Patient Similarity Network of Newly Diagnosed Multiple Myeloma Identifies Patient Sub-groups with Distinct Genetic Features and Clinical Implications

Sherry Bhalla¹, David T. Melnekoff², ³, Jonathan Keats⁴, Kenan Onel², ³, ⁵, ⁶, ⁷, Jeffrey R. Sawyer⁸, Deepu Madduri¹, ⁵, Joshua Richter¹, ⁵, Shambavi Richard¹, ⁵, Ajai Chari¹, ⁵, Hearn Jay Cho¹, ⁵, Joel T. Dudley⁹, Sundar Jagannath¹, ⁵, Alessandro Laganà², ³, *, Samir Parekh¹, ⁵, ¹⁰

¹Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ²Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ³Institute for Data Science and Genomic Technologies, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ⁴Translational Genomics Research Institute, Phoenix, AZ, USA; ⁵Department of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ⁶Department of Pediatric Hematology and Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ⁷Department of Pathology, Molecular and Cell Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; ⁸Myeloma Center, University of Arkansas for Medical Sciences, Little Rock, AR, USA; ⁹Tempus Labs, Inc., Chicago, IL, USA; ¹⁰Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

*Correspondence:
Dr Alessandro Laganà, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. E-mail: alessandro.lagana@mssm.edu.
Dr Samir Parekh, Division of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai, Tisch Cancer Institute, New York, NY, USA. E-mail: samir.parekh@mssm.edu.

ABSTRACT

The remarkable genetic heterogeneity of Multiple Myeloma poses a significant challenge for proper prognostication and clinical management of patients. Here, we introduce MM-PSN, the first multi-omics Patient Similarity Network of Myeloma. MM-PSN enabled accurate dissection of the genetic and molecular landscape of the disease and determined twelve distinct sub-groups defined by five data types generated from genomic and transcriptomic patient profiling of 655 patients. MM-PSN revealed that 1q gain is the most important single lesion conferring high risk of relapse and that it can improve on the current International Staging Systems (ISS and R-ISS). Several sub-groups uncovered novel associations between the gain of 1q and other adverse secondary lesions, thereby identifying the chromosomal hallmarks of sub-clonal heterogeneity and tumor progression in MM. We also determined gene vulnerabilities and potential therapeutic options for each sub-group and validated our prognostic model in an independent dataset.
INTRODUCTION

Multiple Myeloma (MM) is a mostly incurable malignancy of bone marrow terminally differentiated plasma cells, affecting more than 30,000 patients each year in the United States, with a median survival of approximately 6 years\(^1,2\). While the majority of patients initially respond to standard of care treatment, most relapse and become refractory to treatment as they undergo multiple lines of therapy. In particular, about 15% of patients fall in the high-risk category and typically relapse within two years from diagnosis\(^3,4\). MM is characterized by remarkable clinical and genomic heterogeneity\(^5\). Recent studies based on next-generation sequencing have revealed complex patterns of primary and secondary genetic alterations across patients\(^6,7\), and novel precision medicine approaches, where treatment is guided by the genomic profile of the individual patient, are being tested in trials\(^8,9\). Accurate classification of MM patients into biologically homogeneous classes is thus essential for diagnosis, prognosis and clinical management.

Several classifications of MM based on gene expression have been proposed in the past two decades. The TC classification included 8 groups characterized by different chromosomal translocations and the up-regulation of the Cyclin D genes CCND1 and CCND3\(^10\). The UAMS (University of Arkansas for Medical Sciences) classification, based on unsupervised clustering of gene expression data, proposed 7 clusters in part overlapping the TC classes and enriched for clinically relevant features and differential response to therapy\(^11\). A further refinement was proposed by the HOVON study group, which consisted of 6 of the UAMS classes and 4 novel classes enriched for activation of specific genes such as NFkB and PRL3, and a myeloid signature\(^12\).

Our recent network model of newly diagnosed MM based on gene co-expression, MMNet, revealed a clear molecular separation between patients with Ig translocations and hyperdiploidy and identified three novel subtypes characterized by cytokine signaling (CK), immune signatures (IMM) and MYC translocations (MYC)\(^13\). Another recent study investigated novel MM subtypes based on a targeted DNA panel\(^14\). The analysis revealed a large cluster comprising most HD and IgH translocated patients, and two smaller clusters, one enriched for IgH translocations and high number of copy number alterations (CNA), and one mostly composed of HD patients, with the fewest CNAs and mutations. The different results obtained from different data types, e.g. DNA vs RNA, suggest that a more holistic approach including different omics might further improve patient classification and reveal biologically and clinically informative subtypes of the disease.
Recently, Patient Similarity Networks (PSN) have emerged as a powerful tool to capture and structure the complexity and diversity of clinical, genetic and molecular information across a patient population\textsuperscript{15}. In a PSN, patients are represented as nodes, much like in a social network, and connected with one another based on how similar their genomic and transcriptomic profiles are. The network structure enables effective identification of communities of highly similar patients, allowing a more comprehensive classification than other approaches based on a single measurement. PSNs have been successfully employed to dissect the genomic and molecular complexity of several cancers, including medulloblastoma, glioblastoma multiforme, pancreatic ductal adenocarcinoma and metastatic colorectal cancer\textsuperscript{16–19}.

In this study, we generated MM-PSN, the first PSN of newly diagnosed MM based on multi-omics data from the MMRF CoMMpass study\textsuperscript{20}. Clustering of MM-PSN identified 3 patient groups and 12 sub-groups, revealing novel insights into the co-occurrence of primary translocation events such as t(4;14)-MMSET and secondary adverse lesions such as gain of 1q and whole-arm deletions of 16q and 17p. This is the first multi-omics study to classify co-occurring primary and secondary genomic aberrations and assess their clinical significance. Network modeling of comprehensive patient profiles allowed to identify and characterize groups of genetically and molecularly similar patients more accurately than in previous single-omics classifications and revealed remarkable diversity within previously defined subtypes.

RESULTS

Multi-omics patient similarity network of newly diagnosed MM reveals greater genetic and molecular heterogeneity than current classifications

We generated MM-PSN, a multi-omics PSN based on Whole-Exome Seq (WES), Whole-Genome Seq (WGS) and RNA-Seq data from 655 tumor samples from newly diagnosed MM patients enrolled in the MMRF CoMMpass study, using the similarity network fusion (SNF) method\textsuperscript{17} (see Table 1 for summary patient characteristics). In MM-PSN, each node represents a patient and connecting edges represent similarity based on multiple data types. In particular, for each sample we used: 1) gene expression and 2) gene fusion data from RNA-Seq; 3) somatic single nucleotide variations (SNVs) from WES, 4) Copy Number Alterations (CNAs; focal and broad) and 5) translocation calls from WGS (Fig. 1A, 1B). Translocations and CNAs provided the strongest contribution to MM-PSN, followed by gene expression, gene fusions and SNVs (Fig. 1C). We then applied
spectral clustering to determine groups of highly similar patients sharing features across the five data types. Our evaluation of the network using Eigen Gap and Rotation Cost suggested three as the optimal number of clusters (Fig. 1D; see Methods for further details). Differential feature analysis revealed that the three clusters were enriched for (1) Hyperdiploidy (HD) and the t(8;14) translocation of MYC (tMYC); (2) translocations t(4;14) of MMSET/FGFR3 (tMMSET) and t(14;16) of MAF (tMAF); and (3) translocation t(11;14) of CCND1 (tCCND1), respectively (Fig. 1A, 1E). We labeled each group based on these features. Group 1 (HD) included n=357 patients (54.5%) and was further enriched for mutations in NRAS and an LSAMP:RPL18 gene fusion; Group 2 (tMMSET+tMAF) included n=166 patients (25.3%) and was enriched for gain of 1q and mutations in FGFR3, DIS3 and MAX; Group 3 (tCCND1) included n=132 patients (20.15%) and was enriched for mutations in CCND1 and NRAS. To further dissect intra-group heterogeneity, we re-applied spectral clustering within each group, determining a total of 12 sub-groups (Fig.1B, 1E, Table 2, Supp Tables 1-22).

Group 1 was comprised of four sub-groups. Sub-groups 1a (n=100; 28%), 1b (n=105; 29.4%) and 1c (n=103; 28.9%) were mostly characterized by HD with differences in gain of chromosomes 7 and 11, which were almost not detected in sub-groups 1a and 1c, respectively. Sub-groups 1b, c and d were enriched for tMYC, which was virtually absent in sub-group 1a. Additionally, sub-group 1c was significantly enriched for the fusion HLA-C:IGHA1, gain(1q) and del(13q), and had a significantly higher number of female and older patients. Sub-group 1d (n=49; 13.7%) had a weaker HD signal and was significantly enriched for multiple chromosome deletions instead, including del(1p), del(8p), del(17p), del(14q), del(16q) and del(13q).

Group 2 was comprised of five sub-groups. Sub-groups 2a (n=38; 22.9%), 2d (n=24; 14.5%) and 2e (n=40; 24.1%) were enriched for the tMMSET, while sub-group 2b (n=36; 21.7%) was enriched for translocations of MAF (tMAF, tMAF-A, tMAF-B). Sub-group 2a was also enriched for mutations in PRKD2, sub-group 2b had a significantly higher number of female patients, along with sub-group 1c, and was enriched for the IGKJ5:ITGB7 gene fusion and del(16q), and sub-groups 2d and 2e were enriched for mutations in FGFR3. Del(13q) was observed in all 5 sub-groups, while gain(1q) was significantly enriched in all sub-groups with the exception of 2a. Of note, sub-group 2d was also enriched for gain of 15q and showed additional HD features, such as gain of chromosomes 3 and 19. Sub-group 2c (n=28; 16.9%) was the only sub-group in group 2 without any translocation and was mainly characterized by gain(1q), del(14q) and mutations in TRAF3.
Group 3 was comprised of three sub-groups, all enriched for tCCND1. Sub-group 3a (n=36; 27.3%) had virtually no CNAs, while sub-group 3b (n=73; 55.3%) was additionally enriched for del(13q), specifically focused at 13q12.11, TP53 biallelic inactivation (mutation + deletion), gain(11q) and other sparse CNAs. Sub-group 3c (n=23; 17.4%) was enriched for gain(1q). The co-occurrence of gain(1q) and del(16q) was identified in three different subgroups (1c, 2b, 2c), indicating that it is a recurring whole-arm secondary aberration.

We compared MM-PSN with previous classifications based on gene expression, such as the UAMS and our previous MMNet classifications\(^{11,13}\) (Supp Fig. 1A, 1B). The results showed substantial agreement among the three systems, where previous classes defined by HD or specific translocations matched the corresponding enriched sub-groups in MM-PSN. However, while gene expression-based systems identified broad disease sub-types driven by differential expression reflecting major genetic alterations, MM-PSN revealed a more granular stratification with important biological and clinical implications. Both UAMS and MMNet clustered patients with tMMSET in a unique class, while MM-PSN revealed the existence of three separate sub-groups of patients carrying tMMSET, 2a, 2d and 2e, each associated with different genetic and molecular alterations. Similarly, while HD was previously considered as a homogeneous class of patients, MM-PSN revealed three distinct HD sub-groups, 1a, 1b and 1c. Sub-group 1c had significant overlap with the low bone disease (LB) UAMS class, whose genetic distinctive features were not apparent by simple analysis of its gene expression profile and was then considered as a separate class. The CCND1 groups in UAMS (CD1/CD2) and MMNet (CCND1) were represented in the three different MM-PSN sub-groups 3a, 3b and 3c, which revealed significant genetic differences in terms of co-occurring CNAs among patients affected by tCCND1. Lastly, sub-groups 1d, characterized by prominent multiple deletions, and 2c, enriched for gain(1q) and del(14q), were poorly represented in both the UAMS and MMNet classifications and identified novel classes whose genetic features were not apparent by gene expression analysis alone. The clinical and prognostic implications of the novel MM-PSN classes are presented in the next section.

**Survival analysis of MM-PSN reveals novel prognostic findings that improve on current cytogenetic risk classification systems**

Survival analysis of the three main groups showed that patients in group 2(tMMSET/tMAF) had shorter Progression Free Survival (PFS) compared to patients in groups 1(HD) and 3(tCCND1) (p<0.01) (Fig. 2A, 2B)
and shorter Overall Survival (OS) compared to patients in group 1 (p=0.05). Analysis performed within each group revealed significant differences in survival among sub-groups, where those with the shortest median time to relapse (MTR) were all enriched for gain(1q).

While HD has been shown to confer a more favorable prognosis, concurrent gain(1q) identified a sub-group of patients, 1c(HD/tMYC/1q+), at higher risk of relapse within group 1 (MTR: 677 days; Hazard Ratio HR=1.6, p=0.04), and significantly shorter OS than patients in sub-group 1b(HD/tMYC) (HR=2.26, p=0.01) (Fig. 2C, 2D). Stratification of the whole cohort based on the co-occurrence of HD and gain(1q) revealed significantly shorter PFS and OS in patients carrying both aberrations, as compared to patients with HD and no 1q alterations (PFS: HR=1.56, p=0.008; OS: HR=2.04, p=0.004) (Fig. 3A, 3B).

tMMSET is currently considered to confer poor prognosis and identifies high-risk in the revised International Staging System (R-ISS)21. MM-PSN revealed the existence of three separate sub-groups of patients carrying tMMSET, each associated with different genetic and molecular alterations and with significantly different risk profiles. While sub-group 2e(tMMSET/1q+) had significantly worse PFS (MTR: 624 days; HR=2.04, p=0.002) and OS (Median time to death: 1033 days; HR=2.71, p=5e-04) within group 2 and overall, patients in sub-group 2a(tMMSET) had a significantly better prognosis, comparable to that of HD patients (MTR: 917 days; HR=2.35, p=0.005) (Fig. 2E, 2F, 2I, 2J). Of note, sub-group 2d, which was enriched for both tMMSET and gain(1q) and had additional HD features such as gain(15q) and gain(3p/3q), had better prognosis than sub-group 2e (MTR: 1033 days; HR=2.59, p=0.01). Conversely, sub-group 2c(1q+/14q-), which was the only sub-group in group 2 without any translocations, had a PFS profile similar to sub-group 2e and the shortest MTR in group 2 and overall (610 days).

While no significant differences in survival were detected among sub-groups of the tCCND1-enriched group 3, a slightly poorer PFS in patients in sub-group 3c enriched for gain of 1q(tCCND1/1q+) was observed (HR=1.3, p=0.3) (Fig. 2G, 2H). These findings indicate a significant overall role of gain(1q) in driving prognosis.

Multivariate Cox regression analysis including basic demographics, such as gender, age and race, the therapies received and the major alterations enriched in the network, confirmed that gain(1q) alone and co-occurring with tMMSET were both significantly associated with worse PFS and OS (PFS: HR=2.08, p=<0.001,OS: HR=3.49, p=<0.001) (Supp Figures 3 and 4), while tMMSET alone was not. Overall, patients with both tMMSET and gain(1q) (n=45) had much shorter PFS and OS compared to patients with tMMSET alone (n=43) (PFS: HR=2.3,
p=0.005; OS: HR=3.47, p=0.008) (Fig. 3C, 3D). Since the number of 1q copies has been previously reported as being prognostically relevant, we evaluated it across all the sub-groups. Sub-group 2e had the highest number of copies (p=0.0001) (Supp Fig. 2A). Stratification extended to all patients in CoMMpass with CNA data available (n=870), confirmed that patients carrying 4 copies of 1q had worse PFS and OS than patients carrying 3 copies (PFS: HR=1.4, p=0.06; OS: HR=1.6, p=0.04,) (Supp Fig. 2B, 2C). The finding of amp(1q) in group 2e indicates this subgroup identifies the sub-clonal progression of 1q copy number. We have previously shown concomitant multiple whole-arm deletions in patients with gain(1q), including 16q and 17p. Our analysis revealed significant enrichment for del(16q) in sub-group 2b(tMAF) and significant co-occurrence of gain(1q) with del(16q) in sub-groups 1c(HD/tMYC/1q+/), 2b(tMAF) and 2c(1q+/14q−), and of gain(1q) with del(17p) in sub-group 2d(tMMSET/1q+/15q−).

Biallelic inactivation of TP53 (mutation and deletion), which was enriched in sub-group 3b(tCCND1/11q+/13q−), was also significantly associated with worse prognosis (PFS: HR= 2.02, p=0.005; OS: HR= 3.08, p=0.0002), although it was overall present in a small fraction of the patient population (n=22, 3.3%) (Supp Fig. 3, 4 and 6). Of note, the translocation of MAF, which has been considered a high-risk alteration in previous studies, was not significantly associated with worse outcome. The analysis also showed that African American (AA) patients and male patients, independent of ancestry, had also worse OS (AA: HR=1.73, p=0.012; male: HR=1.50, p=0.041). In contrast, gain(15q), which was a widespread alteration in main group 1(HD) and significantly enriched in sub-group 2d(tMMSET/1q+/15q−), was associated with better PFS and OS (PFS: HR=0.66, p=3e-04; OS: HR=0.7, p=0.04) (Fig. 3E, 3F). This might in part explain the better prognosis observed in sub-group 2d compared to 2e, despite the presence of tMMSET and gain(1q).

The analysis also revealed that neither induction therapy (e.g. bortezomib/carfilzomib with lenalidomide) nor autologous stem cell transplant (ASCT) could overcome the negative prognostic impact of gain(1q). While patients with gain(1q) who received ASCT (n=108) had better prognosis compared to patients with gain(1q) who did not receive it (n=217), the outcome was still significantly poorer than in patients without gain(1q) who received ASCT (n=194) (Fig. 3G, 3H, Supp Fig. 3, 4). However, in the whole cohort (n = 655), we observed better PFS in patients treated with carfilzomib-based therapies (HR=0.46, p=0.05) and ASCT was the only treatment that was overall significantly associated with both better PFS and OS (PFS: HR=0.26, p=0.001; OS: HR=0.26, p=0.001) (Supp Fig. 3 and 4).
Given these prognostic findings, we stratified patients based on their risk classes according to the International Staging System (ISS) and the presence/absence of gain(1q). ISS is the is the most used risk score in the clinical setting and is based on serum beta-2 microglobulin and albumin\(^26\). The revised ISS (rISS) additionally includes lactate dehydrogenase (LDH) as well as the high-risk cytogenetic markers tMMSET, tMAF and del(17p), but not gain(1q)\(^21\). Our analysis revealed that gain(1q) could identify patients at higher risk of relapse in ISS classes I and III (ISS I: \(p=0.02\); ISS III:\(p=0.03\)) and rISS classes II and III (rISS II: \(p=0.004\); rISS III: \(p=0.019\)) (Fig. 3I, 3J). Gain(1q) also identified patients with significantly shorter OS in all three ISS classes (ISS I: \(p=0.013\); ISS II: \(p=0.007\); ISS III: \(p=0.02\)) and in rISS classes II and III (rISS II: \(p=0.01\); rISS III: \(p=0.002\)) (Fig. 3K, 3L).

**Gene expression analysis reveals activation of specific oncogenic pathways in MM-PSN sub-groups**

To characterize the 12 MM-PSN sub-groups functionally, we performed sample-based pathway activation analysis followed by sub-group enrichment. Table 2 and Fig. 4 provide a summary and a graphical representation of the results (Supp Table 23). The analysis was performed on a set of 456 pathways mapped onto 10 canonical cancer hallmarks\(^27\).

Group 1(HD) was overall enriched for pathways in the *tumor-promoting inflammation* and *immune evasion* hallmarks. In particular, sub-group 1a(HD/-7\(^-\)) was characterized by significant activation of inflammation and interferon-related pathways, such as IFN alpha/beta/gamma signaling, JAK1-JAK2-STAT1, IL-6 cytokine receptor ligand interactions and TRAF6-mediated induction of proinflammatory cytokines. Sub-group 1b(HD/tMYC) was enriched for telomerase activation, apoptosis and pre-NOTCH processing, while sub-group 1c(HD/tMYC/1q\(^+\)) showed activation of cell cycle. Sub-group 1d(MultiDel) was characterized by increased genome instability through activation of non-homologous end joining.

Group 2(tMMSET/tMAF) was overall enriched for pathways in the *sustaining proliferative signaling* and *deregulating cellular energetics* hallmarks. Sub-groups 2a(tMMSET), 2d(tMMSET/15q\(^+\)) and 2e(tMMSET/1q\(^+\)) were characterized by the activation of the FGF signaling pathway as a result of the t(4;14) translocation involving FGFR3/MMSET and additional activating mutations in the gene. Sub-group 2a was also enriched for several signaling pathways including the RAF/MAPK cascade, JNK signaling, growth hormone receptor, TRAIL, FAS and ErbB4. Sub-group 2b(tMAF) was characterized by activation of cell cycle, interleukin signaling, class I PI3K signaling, ErbB1 signaling, ERK, MAPK and MTOR. Sub-group 2c(1q\(^+\)/14q\(^-\)) was characterized by inhibition of
ErbB and AKT/MTOR signaling. Sub-group 2d(tMMSET/1q+/15q+) was enriched for the insulin receptor (IR), insulin-like growth factor receptor (IGFR), and insulin-receptor-related receptor (IRR) cascade, as well as for the nerve growth factor signature. Sub-group 2e(tMMSET/1q+) was characterized by activation of cell cycle, hypoxia response via HIF and replicative immortality through packaging of telomere ends. Group 3(tCCND1) was overall enriched for pathways in the genome instability, enabling replicative immortality and growth suppression evasion hallmarks. In particular, sub-group 3b(tCCND1/11q+/13q-) was characterized by increased replicative immortality activity related to polymerase switching, DNA damage and telomere stress. Both sub-groups 3a(tCCND1) and 3b(tCCND1/11q+/13q-) were enriched for regulation of TP53 activity and TP53-mediated regulation of death receptors, HDAC activity, ATF2 activity, FAS and JNK signaling, as well as targets of c-MYC transcriptional repression. Sub-group 3c(tCCND1/1q+) was enriched for NOTCH signaling.

**Gene essentiality screenings and multi-omics drug repurposing identify potential sub-group-specific vulnerabilities and small molecule and immuno-oncology candidates for future trials**

To investigate the therapeutic implications of the MM-PSN classification, we matched the up-regulated genes in each sub-group with genes essential for cell survival as determined by CRISPR-Cas9 screens (DepMap) in MM cell lines (CERES score < -0.5; see Methods for details)\(^28,29\). First, we employed the MM-PSN classifier to assign 60 MM cell lines to sub-groups based on their DNA and RNA profiles (Supp Tables 24, 31). We were able to find at least one representative cell line for 9 out of 12 sub-groups and identified essential up-regulated genes in 6 of the 9 sub-groups (1d, 2b, 2c, 2d, 3b, 3c) using the corresponding CRISPR-Cas9 data. We additionally determined essential genes in the remaining sub-groups using data from all available MM cell lines. The analysis revealed a total of 213 essential genes, some of which were potential actionable targets (Fig. 5, Supp Table 25). The overexpression of CCND2 was determined to be the major vulnerability across all group 2 (tMMSET/tMAF) and sub-groups 1c(HD/1q+) and 1d(MultiDel). CCND2 is a transcriptional target of MAF, thus its overexpression in sub-group 2b(tMAF) is a direct consequence of tMAF. We also determined that up-regulation of MAF and CCND2 were vulnerabilities in sub-groups 2d(tMMSET/1q+/15q+) and 2e(tMMSET/1q+), consistently with recent studies that have reported MAF activation also in tMMSET cell lines and primary samples\(^30\). However, MAF was not identified as essential in sub-group 2a(tMMSET), revealing that not all patients with tMMSET have MAF activation. Since the MEK-ERK pathway has been demonstrated to regulate MAF transcription, patients in MAF-
expressing sub-groups could be potential candidates for treatment with MEK inhibitors. CDK6 was another major vulnerability identified in the 2b(tMAF), supporting clinical evaluation of CDK4/CDK6 inhibitors in this class of patients. Another actionable vulnerability identified by our screening was the overexpression of FGFR3 in the tMMSET sub-groups 2a, 2d and 2e. Sensitivity to FGFR inhibitors in patients overexpressing FGFR3 is currently being tested in basket trials. Among the essential genes up-regulated in the high-risk sub-group 2e(tMMSET/1q+), we also identified IGF1R, which has been previously reported as aberrantly expressed in aggressive MM and represents a potential vulnerability that could be exploited therapeutically in this specific class of patients.

We further investigated the therapeutic implications of MM-PSN by applying our multi-omics precision medicine approach previously described to the patients in the network and performing sub-group enrichment. The analysis identified general, group and sub-group specific therapeutic options based on actionable DNA and RNA alterations represented in the database CIViC. Both the MEK inhibitor trametinib and the CDK4/CDK6 inhibitor abemaciclib were among the top 10 candidate options in several different sub-groups because of the enrichment for actionable mutations in either KRAS (trametinib and abemaciclib) or BRAF (trametinib). Trametinib, currently being tested in the MyDRUG trial, was also identified as a top-ranking option in combination with docetaxel, metformin or omipalisib in sub-group 3a(tCCND1), and with docetaxel and pemetrexed in sub-group 3c(tCCND1/1q+), because of the enrichment for NRAS mutations.

Group 1(HD) was enriched for several drugs mostly matched to actionable genes in the chromosomes affected by trisomies. For example, onartuzumab and crizotinib were top-ranking options in sub-groups 1a(HD/-7) and 1c(HD/tMYC/1q+) because of gain of MET (chr7). Other options included the CDK4/CDK6 inhibitor palbociclib associated with CCND1 gain in sub-groups 1a, 1b and 1d, the NOTCH1 antibody PF-06293622 associated with gain of NOTCH1 in all four sub-groups, the PI3K inhibitors alpelisib and taselisib associated with gain of PIK3CA and pictlisib associated with gain of PIK3CA or ERBB2 and mutations in BRAF. Additionally, sub-group 1d was associated with irinotecan because of amplification of TOP1.

In group 2(tMMSET+tMAF), tazemetostat and selumetinib were significant top-ranking options in sub-group 2a(tMMSET) because of deletion and/or under-expression of SMARCB1 and deletion of NF2, respectively. The pan-FGFR inhibitors erdafitinib and infgratinib and the cisplatin/gemcitabine combination were significantly recommended for sub-group 2d(tMMSET/1q+/15q+) because of the enrichment for FGFR3 mutations. Potential
options in sub-group 2e(tMMSET/1q⁺) included the MAPK11 inhibitor SB202190 because of under-expression of STK11, the PARP inhibitor Olaparib because of over-expression of PARP and/or under-expression of ATM, and the FGFR inhibitor erdafitinib because of mutations in FGFR3. Lastly, since immuno-oncology therapy is emerging as a promising approach in MM, we matched up- and down-regulated genes in each sub-group with current CAR-T and immuno-oncology treatments targets. While transcript expression of such genes does not necessarily correlate with the corresponding surface protein levels, it may still be an important indication of potential actionable targets specific to each sub-group. The CD19 gene, which encodes a surface glycoprotein that functions as co-receptor for the B-cell antigen receptor complex (BCR), was identified as up-regulated across groups 1(HD) and 3(tCCND1), with the exception of the gain(1q)-enriched sub-groups 1c(HD/tMYC/1q⁺) and 3c(tCCND1/1q⁺). CAR-T therapy targeting CD19 was recently approved by FDA in relapsed/refractory pediatric B-cell acute lymphoblastic leukemia (B-ALL) and for adult patients with relapsed/refractory diffuse large B-cell lymphoma (DLBCL)³⁶. A small clinical trial has recently determined that anti-CD19 CAR T-cells with high-dose melphalan and ASCT may yield clinical benefits in relapsed/refractory MM³⁷. The mucin gene MUC1, which is encoded in chromosome 1q22, was a recurrent target identified in the gain(1q)-enriched sub-groups 1c (HD/tMYC/1q⁺), 2c(1q⁺) and 2e(tMMSET/1q⁺), while MUC16 was additionally found up-regulated in 2a(tMMSET), 2d(tMMSET/1q⁺/15q⁺) and 2e(tMMSET/1q⁺). Other relevant targets in group 2 included IGF1R, which was also identified as an essential gene in 2e, CTAG2, up-regulated in 2e and ITGB7, up-regulated in 2b(tMAF) and 2e. In group 3(tCCND1), the gene MS4A1, encoding the B-cell surface protein CD20, was identified as a potential target in sub-groups 3b(tCCND1/11q⁺/13q⁻) and 3c(tCCND1/1q⁺). These results may help inform patient selection in future clinical trials.

A prognostic gene expression signature induced by MM-PSN predicts survival in an independent dataset
To validate our MM-PSN model, we first devised a gene expression classifier trained on the 12 MM-PSN sub-groups, employing an approach based on support vector machine (SVM) (Supp Table 30). Then, we tested the classifier on a gene expression dataset with 559 tumor samples from newly diagnosed patients pre-TT2 and -TT3 treatments. The classifier was able to predict patients in the 3 main groups and to replicate the prognostic findings from MM-PSN, with group 2 having significantly shorter PFS (HR=2.25, p=3e-07) and OS (HR=1.8, p=8e-05) than groups 1 and 3 (Supp Fig. 7A, 7B). Furthermore, we validated sub-group 2e as having significantly
shorter PFS (HR=3.08, p=0.004) and OS (HR=3.64, p=0.002) than 2a (Supp Fig. 7C, 7D), thus supporting the existence of two tMMSET-associated disease subtypes with different risk profiles and, specifically, the negative prognostic impact of co-occurrence of tMMSET and gain(1q).

DISCUSSION

In this study we have generated a PSN of newly diagnosed MM patients, MM-PSN, using five different data types derived from WGS, WES and RNAseq, and have determined a novel classification of MM consisting of 3 main groups and 12 sub-groups. While several studies have focused on specific genetic lesions to determine their prognostic implications, our analysis aimed at dissecting the inter-tumor heterogeneity of MM and investigating the co-occurrence of multiple alterations in an integrated fashion. Overall, our MM-PSN model organized previous knowledge on genetic markers into broad patient groups defined by translocations or hyperdiploidy, then further refined and harmonized such knowledge cohesively with other genetic and molecular markers to determine sub-groups of highly similar patients. Our model identified two novel sub-groups, one (1d) characterized by multiple chromosome deletions and one (2c) by co-occurrence of gain(1q) with del(14q). Most importantly, the MM-PSN classification uncovered novel associations between distinct MM hallmarks with powerful prognostic implications and enabled further refinement of risk stratification.

Close inspection of the MM-PSN groups identified a small number of outlier patients who lacked some of the defining features of the groups they were assigned to (Supp Fig 8). This confirms the exceptional heterogeneity of MM and demonstrates one limitation of the PSN approach. Nevertheless, this methodology enables easy identification of these outliers, which can then be subjected to further analysis to determine their unique characteristics.

The most significant findings of our study involve the gain of the long arm of chromosome 1, a high-risk feature observed in ~40% of MM cases and in several other types of cancer, including breast cancer, hepatocellular carcinoma and myeloproliferative neoplasms38–41. Patients affected by monoclonal gammopathy of undetermined significance (MGUS), a precursor of MM, and smoldering MM (SMM) who carry gain(1q) have higher risk of progression to MM, with a median time to progression of two years42,43. Other studies with newly diagnosed and relapsed/refractory MM patients have confirmed the negative prognostic impact of gain(1q) in different therapeutic regimens22,44. MM-PSN identified six sub-groups (1c, 2b, 2c, 2d, 2e, 3c) enriched for
gain(1q), all of them associated with other recurrent lesions and having the shortest median time to relapse and death compared to the other sub-groups. We have previously described 1q12 pericentromeric instability of satellite DNA as a major cause of many of the secondary sub-clonal karyotypic events in MM45. In this regard, a recent study of double-refractory MM found that karyotypic events influenced clustering of patients more than treatment or mutations, and that amp(1q) was the only high-risk feature predicting survival46. These authors suggest that chromosome instability enables sub-clones to enter different evolutionary trajectories and adapt to selective pressure of therapies and underlying treatment failures. The types of secondary karyotypic associations found in MM-PSN add significantly to the genomic heterogeneity of the disease and are most likely at least partially responsible for the resistance to therapy. The subgroup of patients having the poorest outcome in terms of both PFS and OS had co-occurrence of gain(1q) with the t(4;14) translocation involving MMSET. We further confirmed this prognostic implication in an independent dataset. Notably, while the presence of gain(1q) alone was still a significantly deleterious event, the presence of tMMSET alone was not. This is an important result, as tMMSET has always been considered a high-risk feature regardless of other co-occurring lesions21,47. In contrast, and consistent with findings from the Myeloma XI and Myeloma IX trials, gain(1q) appeared to confer poorer prognosis also to patients usually considered at lower risk, such as those carrying hyperdiploidy23. Hyperdiploid patients with concurrent gain(1q) had, indeed, significantly shorter PFS and OS than hyperdiploid patients with no gain(1q). A similar trend, although not statistically significant, was also observed in patients with tCCND1 and concurrent gain(1q), who had a median time to progression close to patients with tMMSET and gain(1q).

Our multivariate analysis also revealed a protective effect conferred by gain(15q), whose presence determined a significantly longer PFS and OS. This might also explain the better outcome observed in patients with tMMSET, gain(1q) and gain(15q) (sub-group 2d). Further investigations aimed at dissecting this potential protective effect are currently underway.

Despite increasing evidence supporting the prognostic relevance of gain(1q), current staging systems such as the International Staging System (ISS) and its revised version, rISS, do not include it21,48. Our results show that gain(1q) could significantly stratify patients in almost all risk classes into high vs low risk sub-classes, in terms of both PFS and OS. Our results independently confirm findings by Walker et al47 and suggest that gain(1q) should be incorporated into staging systems and used in the clinic to determine patient risk.
Functional characterization of the MM-PSN sub-groups through pathway activation analysis and drug repurposing based on both DNA and RNA alterations have further revealed meaningful insights with important biological and clinical implications. These findings have immediate implication for precision medicine and clinical trials, as different sub-groups of patients may respond to different targeted and immuno-oncology therapies based on their genomic and transcriptomic profiles.

Our study confirms the advantages of employing multiple features to dissect cancer heterogeneity and the ability of PSNs to handle multiple data types to generate clear and interpretable disease models. MM-PSN structures and harmonizes the complexity of MM by associating patients with highly similar genomic and transcriptomic profiles to form more granular and homogeneous classes than achieved by previous classifications. The MM-PSN classification is a valuable and accessible resource that can be employed in most clinical settings, since the features defining high-risk can be easily detected by FISH/Cytogenetics.

While the prognostic impact of gain(1q) has been previously investigated and established in numerous studies, our network model and analysis have revealed a much higher significance and centrality of this genetic lesion in risk assessment of treatment-naive MM patients. On-going research is now focused on a deeper characterization of the MM-PSN sub-groups, and in particular those enriched for gain(1q), in order to gain novel insights into the molecular mechanisms and pathways that drive each disease subtype. These studies are fundamental to advance our understanding of MM pathology and paves the way for future research into drug repurposing approaches aimed at novel therapies tailored to specific patient sub-groups.

METHODS

Dataset acquisition and primary data generation

Whole-Genome Seq (WGS), Whole-Exome Seq (WES) and RNA-Seq data were generated from bone marrow aspirates (tumor CD138+ cells) and peripheral blood (control) of 655 treatment-naive newly diagnosed MM patients enrolled in the MMRF CoMMpass study. Data was provided by MMRF and is available on the dbGaP database (http://www.ncbi.nlm.nih.gov/gap) under accession number ‘phs000748’.

All RNA and DNA samples were isolated at the North American Biobank (VARI, Grand Rapids, Michigan) and provided to the genomic characterization center at the Translational Genomics Research Institute (TGen), Phoenix, Arizona. All libraries were sequenced on Illumina HiSeq2000 or HiSeq2500 using Illumina HiSeq v3 or
v4 chemistry. Whole-Exome and RNA assays were sequenced using paired-end 83x83bp (TruSeqExome/RNA and Agilent V5+UTR single-plex) or 82x82bp (Agilent V5+UTR 8-plex pools) sequencing. Whole-Genome Long-insert assays were sequenced using 86x86bp sequencing. mRNA unstranded libraries were created from tumor RNA only (CD138+ cells). Whole-Exome and Whole-Genome libraries were created from matched tumor (CD138+ cells) and peripheral blood samples.

All primary data were generated at TGen as follows. Fastq files were aligned to the GRCh37 reference human genome and all annotation and gene models were based on Ensembl version 74. Whole-Exome and Whole-Genome fastq files were aligned to the reference genome using BWA (0.7.8) \(^{49}\). Sequence alignment map (SAM) files were converted to binary alignment maps (BAM) files using Samtools (0.1.19) \(^{50}\). Lane level quality recalibration was performed using PICARD (1.111). Duplicate fragments were identified and marked using PICARD (1.111). Joint-Indel realignment was performed on the matched tumor and constitutional bam files using GATK (GenomeAnalysisTK-3.1-1) \(^{51}\).

RNAseq fastq files were aligned to the reference genome using STAR (2.3.1z) \(^{52}\). Gene expression estimates were calculated using HT-Seq (0.6.0) counts \(^{53}\).

Somatic single nucleotide variants (SNVs) and small INDELs were identified from matched tumor-normal WES data using an integrated approach leveraging three different somatic variant callers: Seurat v2.6, Strelka v1.0.13, and MuTect v1.1.4 \(^{54-56}\). Somatic mutations were filtered to include calls made by at least two of the three callers.

Copy number alterations (CNAs) were identified using a CoMMpass-specific version of TGen’s in-house copy number tool tCoNuT_COMMPASS. Relative copy number was determined as the log2 difference between the normal and tumor normalized coverage, where normalization is the mean coverage across a 2 kb window divided by the genome-wide mode coverage. We additionally used the tool Gistic 2.023 to identify regions of the genome that were significantly amplified or deleted across multiple samples \(^{57}\). CNAs were classified as broad (arm-level) or focal (band/sub-band level).

Structural variants were identified from the whole-genome long-insert assay using the tools DELLY and an internally developed TGen caller \(^{58}\).

To detect gene fusion transcripts RNA-Seq data was aligned using TopHat2 v2.0.11 \(^{59}\). TopHat-Fusion post was set to report fusions \(^{60}\). Potential fusions called by TopHat-Fusion were subject to a guided-fusion assembly approach to assemble each of the potential fusion transcripts using the tool Trinity 2.2 for de novo assembly of RNA-Seq data and bwa-mem version 0.7.8 for alignment to the fusion reference \(^{61}\). Potential fusions were also
independently validated in the long-insert WGS data by traversing through a series of windows initially set to a size of 1000 base pairs.

**Generation and clustering of Patient Similarity Network**

For the generation of MM-PSN we applied the Similar Network Fusion (SNF) method as it did not require *a priori* feature selection and was shown to outperform methods based on single data types as well as other multi-omics approaches such as iCluster\(^{17,19,62}\). We additionally tested the methods iClusterBayes\(^{63}\), which was discarded because it required feature selection and thus may significantly introduce bias in the results, and ANF (Affinity Network Fusion)\(^{64}\), whose results were almost identical to those obtained by SNF. MM-PSN was generated using the SNF method implemented in the R package SNFtool (v2.3.0)\(^{17}\). The tool was run on 655 MM tumor samples using gene expression (50,495 genes), SNV (57,736 mutations), gene fusion (13,682 fusions), focal CNA (93 features) and Broad CNA data (39 features), using the parameters K = 50, alpha = 0.6, T = 50. The fused matrix obtained from the SNF function was clustered using the Spectral Clustering method implemented in the SNFtool package, setting k=2 to 15 (k = number of clusters). We selected three as the optimal number of clusters which maximized the Eigen Gap and minimized the Rotation Cost, as suggested by the authors of SNF (Fig. 1). For each of these three groups, we ran SNFtool again independently with the following parameters:

- **Group 1**: K = 40; alpha = 0.6; T = 50
- **Group 2**: K = 60; alpha = 0.6; T = 50
- **Group 3**: K = 20; alpha = 0.6; T = 50

We then employed spectral clustering on the fused similarity matrices, setting k=2 to 8, and selected four, five and three as the optimal number of sub-groups for groups 1, 2 and 3, respectively. Outliers were defined as patients changing group for increasing values of k.

**Network Visualization**

We retrieved all the patient pairs from the fused similarity matrix W returned by SNF and retained only those with similarity greater than the third quartile values of all possible pairs, for improved visualization. These filtered pairs
were then imported in Cytoscape (v3.7.2)\textsuperscript{65}. We used the edge-weighted spring embedded layout for visualizing the edges shown in Fig. 1A.

**Survival analysis**

Progression-free survival (PFS) and overall survival (OS) were analyzed by the Kaplan-Meier method and p-values were calculated using the log-rank test. The Hazard Ratio (HR) was calculated using the Cox proportional hazards method (coxph from the R package 'survival').

**Differential feature analysis and pathway enrichment**

Significant features across MM-PSN groups and sub-groups were identified by Kruskal-Wallis test (numerical features) and chi square test (categorical features). Sub-group enrichment for significant features was determined by Wilcoxon Rank Sum Test (numerical features) and chi square test (categorical features). CNA data were considered as numerical (GISTIC scores), while SNV and fusion data were considered binary (presence or absence). Clinical data were split into numerical (e.g. blood markers) and categorical (e.g. sex) variables and analyzed separately. Differential gene expression analysis was performed using the R package edgeR\textsuperscript{66}, comparing each sub-group of patients with all the other patients.

We calculated patient-specific pathway activation by performing single-sample gene signature enrichment using the tool GSVA (Gene Set Variation Analysis) on a set of 456 pathway signatures previously organized in 10 cancer hallmarks\textsuperscript{27}. For each sample, pathways with a GSVA score > 0.2 were considered activated. Sub-group enrichment for activated pathways was assessed by applying Fisher’s exact test to activated pathways in patients in each sub-group. P-values were corrected for multiple testing using the Benjamini-Hochberg procedure, and 0.05 was set as cut-off value for significance (Supp Table 1). The enrichment map was created using the tool Cytoscape and the plugin enhancedGraphics\textsuperscript{65,67}.

**MM-PSN classifier**

Data from the 655 samples was divided into training and validation sets in the ratio of 70:30. The classifier was built by combining three types of features, i.e. actual copy number of focal chromosome bands, translocation calls and gene expression. First, feature selection was performed on copy number values using a Recursive
Feature Elimination (RFE) strategy. XGBoost was used as an external estimator that assigned weights to the features. First, we filtered out highly correlated focal bands (correlation > 0.95) and trained XGBoost on the remaining 85 bands. The importance of each feature was measured in terms of the weights assigned by the classifier. Then, the five least important features with minimum weight were pruned from the current set of features. This process was recursively reiterated on the trimmed set until the required number of features was eventually reached. We selected the top 50 features which maximized the performance in terms of weighted recall. Then, the selected features were converted into categorical variables according to the following criteria: Between 0 and 0.5: Loss; Between 0.5 and 1.5: Deletion; Between 1.5 and 2.5: Normal; Between 2.5 and 3.5: Gain; >3.5: Amplification. The features were then further one-Hot encoded to be fed into the classifier.

For gene expression, we first selected the top 100 genes with the highest NMI (Normalized mutual information) for each of the three main groups and their subgroups. The VST (Variance Stabilizing Transformation) normalized gene expression features of the training dataset were scaled to z-scores\textsuperscript{66}. The validation gene expression dataset was normalized using the mean and standard deviation of training data. Then, we used a two-step feature selection method: first, the features were filtered based on a linear support vector machine with L1 norm (SVC-L1), then an RFE approach using XGBoost was applied similarly to as described above for CNA data. Finally, a total of 109 gene expression features were selected and converted to categorical variables as per the following criteria: Z-score >1.5: Upregulated; Z score < -1.5: Downregulated; Z-Score >-1.5 and Zscore <1.5: Normal expression. The categorical expression features were then OneHot encoded.

We then combined the one hot encoded CNA, Expression and all the 8 translocation calls side by side and used a stacking classifier with Random forest, Xgboost and linear SVC as the base learners. Then, a linear support vector machine was used as a final estimator to compute the final prediction of sub-groups taking the input from the three base classifiers. The performance on training data was calculated using 3-fold cross validation. The final model was developed using full training data and tested on 30% validation dataset separated upfront from the full dataset. Supp Fig. 9A and 9B show the Precision vs Recall curves for the training and test sets, respectively.

**Gene essentiality screening and drug repurposing analysis**

To identify essential genes in MM-PSN sub-groups, we first applied the MM-PSN classifier to sequencing data from 60 MM cell lines and assigned them to 9 different sub-groups. No cell lines were matched with sub-groups
1b, 2a and 3a. The results are given in Supp Table 2. We then retrieved the Achilles dataset, which contains the results of genome-scale CRISPR knockout screens for 18,119 genes in 793 cell lines, from the Cancer Dependency Map (DepMap) portal. We identified MM cell line-specific essential genes as those with a CERES score < -0.5, as suggested in DepMap\textsuperscript{28,29}. We matched the essential genes with significantly up-regulated genes in each sub-group, which were identified by differential expression analysis (DE) performed with the R package edgeR\textsuperscript{69}. Immunoglobulin genes were excluded. Genes with logFC $\geq$ 1.5 and False Discovery Rate (FDR) $< 0.05$ were considered up-regulated. Specifically, up-regulated genes in each sub-group were matched with genes with a CERES score $<$ -0.5 in at least one of the cell lines representative of the sub-group. For sub-groups with no representative cell lines, we considered an average CERES score of all MM cell line data available. Overall, we identified 213 essential genes (Supp Table 3).

Drug repurposing analysis was performed using an updated version of our multi-omics precision medicine platform previously described. For each patient in MM-PSN, we matched SNVs, CNAs and differentially expressed genes with the actionable alterations reported in the database CIViC (Clinical Interpretation of Variants in Cancer)\textsuperscript{35}. Patient-specific differentially expressed genes were identified by calculating z-scores from log2 VST-normalized gene expression data and selecting genes with $|z\text{-scores}| \geq 2$. Actionable alterations in CIViC are associated with drugs and/or drug combinations curated from the literature and clinical trials in MM and other cancers. We only considered associations with evidence level of A (validated), B (clinical evidence), C (case study) or D (pre-clinical). Inferential associations (level E) were discarded. Sub-group enrichment for drugs and drug combinations was assessed by applying Fisher’s exact test to drugs and drug combinations associated with patients in the sub-group (Supp Table 4).

CAR-T and immuno-oncology targets were curated from the literature and sources in our MM clinic\textsuperscript{36}. Sub-group specific targets were identified by matching this gene list with significantly up-regulated genes in each sub-group according to the analysis described above.

**Network validation**

To validate MM-PSN groups and sub-groups on an independent gene expression dataset, we trained classification models on the 655 samples using gene expression data only. The 12 sub-groups obtained by spectral clustering on the SNF fused similarity matrix were taken as the ground truth for label assignment. The classifier was developed employing a two-step process. First, an SVM classifier was developed to classify the 3
main groups using the z-scored expression of 134 genes. The 134 gene expression features were selected using SVC-L1 feature selection method implemented in the Python package Scikit-learn. The samples predicted to be in Group 2 were further sent through another SVM classifier based on 1200 features which were selected using univariate feature selection method f_classif in Scikit-learn, in order to predict membership to the five sub-groups of Group 2 (2a, 2b, 2c, 2d and 2e). The best models were selected using 10-fold cross validation at the group level and 5-fold cross validation at the sub-group level on the training data. The validation data was generated with Affymetrix GeneChip U133 plus 2.0 arrays from Shaughnessy et al and was retrieved from NCBI GEO (GSE2658).

Data availability

Sequencing data are available through protected access on the dbGaP database (http://www.ncbi.nlm.nih.gov/gap) under accession number phs000748. The MM-PSN classifier is freely available at https://github.com/laganalab/MM-PSN. The dataset used for validation was retrieved from NCBI GEO (Acc. Number GSE2658).

REFERENCES

1. Kumar, S. K. et al. Multiple myeloma. Nat Rev Dis Primers 3, 17046 (2017).
2. American Cancer Society. Cancer Facts and Figures 2019. (2019).
3. Bazarbachi, A. H., Al Hamed, R., Malard, F., Harousseau, J.-L. & Mohty, M. Relapsed refractory multiple myeloma: a comprehensive overview. Leukemia 33, 2343–2357 (2019).
4. Rajkumar, S. V. & Vincent Rajkumar, S. Multiple myeloma: 2020 update on diagnosis, risk-stratification and management. American Journal of Hematology vol. 95 548–567 (2020).
5. Morgan, G. J., Walker, B. A. & Davies, F. E. The genetic architecture of multiple myeloma. Nat. Rev. Cancer 12, 335–348 (2012).
6. Lohr, J. G. et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. Cancer Cell 25, 91–101 (2014).
7. Bolli, N. et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat. Commun. 5, 2997 (2014).
8. Laganà, A. et al. Precision Medicine for Relapsed Multiple Myeloma on the Basis of an Integrative Multiomics Approach. JCO Precis Oncol 2018, (2018).
9. Study tests targeted drugs for multiple myeloma. Cancer Discov. 9, 459 (2019).
10. Bergsagel, P. L. & Kuehl, W. M. Molecular pathogenesis and a consequent classification of multiple
11. Zhan, F. et al. The molecular classification of multiple myeloma. Blood 108, 2020–2028 (2006).
12. Broyl, A. et al. Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. Blood 116, 2543–2553 (2010).
13. Laganà, A. et al. Integrative network analysis identifies novel drivers of pathogenesis and progression in newly diagnosed multiple myeloma. Leukemia 32, 120–130 (2018).
14. Bolli, N. et al. Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. Leukemia (2017) doi:10.1038/leu.2017.344.
15. Pai, S. & Bader, G. D. Patient Similarity Networks for Precision Medicine. J. Mol. Biol. 430, 2924–2938 (2018).
16. Cavalli, F. M. G. et al. Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell 31, 737-754.e6 (2017).
17. Wang, B. et al. Similarity network fusion for aggregating data types on a genomic scale. Nat. Methods 11, 333–337 (2014).
18. Cancer Genome Atlas Research Network. Electronic address: andrew_aguirre@dfci.harvard.edu & Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell 32, 185-203.e13 (2017).
19. Pitroda, S. P. et al. Integrated molecular subtyping defines a curable oligometastatic state in colorectal liver metastasis. Nat. Commun. 9, 1793 (2018).
20. Keats, J. J. et al. Interim analysis of the mmrf CoMMpass trial, a longitudinal study in multiple myeloma relating clinical outcomes to genomic and immunophenotypic profiles. Blood 122, 532–532 (2013).
21. Palumbo, A. et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. J. Clin. Oncol. 33, 2863–2869 (2015).
22. Schmidt, T. M. et al. Gain of Chromosome 1q is associated with early progression in multiple myeloma patients treated with lenalidomide, bortezomib, and dexamethasone. Blood Cancer J. 9, 94 (2019).
23. Shah, V. et al. Prediction of outcome in newly diagnosed myeloma: a meta-analysis of the molecular profiles of 1905 trial patients. Leukemia 32, 102–110 (2018).
24. Sawyer, J. R. et al. Jumping translocations of 1q12 in multiple myeloma: a novel mechanism for deletion of 17p in cytogenetically defined high-risk disease. Blood 123, 2504–2512 (2014).
25. Sawyer, J. R. et al. An acquired high-risk chromosome instability phenotype in multiple myeloma: Jumping 1q Syndrome. Blood Cancer J. 9, 62 (2019).
26. Greipp, P. R. et al. International staging system for multiple myeloma. J. Clin. Oncol. 23, 3412–3420 (2005).
27. Iorio, F. et al. Pathway-based dissection of the genomic heterogeneity of cancer hallmarks’ acquisition with SLAPenrich. Sci. Rep. 8, (2018).
28. Tsherniak, A. et al. Defining a Cancer Dependency Map. Cell 170, 564-576.e16 (2017).
29. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. 49, 1779–1784 (2017).
30. Hurt, E. M. et al. Overexpression of c-maf is a frequent oncogenic event in multiple myeloma that promotes proliferation and pathological interactions with bone marrow stroma. *Cancer Cell* 5, 191–199 (2004).
31. Annunziata, C. M. et al. A mechanistic rationale for MEK inhibitor therapy in myeloma based on blockade of MAF oncogene expression. *Blood* 117, 2396–2404 (2011).
32. Liu, F.-T. et al. Recent advance in the development of novel, selective and potent FGFR inhibitors. *Eur. J. Med. Chem.* 186, 111884 (2020).
33. Chng, W. J., Gualberto, A. & Fonseca, R. IGF-1R is overexpressed in poor-prognostic subtypes of multiple myeloma. *Leukemia* 20, 174–176 (2006).
34. Bataille, R., Robillard, N., Avet-Loiseau, H., Harousseau, J.-L. & Moreau, P. CD221 (IGF-1R) is aberrantly expressed in multiple myeloma, in relation to disease severity. *Haematologica* 90, 706–707 (2005).
35. Griffith, M. et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* 49, 170–174 (2017).
36. Chu, F., Cao, J. & Neelalpu, S. S. Versatile CAR T-cells for cancer immunotherapy. *Contemp. Oncol. (Pozn.)* 2018, 73–80 (2018).
37. Garfall, A. L. et al. Anti-CD19 CAR T cells with high-dose melphalan and autologous stem cell transplantation for refractory multiple myeloma. *JCI Insight* 3, (2018).
38. Muthuswami, M. et al. Breast tumors with elevated expression of 1q candidate genes confer poor clinical outcome and sensitivity to Ras/PI3K inhibition. *PLoS One* 8, e77553 (2013).
39. Chen, L., Chan, T. H. M. & Guan, X.-Y. Chromosome 1q21 amplification and oncogenes in hepatocellular carcinoma. *Acta Pharmacol. Sin.* 31, 1165–1171 (2010).
40. Najfeld, V. et al. Jumping translocations of the long arms of chromosome 1 in myeloid malignancies is associated with a high risk of transformation to acute myeloid leukaemia. *Br. J. Haematol.* 151, 288–291 (2010).
41. Marcellino, B. K. et al. Advanced forms of MPNs are accompanied by chromosomal abnormalities that lead to dysregulation of TP53. *Blood Adv* 2, 3581–3589 (2018).
42. Hanamura, I. et al. Frequent gain of chromosome band 1q21 in plasma-cell dyscrasias detected by fluorescence in situ hybridization: incidence increases from MGUS to relapsed myeloma and is related to prognosis and disease progression following tandem stem-cell transplantation. *Blood* 108, 1724–1732 (2006).
43. Rajkumar, S. V., Landgren, O. & Mateos, M.-V. Smoldering multiple myeloma. *Blood* 125, 3069–3075 (2015).
44. Mohan, M. et al. Daratumumab in high-risk relapsed/refractory multiple myeloma patients: adverse effect of chromosome 1q21 gain/amplification and GEP70 status on outcome. *Br. J. Haematol.* 189, 67–71 (2020).
45. Sawyer, J. R., Tricot, G., Mattox, S., Jagannath, S. & Barlogie, B. Jumping translocations of chromosome 1q in multiple myeloma: evidence for a mechanism involving decondensation of pericentromeric heterochromatin. *Blood* 91, 1732–1741 (1998).
46. Ziccheddu, B. et al. Integrative analysis of the genomic and transcriptomic landscape of double-refractory multiple myeloma. *Blood Adv.* **4**, 830–844 (2020).

47. Chng, W. J. et al. IMWG consensus on risk stratification in multiple myeloma. *Leukemia* vol. 28 269–277 (2014).

48. Abdallah, N. et al. Clinical characteristics and treatment outcomes of newly diagnosed multiple myeloma with chromosome 1q abnormalities. *Blood Adv.* **4**, 3509–3519 (2020).

49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

50. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

51. Van der Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11–10 (2013).

52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

53. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

54. Christoforides, A. et al. Identification of somatic mutations in cancer through Bayesian-based analysis of sequenced genome pairs. *BMC Genomics* **14**, 302 (2013).

55. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).

56. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).

57. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).

58. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).

59. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).

60. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12**, R72 (2011).

61. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

62. Shen, R. et al. Integrative subtype discovery in glioblastoma using iCluster. *PLoS One* **7**, e35236 (2012).

63. Mo, Q. et al. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics* **19**, 71–86 (2018).

64. Ma, T. & Zhang, A. Integrate multi-omic data using affinity network fusion (ANF) for cancer patient clustering. in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 398–403 (2017).

65. Su, G., Morris, J. H., Demchak, B. & Bader, G. D. Biological network exploration with Cytoscape 3. *Curr. Protoc. Bioinformatics* **47**, 8.13.1-24 (2014).
66. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

67. Morris, J. H., Kuchinsky, A., Ferrin, T. E. & Pico, A. R. enhancedGraphics: a Cytoscape app for enhanced node graphics. *F1000Res.* **3**, 147 (2014).

68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

69. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

ACKNOWLEDGEMENTS

The authors thank all the patients enrolled in the MMRF CoMMpass study. This work was supported by NIH-NCI (R21-CA209875-01A1; R01-1R01CA244899-01A1), Tisch Cancer Institute (TCI) (NCI Support Grant: P30 CA196521) and the Multiple Myeloma philanthropic fund. This work was also supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

AUTHOR CONTRIBUTIONS

Contribution: S.B., A.L. and S.P. conceived and designed the study; S.B. led the study and performed data analysis; D.T.M. performed pathway and drug repurposing analysis; J.K. provided sequencing data; D.M., J.R., S.R., A.C., H.J.C., S.J. and S.P. contributed patient samples and clinical data; K.O., J.R.S. and J.T.D. provided scientific expertise; S.B., A.L., J.R.S. and S.P. wrote the manuscript; all authors revised and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.
Table 1. Patient cohort description and demographics

| Characteristic     | Value     |
|-------------------|-----------|
| No. of Patients   | 655       |
| Age               | 63 (31-93)|
| **Gender**        |           |
| M                 | 382 (50.1%)|
| F                 | 218 (33.3%)|
| N/A               | 109 (16.6%)|
| **Ethnicity**     |           |
| White             | 420 (64.1%)|
| Black or African American | 83 (12.7%)|
| Asian             | 12 (1.8%)  |
| Filipino          | 1 (0.2%)   |
| Honduran          | 1 (0.2%)   |
| Middle Eastern    | 1 (0.2%)   |
| N/A               | 137 (20.9%)|
| **Disease Stage (ISS)** |       |
| I                 | 227 (34.7%)|
| II                | 226 (34.5%)|
| III               | 182 (27.8%)|
| N/A               | 20 (3.1%)  |
| **Disease Stage (riSS)** |       |
| I                 | 128 (19.5%)|
| II                | 339 (51.8%)|
| III               | 63 (9.6%)  |
| N/A               | 125 (19.1%)|
| **Translocations**|           |
| MMSET             | 88 (13.4%) |
| CCND3             | 9 (1.4%)   |
| MYC               | 97 (14.8%) |
| MAFA              | 4 (0.6%)   |
| CCND1             | 126 (19.2%)|
| CCND2             | 6 (0.9%)   |
| MAF               | 27 (4.1%)  |
| MAFB              | 10 (1.5%)  |
| None              | 307 (46.9%)|
| Multiple *        | 19 (2.9%)  |
| **Disease Status**|           |
| PFS < 1yr         | 225 (34.4%)|
| PFS < 2yr         | 365 (55.7%)|

*CCND1 + MYC: 5; CCND2 + MYC: 2; MAF + MYC: 5; MAFB + MYC: 2; MMSET + MAF: 1; MMSET + MYC: 4
### Table 2. Summary of the MM-PSN sub-groups.

| Group/Sub-group | Selected enabling features | Median time to relapse (years) | Prognostic notes | Selected activated pathways | Selected drug candidates | Selected essential genes |
|-----------------|----------------------------|-------------------------------|------------------|---------------------------|------------------------|-------------------------|
| 3-16E           | Hypomethylating Mutations in NRAS, ARID1B, LILRDAP, CDS1, DCAF1, and PTP2B fusion | 1004 (N/A) | - | Trampt-promoting inflammation, immune evasion | NTRK3 inhibitors, Immunotherapies, Chemotherapies, Immunotherapies, Monoclonal Antibodies | APOT, EPL, MTOL, PAK5 |
| 5A, 5B, 9B     | Hypomethylating with regards of cl: 7, TPN4, BRD1 | 1082 (N/A) | - | Interferon signaling, IL-6 signaling/immunosuppression, induction of proinflammatory cytokines, NF-κB signaling | Cachectin-Induced Signaling, Stat3/4, PI3K/Akt | APO1, EPL, MTOL, PAK5 |
| 1B, 1C, 2A, 4B | Hypomethylating with MTC transcription, Mutations in MMAS | 1297 (N/A) | - | Telomere extension, Apoptosis, Phospho-MTOR signaling, Regulation of innate immune response | Distal chromatin, Chromatin, Histone, NF-κB, P53, P73 | CD138, CD164, CD141, CD38 |
| 3C, 4C, 5C, 6A, 7B | Hypomethylating with PSN translocation and Tg730 gain | 677 (N/A) | Shorter PFS than 2B | Cell cycle | MMPS inhibitors, Distal chromatin, Histone, NF-κB, P53, P73 | CD138, CD164, CD141, CD38 |
| 3D, 4C, 5C, 6A | Multiple chromosome alterations including 1q, 13q, 9q, 6q, 8q, 9p, 10q, 19q, 21q, 1p46 (N/A) | 1246 (1779) | - | Non-homologous end joining | Inflammation, Distal chromatin | CD138, CD164 |
| 2X, 3A, 4A, 5A | Translocations of MMAT7 and MAF; Gain of chromosome 1q, 13q, and 21q Mutations in KIT, PDGFRD, and CSF1R | 751 (N/A) | Shorter PFS than 3, and 1 | Sustaining proliferative signaling, Disrupting cellular energetics | Endothelial, Inflammation, Signaling, Stat3/4, PI3K/Akt, Immunotherapies, Monoclonal Antibodies | Distal chromatin, Histone, NF-κB, P53, P73 |
| 2A, 3A, 4A, 5A | Translocation of MMAT7, deletion of 13q33, 13q4, 14q, and 21q Mutations in KIT | 517 (N/A) | - | NF-kB/MAPK cascade, downstream genetic events, signaling, JAK/STAT activating receptor, MAPK signaling, PI3K/Akt signaling | Tumor suppressors, Stat3/4, NF-kB | CD138, CSK, PAK1, PAK4 |
| 2B, 3B, 4B, 5B | Translocation of MMAT7, Gain of 1p46, Gain of 1q, Gain of 14q, 14q, and 21q Mutations | 960 (1500) | - | ERK cascade, MAPK; NF-kB, interleukin signaling, Class I P53 signaling, JAK/STAT signaling, JAK3 downstream pathway | Tumor suppressors, Stat3/4 | CD138, CSK, PAK1, PAK4 |
| 3A, 3C, 4A, 5A | Gain of 5q; Deletion of 13q and 21q Mutations in MMAT7 | 610 (N/A) | - | Inhibition of AKT, MDM2, and NF-κB signaling and inhibition of Stat3 signaling | Tumor suppressors, Stat3/4 | CD138, CD164, CD141, CD38 |
| 3C, 5C, 6A, 7B | Translocation of MMAT7, Mutations in KIT and PDGFRD; Gain of 5q, 3p, 3q, and 14q, deletion of 13q | 1603 (N/A) | - | MAPK signaling, FGF/RAS/IGF1 signaling and activation, Stat3 signaling, Hypoxia response, Cell cycle | Integrins, Endothelial, Disrupting Cytoskeleton | CD138, CSK, PAK1, PAK4 |
| 3C, 5C, 6A, 7B | Translocation of MMAT7, Mutations in PTPRD and PDGFRD; Gain of 5q (highest number of copies), Deletion of 13q | 616 (1659) | Shorter PFS and OS than 2A, 3A, 3B, 3C, 3D, 4A, 4B, and 5B | FGF signaling, Cell cycle, Hypoxia response, Packaging of tollerance ends | Stat3/4, Anti-FGFR1/2/3, P2G79/187 | CD138, CSK, PAK1, PAK4 |
| 3D, 4C, 5C, 6A | Translocation of MMAT7, Gain of 1q and 14q, and 21q Mutations in KIT and NF-κB signaling | 737 (N/A) | - | Disrupting proliferative signaling, Disrupting cellular energetics | Endothelial, Inflammation, Stat3/4, NF-kB | CD138, CSK, PAK1, PAK4 |
| 3D, 4C, 5C, 6A | Translocation of MMAT7, Gain of 1q and 14q; Activation of BRAF; Inactivation of PTPRD, PTPRD and NF-κB | 1239 (N/A) | - | BRAF, PI3K, AKT, NF-κB, Stat3/4, PI3K/Akt | Tumor suppressors, Stat3/4, NF-kB | CD138, CSK, PAK1, PAK4 |
| 3C, 4C, 5C, 6A | Translocation of MMAT7, Gain of 1q and 14q, and 21q Mutations in KIT and PDGFRD, Gain of 5q, 3p, 3q, and 14q, deletion of 13q | 832 (3305) | - | Disrupting proliferative signaling, Disrupting cellular energetics | Endothelial, Inflammation, Stat3/4, NF-kB | CD138, CSK, PAK1, PAK4 |
Fig. 1. Network generation and identification of groups and sub-groups. A. Somatic genetic variants and transcriptomic features from Whole Exome (WES), Whole Genome (WGS) and RNAseq data from 655 patients in the MMRF CoMMpass study was used to generate a Patient Similarity Network (MM-PSN) using the Similarity Network Fusion approach. Edges connecting patients in the network represent similarity based on one or more feature type (e.g. orange edges in the sample network represent similarity based on SNVs, magenta edges represent similarity based on all the types of features). Spectral clustering was employed to identify patient groups and then re-applied to identify sub-groups enriched for specific features. B. Representation of MM-PSN where nodes (patients) are colored according to the three main groups identified by spectral clustering. C. The plot shows contribution of the different data types to the fused matrix, in terms of Normalized Mutual Information (NMI). D. Eigen Gap (max) and Rotation Cost (min) were used to determine 3 as the optimal number of clusters. E. Overview of MM-PSN patient groups and sub-groups. The heatmap shows characterization of the three main groups and twelve sub-groups of MM-PSN based on their enrichment for the different genomic and transcriptomic features.
Fig. 2. Survival analysis of MM-PSN identifies high-risk sub-groups. A, B. Progression Free Survival (PFS) and Overall Survival (OS) plots for the three main patient groups identified by MM-PSN, showing significant poorer outcome for the tMMSET+tMAF group. C, D. Survival plots for sub-groups of group 1 show shorter PFS in patients from sub-group 1C characterized by hyperdiploidy, tMYC and 1q gain, compared to patients in sub-group 1B which do not have 1q gain. E, F. Survival plots for sub-groups of group 2 show shorter PFS and OS in patients from sub-group 2E, enriched for tMMSET and 1q gain. G, H. Survival plots for sub-groups of group 3 do not show significant differences in either PFS or OS. I, J. Survival plots for all twelve sub-groups of MM-PSN, indicating poorer outcome of patients in sub-group 2E.
Fig. 3. Prognostic implications of gain(1q), tMMSET and gain(15q). A, B. Survival plots show 1q gain identifies a subset of hyperdiploid patients with significantly shorter PFS and OS. C, D. Survival plots show that patients with 1q gain with or without tMMSET have poorer outcome than patients with tMMSET alone. E, F. Gain of 15q is associated with better PFS and OS. G, H. Survival plots show that patients with 1q gain that received ASCT have significantly better PFS and OS compared to patients that didn’t receive ASCT. I, J. Gain of 1q significantly stratifies risk for relapse and mortality in patients in ISS classes I and III, and risk of mortality in patients in ISS class II. K, L. Gain of 1q significantly stratifies risk for relapse and mortality in patients in rISS class II, and risk of mortality in patients in rISS class III.
Fig. 4. Pathway activation in MM-PSN sub-groups. Enrichment map for selected pathways that are significantly activated in MM-PSN sub-groups. Each circle indicates a pathway and the colors represent the sub-groups with significant activation of the pathway. Edges connect pathways that share genes and edge thickness is proportional to the number of shared genes. Pathways are organized into cancer hallmarks (yellow areas), which are identified by the large blue labels.
Fig. 5. Gene essentiality screening identifies potential vulnerabilities in MM-PSN sub-groups. Selected genes that are considered essential for cell survival according to the CRISPR/Cas9 screening data retrieved from DepMap. Lower CERES scores indicate higher essentiality. A green star symbol in the heatmap header indicates that the gene was identified based on cell lines that were also specifically matched to the corresponding sub-groups. For the other sub-groups genes were identified using all the available MM cell lines in DepMap.
Fig. 6. Multi-omics drug repurposing identifies candidate therapeutic options in MM-PSN sub-groups based on enriched actionable alterations. A. Schema of the drug repurposing analysis. Somatic SNVs, CNAs and gene expression profiles are annotated with the pan-cancer database CIVIC (Clinical Interpretation of Variants in Cancer) to determine actionable alterations and the associated drugs. B. Sub-groups are annotated with the drugs associated with enriched actionable alterations. C. Targets of immuno-oncology therapies up and down regulated in each sub-group.
Supp Fig. 1. Comparison of MM-PSN sub-groups with UAMS and MMNet classes. The comparison was performed using Pearson residuals. Positive residuals are in blue and specify a positive association between the corresponding classes. Negative residuals are in red and imply a negative association between the corresponding classes. 

**A.** Comparison between the MM-PSN sub-groups and the UAMS classes (CD1/CD2: CCND1; HY: hyperdiploid; LB: low bone disease; MF: MAF; MS: MMSET; PR: proliferative).

**B.** Comparison between the MM-PSN sub-groups and the MMNet classes (CC: cell cycle; CK: cytokines; IMM: immune).
Supp Fig. 2. Prognostic implications of 1q gain and amplification. A. The distribution of the number of copies of 1q across MM-PSN sub-groups show a significantly higher number of copies in patients in 2e. The stars on top of each bar indicate significance compared to 2e (ns: P > 0.05, *: P ≤ 0.05, **: P ≤ 0.01, ***: P ≤ 0.001, ****, P ≤ 0.0001). B-C. Number of 1q copies significantly stratify PFS and OS. Amplification of 1q (4 or more copies) confers much worse prognosis than gain (3 copies).
**Supp Fig. 3. Multivariate cox-regression analysis of progression free survival.** The analysis reveals that 1q gain, its combination with tMMSET and biallelic inactivation of TP53 are significantly associated with shorter PFS, while gain of 15q is associated with better PFS. Among treatments, autologous Stem Cell Transplant (ASCT) is significantly associated with better survival in the context of all the high-risk factors included in the model. Carfilzomib-based treatments have borderline significant benefits.
### Supp Fig. 4. Multivariate cox-regression analysis of overall survival.

The analysis reveals that 1q gain, its combination with tMMSET and biallelic inactivation of TP53 are significantly associated with shorter OS. Gain of 15q has borderline significant benefit in terms of OS. Younger patients (age < median = 65 yo) have also significantly longer OS. Male and black african-american patients have significantly shorter OS. Among the treatments, only ASCT is significantly associated with better OS.

#### Overall Survival

| Variable                      | Hazard Ratio | p-value |
|-------------------------------|--------------|---------|
| tMAF                          |              |         |
| No (N=610)                    | reference    |         |
| Yes (N=26)                    | 1.62 (0.657 – 3.75) | 0.263   |
| tMMSET/1q                     |              |         |
| WT (N=288)                    | reference    |         |
| 1q gain only (N=165)          | 2.23 (1.457 – 3.41) | <0.001 *** |
| tMMSET + 1q gain (N=46)       | 3.49 (2.028 – 6.01) | <0.001 *** |
| tMMSET only (N=39)            | 0.72 (0.279 – 1.84) | 0.487   |
| TP53                          |              |         |
| WT (N=423)                    | reference    |         |
| Biallelic (N=22)              | 3.66 (1.913 – 7.00) | <0.001 *** |
| Deletion (N=63)               | 1.39 (0.818 – 2.37) | 0.222   |
| Mutation (N=8)                | 1.64 (0.351 – 7.71) | 0.528   |
| 15q                           |              |         |
| WT (N=306)                    | reference    |         |
| 15q Gain (N=330)              | 0.71 (0.465 – 1.03) | 0.067   |
| Age                           |              |         |
| Age Greater than Median (N=254) | reference    |         |
| Age Less than Median (N=345)  | 0.63 (0.406 – 0.98) | 0.04 *   |
| Median Age (N=327)            | 1.03 (0.452 – 2.64) | 0.843   |
| Gender                        |              |         |
| Female (N=257)                | reference    |         |
| Male (N=279)                  | 1.50 (1.017 – 2.22) | 0.041 *   |
| Race                          |              |         |
| White (N=412)                 | reference    |         |
| Asian (N=12)                  | 0.93 (0.445 – 2.45) | 0.279   |
| Black or african american (N=92) | 1.129 (0.62 – 2.165) | 0.012 *   |
| BTZ-based                     |              |         |
| No (N=394)                    | reference    |         |
| Yes (N=242)                   | 1.10 (0.743 – 1.63) | 0.634   |
| CFZ-based                     |              |         |
| No (N=545)                    | reference    |         |
| Yes (N=81)                    | 1.11 (0.659 – 1.87) | 0.701   |
| BTZ–IMiD–based                |              |         |
| No (N=304)                    | reference    |         |
| Yes (N=322)                   | 1.10 (0.740 – 1.64) | 0.635   |
| BTZ–IMiD–CFZ–based            |              |         |
| No (N=420)                    | reference    |         |
| Yes (N=96)                    | 1.00 (0.455 – 2.23) | 0.991   |
| CFZ–IMiD–based                |              |         |
| No (N=476)                    | reference    |         |
| Yes (N=158)                   | 0.90 (0.540 – 1.51) | 0.691   |
| IMiDs–based                   |              |         |
| No (N=462)                    | reference    |         |
| Yes (N=174)                   | 0.75 (0.515 – 1.09) | 0.136   |
| ASCT                          |              |         |
| No (N=305)                    | reference    |         |
| Yes (N=331)                   | 0.26 (0.163 – 0.460) | <0.001 *** |

# Events: 132; Global p-value (Log-Rank): 1.8453e–17
AIC: 1452.35; Concordance Index: 0.76
Supp Fig. 5. Multivariate cox-regression analysis of overall survival. A, B. Survival plots show that patients with 1q gain and tMMSET who received ASCT have significantly better PFS and OS compared to patients that didn’t receive ASCT. C, D. Survival plots show that patients with 1q gain and tMMSET who received ASCT have significantly better PFS but no significant difference in OS.
Supp Fig. 6. Biallelic inactivation of TP53. The plots show that biallelic inactivation of TP53, i.e. both deletion and mutation, is significantly associated with worse PFS and OS.
Supp Fig. 7. Validation of MM-PSN sub-groups in an independent dataset. A gene expression classifier was generated based on the 12 MM-PSN sub-groups and used to predict sub-groups in a cohort of newly diagnosed MM patients pre-TT2 and pre-TT3 treatment. A, B. Survival plots of the predicted three main groups show significantly worse PFS and OS of group 2, concordantly with MM-PSN findings. C, D. Survival plots of the predicted sub-groups in group 2 indicate worse PFS and OS of sub-group 2e, concordantly with MM-PSN findings.
Supp Fig. 8. Outliers in MM-PSN. The three main groups and their outliers, which were identified by re-applying spectral clustering with increased number of groups.
Supp Fig. 9. Performance of the MM-PSN classifier. Precision vs Recall curves for (A) training set and (B) test set.