**Supplemental Data**

**Title: Impact of genes highly correlated with *MMSET* myeloma on survival in Non-*MMSET* myeloma patients**

S. Peter Wu1, Ruth M. Pfeiffer2, Inhye Ahn1, Sham Mailankody1, Pieter Sonneveld3, Mark van Duin3, Nikhil Munshi4, Brian Walker5, Gareth Morgan6, Ola Landgren7

1 Multiple Myeloma Section, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

2 Department of Cancer Epidemiology and Genetics, Biostatistics Branch, National Cancer Institute, National Institutes of Health, Rockville, MD, USA

3 Department of Hematology, Erasmus University Medical Center, Rotterdam, Netherlands

4 Lebow Institute of Myeloma Therapeutics and Jerome Lipper Multiple Myeloma Center, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA

5 Section of Haemato-Oncology, The Institute of Cancer Research, London, UK

6 Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences, Little Rock, AR, USA

7 Myeloma Service, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

**Table: 3**

**Appendix: 1**

**Supplemental Table 1. Comparison of hazard ratios of patients in the quartile four of Non-*MMSET* patients and *MMSET* patients by FISH phenotype, the main analysis, and three other sensitivity analyses**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Category** | **HR (95% CI)** | **p-value** |
| Main Analysis | Q4 Risk | 2.0 (1.5-2.8) | <0.001 |
| MMSET | 2.3 (1.6-3.3) | <0.001 |
| No Proteasome | Q4 Risk | 2.1 (1.4-3.1) | <0.001 |
| MMSET | 2.9 (1.9-4.3) | <0.001 |
| 12 Treatment Groups | Q4 Risk | 1.8 (1.3-2.6) | <0.001 |
| MMSET | 2.3 (1.6-3.3) | <0.001 |
| Overall Survival | Q4 Risk | 2.0 (1.4-2.8) | <0.001 |
| MMSET | 2.6 (1.8-3.7) | <0.001 |
| FISH Phenotype | Q4 Risk | 4.3 (3.3-5.1) | <0.001 |
| MMSET | 3.6 (2.5-4.7) | 0.005 |

Caption: Supplemental Table 1 contains a comparison of hazard ratios of Non-*MMSET* patients at the highest quartile (Q4 Risk) and *MMSET* patients.

**Supplemental Table 2. All gene signatures from all sensitivity analyses**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Main Analysis | | No Proteasome | | 12 Treatments | | Overall Survival | | FISH | |
| CCND1 | CCND1 | CCND1 | 208712\_at | CCND1 | 208712\_at | CCND1 | 208712\_at | --- | 227290\_at |
| AZGP1 | AZGP1 | AZGP1 | 209309\_at | AZGP1 | 209309\_at | AZGP1 | 209309\_at | MAGED4 | 221261\_x\_at |
| MAGED4 | MAGED4 | SGCB | 205120\_s\_at | SGCB | 205120\_s\_at | SGCB | 205120\_s\_at | MITF | 207233\_s\_at |
| SGCB | SGCB | PTP4A3 | 206574\_s\_at | PTP4A3 | 206574\_s\_at | MYBL1 | 213906\_at | RMST | 222325\_at |
| PTP4A3 | PTP4A3 | RNF130 | 217865\_at | MYBL1 | 213906\_at | IGFBP7 | 201163\_s\_at | MITF | 226066\_at |
| MYBL1 | MYBL1 | PPIC | 204518\_s\_at | RNF130 | 217865\_at | RNF130 | 217865\_at | DSG2 | 1553105\_s\_at |
| RNF130 | RNF130 | PXDN | 212012\_at | PPIC | 204518\_s\_at | PPIC | 204518\_s\_at | CLIC6 | 227742\_at |
| MPPE1 | MPPE1 | C9ORF122 | 1557014\_a\_at | MPPE1 | 213924\_at | MPPE1 | 213924\_at | --- | 229874\_x\_at |
| ROBO1 | ROBO1 |  |  | PXDN | 212012\_at | PXDN | 212012\_at | FZD8 | 227405\_s\_at |
| PXDN | PXDN |  |  | EDNRB | 204273\_at | KLF4 | 220266\_s\_at | TCEAL2 | 211276\_at |
|  |  |  |  | GOLPH2 | 217771\_at | CXADR | 226374\_at | PBX1 | 212148\_at |
|  |  |  |  | C9ORF122 | 1557014\_a\_at | C9ORF122 | 1557014\_a\_at | C10ORF38 | 212771\_at |
|  |  |  |  |  |  |  |  | --- | 213484\_at |

Caption: Supplemental Table 2 contains a list of gene signatures from sensitivity analyses.

**Supplemental Table 3. All 71 genes with 2-fold or greater change in log-expression in *MMSET* phenotype vs Non-*MMSET* phenotype, passing Bonferroni**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **gene** | **probes** | **lin reg coef** | **p-value** | **annotation** |
| CCND2 | 200951\_s\_at | 3.7 | 5.20E-36 | cyclin D2 |
| CCND2 | 200953\_s\_at | 3.7 | 6.00E-36 | cyclin D2 |
| KLF4 | 221841\_s\_at | 3.7 | 3.90E-35 | Kruppel-like factor 4 (gut) |
| CCND1 | 208712\_at | -3.6 | 1.70E-33 | cyclin D1 |
| UCHL1 | 201387\_s\_at | 3.3 | 2.50E-28 | ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase) |
| AZGP1 | 209309\_at | 3.1 | 6.40E-25 | alpha-2-glycoprotein 1 zinc |
| MAGED4 | 223313\_s\_at | 2.9 | 5.00E-23 | melanoma antigen family D4 |
| KIAA0523 | 213155\_at | 2.8 | 1.60E-21 | KIAA0523 protein |
| C18ORF4 | 232235\_at | 2.8 | 1.80E-21 | chromosome 18 open reading frame 4 |
| GNAI1 | 227692\_at | 2.8 | 1.70E-20 | guanine nucleotide binding protein (G protein) alpha inhibiting activity polypeptide 1 |
| SH3BP4 | 222258\_s\_at | 2.7 | 6.20E-20 | SH3-domain binding protein 4 |
| PKP2 | 207717\_s\_at | 2.7 | 9.30E-20 | plakophilin 2 |
| NOL4 | 206045\_s\_at | -2.6 | 4.70E-18 | nucleolar protein 4 |
| TMEM47 | 209656\_s\_at | 2.6 | 7.50E-18 | transmembrane protein 47 |
| MAGED4 | 221261\_x\_at | 2.6 | 8.10E-18 | melanoma antigen family D4 |
| ETV1 | 221911\_at | 2.6 | 1.10E-17 | ets variant gene 1 |
| APP | 200602\_at | 2.6 | 1.20E-17 | amyloid beta (A4) precursor protein (peptidase nexin-II Alzheimer disease) |
| C10ORF30 | 227341\_at | 2.5 | 2.10E-17 | Chromosome 10 open reading frame 30 |
| BAIAP2L1 | 227372\_s\_at | 2.5 | 2.90E-17 | BAI1-associated protein 2-like 1 |
| SGCB | 205120\_s\_at | 2.5 | 3.90E-17 | sarcoglycan beta (43kDa dystrophin-associated glycoprotein) |
| PTP4A3 | 206574\_s\_at | 2.5 | 4.70E-17 | protein tyrosine phosphatase type IVA member 3 |
| MYBL1 | 213906\_at | 2.5 | 5.30E-17 | v-myb myeloblastosis viral oncogene homolog (avian)-like 1 |
| IGFBP7 | 201163\_s\_at | 2.5 | 8.10E-17 | insulin-like growth factor binding protein 7 |
| DYNLRB2 | 238116\_at | 2.5 | 1.10E-16 | dynein light chain roadblock-type 2 |
| C10ORF38 | 212771\_at | 2.5 | 1.40E-16 | chromosome 10 open reading frame 38 |
| RNF130 | 217865\_at | 2.5 | 1.40E-16 | ring finger protein 130 |
| SOX4 | 201417\_at | 2.4 | 2.70E-16 | SRY (sex determining region Y)-box 4 |
| CCND1 | 208711\_s\_at | -2.4 | 6.00E-16 | cyclin D1 |
| MPPED2 | 205413\_at | 2.4 | 6.90E-16 | metallophosphoesterase domain containing 2 |
| AKR1C3 | 209160\_at | 2.4 | 7.00E-16 | aldo-keto reductase family 1 member C3 |
| GNAI1 | 209576\_at | 2.4 | 8.70E-16 | guanine nucleotide binding protein (G protein) alpha inhibiting activity polypeptide 1 |
| PPIC | 204518\_s\_at | 2.4 | 1.00E-15 | peptidylprolyl isomerase C (cyclophilin C) |
| SEP9 | 208657\_s\_at | -2.4 | 1.20E-15 | septin 9 |
| RP11-321G1.1 | 238067\_at | 2.4 | 1.70E-15 | FLJ20298 protein |
| CDC42BPA | 214464\_at | 2.4 | 2.20E-15 | CDC42 binding protein kinase alpha (DMPK-like) |
| CTHRC1 | 225681\_at | 2.3 | 7.80E-15 | collagen triple helix repeat containing 1 |
| FSTL5 | 232010\_at | -2.3 | 7.90E-15 | follistatin-like 5 |
| MPPE1 | 213924\_at | 2.3 | 2.10E-14 | Metallophosphoesterase 1 |
| CDH2 | 203440\_at | 2.3 | 2.30E-14 | cadherin 2 type 1 N-cadherin (neuronal) |
| ABP1 | 203559\_s\_at | 2.3 | 3.30E-14 | amiloride binding protein 1 (amine oxidase (copper-containing)) |
| MAL2 | 224650\_at | 2.3 | 3.40E-14 | mal T-cell differentiation protein 2 |
| CXCR7 | 212977\_at | 2.3 | 4.00E-14 | chemokine (C-X-C motif) receptor 7 |
| CRIM1 | 202551\_s\_at | 2.3 | 4.60E-14 | cysteine rich transmembrane BMP regulator 1 (chordin-like) |
| LOC401074 | 1559827\_at | 2.3 | 4.70E-14 | hypothetical LOC401074 |
| --- | 225710\_at | 2.2 | 7.90E-14 | CDNA FLJ34013 fis clone FCBBF2002111 |
| LRIG1 | 211596\_s\_at | 2.2 | 8.40E-14 | leucine-rich repeats and immunoglobulin-like domains 1 |
| ROBO1 | 213194\_at | 2.2 | 8.60E-14 | roundabout axon guidance receptor homolog 1 (Drosophila) |
| PXDN | 212012\_at | 2.2 | 1.30E-13 | peroxidasin homolog (Drosophila) |
| CSDA | 201161\_s\_at | 2.2 | 1.60E-13 | cold shock domain protein A |
| WWC2 | 218775\_s\_at | 2.2 | 1.80E-13 | WW C2 and coiled-coil domain containing 2 |
| MAP1B | 226084\_at | 2.2 | 3.40E-13 | microtubule-associated protein 1B |
| C20ORF103 | 219463\_at | -2.2 | 3.50E-13 | chromosome 20 open reading frame 103 |
| SELL | 204563\_at | 2.2 | 4.10E-13 | selectin L (lymphocyte adhesion molecule 1) |
| MPPE1 | 214071\_at | 2.2 | 4.60E-13 | Metallophosphoesterase 1 |
| SEP9 | 41220\_at | -2.2 | 5.20E-13 | septin 9 |
| ELOVL4 | 219532\_at | 2.2 | 5.60E-13 | elongation of very long chain fatty acids (FEN1/Elo2 SUR4/Elo3 yeast)-like 4 |
| EDNRB | 204273\_at | -2.1 | 7.20E-13 | endothelin receptor type B |
| NOL4 | 238605\_at | -2.1 | 1.40E-12 | nucleolar protein 4 |
| SLITRK5 | 214930\_at | 2.1 | 1.50E-12 | SLIT and NTRK-like family member 5 |
| GOLPH2 | 217771\_at | 2.1 | 2.00E-12 | golgi phosphoprotein 2 |
| SNX7 | 205573\_s\_at | 2.1 | 2.20E-12 | sorting nexin 7 |
| KLF4 | 220266\_s\_at | 2.1 | 2.20E-12 | Kruppel-like factor 4 (gut) |
| RBM35A | 225846\_at | 2.1 | 2.60E-12 | RNA binding motif protein 35A |
| GABRB2 | 242344\_at | 2.1 | 3.60E-12 | gamma-aminobutyric acid (GABA) A receptor, beta 2 |
| MYADM | 225673\_at | 2.1 | 4.10E-12 | myeloid-associated differentiation marker |
| CD99 | 201028\_s\_at | 2.1 | 5.60E-12 | CD99 molecule |
| PFKP | 201037\_at | 2 | 1.00E-11 | phosphofructokinase platelet |
| CXADR | 226374\_at | 2 | 1.00E-11 | Coxsackie virus and adenovirus receptor |
| AHR | 202820\_at | 2 | 1.60E-11 | aryl hydrocarbon receptor |
| C9ORF122 | 1557014\_a\_at | 2 | 1.80E-11 | chromosome 9 open reading frame 122 |
| GNAQ | 224862\_at | 2 | 1.90E-11 | Guanine nucleotide binding protein (G protein) q polypeptide |

Caption: Supplemental Table 3 contains a list of 71 genes differentially expressed in *MMSET* patients vs. Non-*MMSET* patients.

**APPENDIX 1. Statistical analysis: gene probes associated with survival in Non-*MMSET* patients**

*Treatment classification*

To accommodate treatment, we combined DT-PACE and D-PACE induction arms of TT2 into one treatment category as there was no five year survival difference between the two. 1 The four induction regimens of the Myeloma IX study (CTD, CVAD, CTD(a)ttenuated, Melphalan) were not further stratified by thalidomide maintenance therapy, because there was no 5-year overall survival difference with maintenance therapy (i.e., no 5-year OS difference between CTD with thalidomide maintenance vs. CTD without thalidomide maintenance, CVAD with thalidomide maintenance vs. CVAD without thalidomide maintenance).2 This resulted in four treatment groups: CVAD, CTD, CTDa, and Melphalan induction regimens that were used as adjustment variables in various models.

*Score computation and validation*

The risk score was calculated based on results from the adjusted Cox regression models fit to 5-year survival outcomes by multiplying probe expression values by the coefficients from the final Cox model, and summing these products over all probes identified by the model selection procedure,

*­* 

where *pki* is the expression value for patient *k for* probe *i*, and *c*i is the Cox model regression coefficient (the log-hazard ratio) corresponding to that probe in the final model. To assess if the risk score (*Sk)* for each patient was associated with survival in an unbiased manner, we conducted a 5-fold cross validation.3 This avoids using the same data to develop the signature and assess its performance which would lead to an overly optimistic assessment of the predictive performance of the score. Briefly, the original data set was divided into five equal parts, with equal numbers of patients from each study in each part. Four of the five parts were then used to develop a gene risk signature following the above procedure (the training set). We then computed the risk score for each patient in the remaining fifth (the test set) and computed the association of the risk score (in quartiles) with survival using Cox regression models. This was repeated five times with each piece serving as the test set once. The risk scores from the five test sets were median-centered and combined to form an independently scored measure of risk

Caption: Appendix 1 contains detailed methods of statistical analysis used to determine gene probes associated with survival in Non-*MMSET* patients.