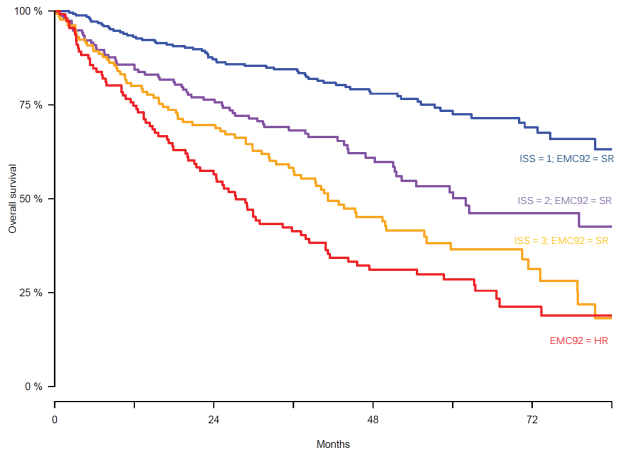

IGH split
del13q
MRCIX6
del17p
t(4;14)
IFM15
gain1q
GPI50
HR.FISH.A
ISS
UAMS80
UAMS70
(del17p − gain1q)
(del17p − ISS)
HM19
UAMS17
(del13q − gain1q)
(HM19 − UAMS80)
(EMC92 − UAMS70)
(gain1q − ISS)
(ISS − t(4;14))
(UAMS17 − UAMS80)
(GPI50 − ISS)
(del13q − ISS)
(del13q − t(4;14))
EMC92
(HR.FISH.A − ISS)
HR.FISH.B/ISS
(ISS − UAMS70)
(HM19 − ISS)
(ISS − UAMS17)
(EMC92 − ISS)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Rank score

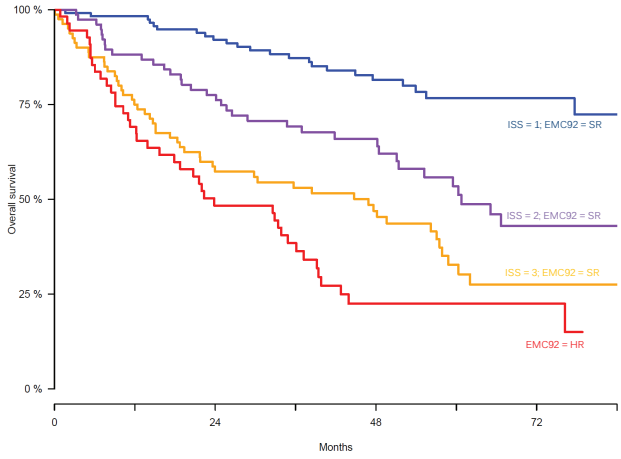
Figure 4A



	Hazard ratio	95%CI	p	Proportion
Low	1			39%
Intermediate−low	1.6	(1.1 − 2.2)	0.016	24%
Intermediate−high	2.3	(1.6 − 3.2)	3.9e−06	20%
High	4.5	(3.2 − 6.3)	. ≤ 1e−15	17%

Likelihood ratio test ≤ 1e−15, n : 645, number of events : 286

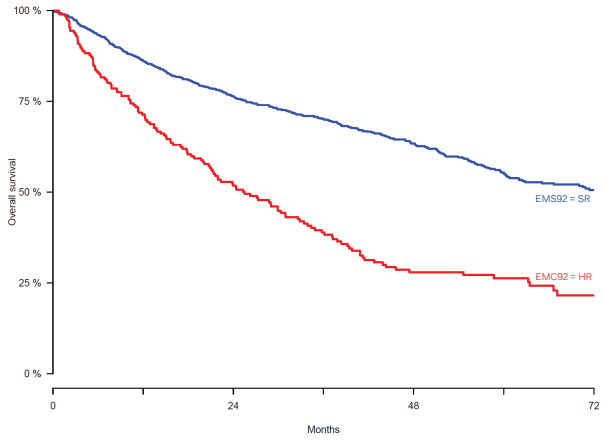
Figure 4B



	Hazard ratio	95%CI	p	Proportion
Low	1			35%
Intermediate−low	2.6	(1.6 − 4.5)	0.00033	23%
Intermediate−high	3.2	(1.9 − 5.4)	5.9e−06	24%
High	6.9	(4.1 − 12)	5.9e−13	17%

Likelihood ratio test 2.70e−12; n: 328; number of events: 149

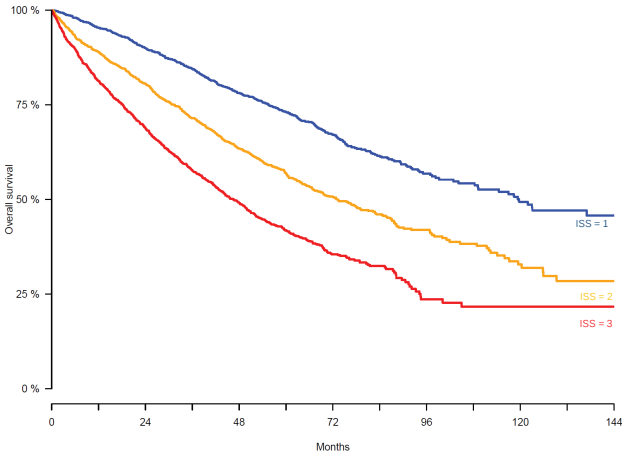
Figure 4C



	Hazard ratio	95%CI	p	Proportion
Standard	1			82%
High	2.8	(2.3 − 3.4)	. ≤ 1e−15	18%

Likelihood ratio test ≤ 1e−15, n : 1094, number of events : 504

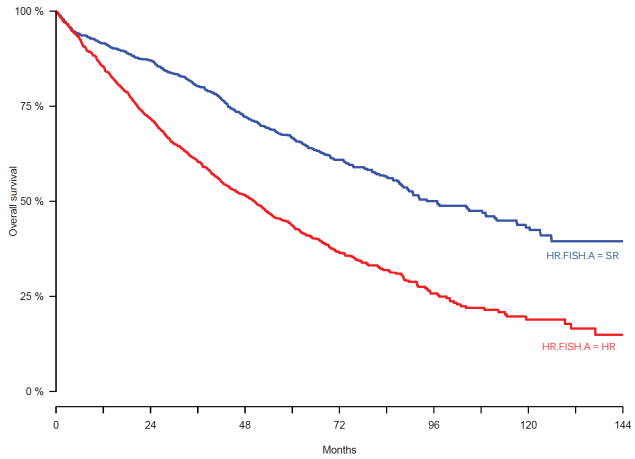
Figure 4D



	Hazard ratio	95%CI	p	Proportion
Low	1			34%
Intermediate	1.6	(1.4 − 1.8)	1e−14	37%
High	2.3	(2.1 − 2.6)	. ≤ 1e−15	30%

Likelihood ratio test ≤ 1e−15, n : 4077, number of events : 1925

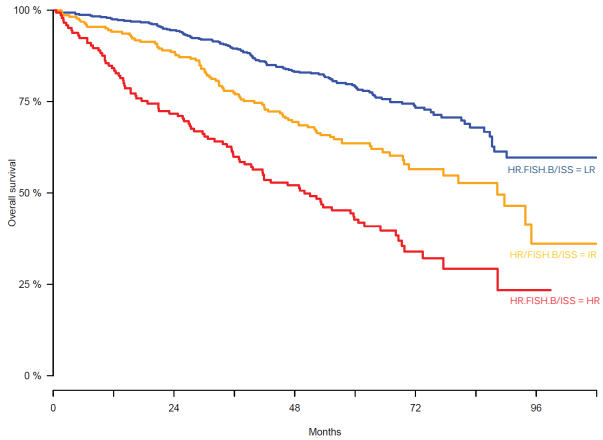
Figure 4E



	Hazard ratio	95%CI	p	Proportion
Standard	1			43%
High	2.3	(2 − 2.5)	. ≤ 1e−15	57%

Likelihood ratio test ≤ 1e−15, n : 2395, number of events : 1309

Figure 4F



	Hazard ratio	95%CI	p	Proportion
Low	1			57%
Intermediate	1.8	(1.4 − 2.4)	2.1e−05	26%
High	3.6	(2.7 − 4.7)	. ≤ 1e−15	17%

Likelihood ratio test ≤ 1e−15, n : 850, number of events : 309