Overexpression of transcripts originating from the MMSET locus characterizes all t(4;14)(p16;q32)-positive multiple myeloma patients

Jonathan J. Keats, Christopher A. Maxwell, Brian J. Taylor, Michael J. Hendzel, Marta Chesi, P. Leif Bergsagel, Loree M. Larratt, Michael J. Mant, Tony Reiman, Andrew R. Belch, and Linda M. Pilarski

Multiple myeloma (MM) is a B-lineage malignancy characterized by diverse genetic subtypes and clinical outcomes. The recurrent immunoglobulin heavy chain (IgH) switch translocation, t(4;14)(p16;q32), is associated with poor outcome, though the mechanism is unclear. Quantitative reverse-transcription–polymerase chain reaction (RT-PCR) for proposed target genes on a panel of myeloma cell lines and purified plasma cells showed that only transcripts originating from the WHSC1/MMSET/NSD2 gene are uniformly dysregulated in all t(4;14)POS patients. Different transcripts detected, multiple myeloma SET domain containing protein (MMSET I), MMSET II, Exon 4a/MMSET III, and response element II binding protein (RE-IIBP), are produced by alternative splicing and alternative transcription initiation events. Translation of the various transcripts, including those from major breakpoint region 4-2 (MB4-2) and MB4-3 breakpoint variants, was confirmed by transient transfection and immunoblotting. Green fluorescent protein (GFP)-tagged MMSET I and II, corresponding to proteins expressed in MB4-1 patients, localized to the nucleus but not nucleoli, whereas the MB4-2 and MB4-3 proteins concentrate in nucleoli. Cloning and localization of the Exon 4a/MMSET III splice variant, which contains the protein segment lost in the MB4-2 variant, identified a novel protein domain that prevents nucleolar localization. Kinetic studies using photobleaching suggest that the breakpoint variants are functionally distinct from wild-type proteins. In contrast, RE-IIBP is universally dysregulated and also potentially functional in all t(4;14)POS patients irrespective of fibroblast growth factor receptor 3 (FGFR3) expression or breakpoint type. (Blood. 2005;105:4060-4069) © 2005 by The American Society of Hematology

Introduction

Many hematologic malignancies are characterized by unique, and often diagnostic, translocations. Multiple myeloma (MM) has no unique translocation, although translocations involving the immunoglobulin heavy chain (IgH) locus on chromosome 14 and multiple partner chromosomes are present in 70% to 80% of patients. The translocation mechanism appears to be linked to isotype switching, as most breakpoints are located in IgH switch regions. The recurrent translocations t(11;14)(q13;q32), t(4;14)(p16;q32), and t(14;16)(q32;q23) are cumulatively present in approximately 40% of patients. These translocations also predict outcome, as t(11;14)POS patients have an improved prognosis while t(4;14)POS and t(14;16)POS patients have a worse prognosis compared with patients without these genetic events.

The switch translocations in MM separate the strong 3’ alpha and mu enhancers of the IgH locus onto different derivative chromosomes. For t(4;14), this results in the expression of fibroblast growth factor receptor 3 (FGFR3) by the 3’ alpha enhancers, while the mu enhancer increases the expression of WHSC1/MMSET/NSD2. This coordinate dysregulation of at least 2 genes makes the identification of a true target gene difficult. We and others have shown that FGFR3 is expressed in only 70% to 75% of t(4;14)POS patients, and the lack of FGFR3 expression generally correlates with the loss of the der(14) chromosome. The poor outcome associated with t(4;14)POS patients lacking FGFR3 expression suggests that FGFR3 may not be the only relevant target gene. The dysregulation of multiple myeloma SET domain-containing protein (MMSET) on der(4) by the mu enhancer is a well-characterized event of t(4;14). However, in 30% to 65% of patients, the breakpoints are downstream of the proper translation initiation site, making the potential contribution of MMSET unclear.

The 4p16 genomic region involved in t(4;14) is linked with a number of human genetic disorders including Huntington disease, Wolf-Hirschhorn syndrome (WHS), and the autosomal dominant skeletal disorders (hypoachondroplasia, achondroplasia, thanatophoric dysplasia types I and II). The skeletal disorders are linked to activating mutations of FGFR3. Some of these mutations are found in 5% to 10% of the FGFR3-expressing t(4;14)POS MM patients. The minimally deleted WHS critical region of 165 kb contains 2 genes: WHSC1/MMSET/NSD2 and WHSC2. There are 2 different transcripts transcribed from the MMSET gene. The first originates upstream of the proper translation initiation site and the second originates downstream.
is alternatively spliced into 3 mRNA species: MMSET type I, MMSET type II, and MMSET type III.13,15,22 MMSET II, the full-length protein, likely regulates gene expression by modifying histone methylation patterns.24 The second transcript originates within intron 9 of the MMSET gene.25 Translation of this transcript initiates in exon 15 and produces the protein response element II binding protein (RE-IIBP), which is identical to the C-terminus of MMSET II.

Here we show by quantitative reverse-transcription–polymerase chain reaction (qRT-PCR) analysis that the overexpression of both MMSET splice variants and RE-IIBP is the only unique characteristic of all t(4;14)P(OS) patients. Furthermore, we show that all MMSET breakpoint variants and RE-IIBP transcripts produce protein products. However, the major breakpoint region 4-2 (MB4-2)/MB4-3 breakpoint variants localize differently from the wild-type/MB4-1 proteins and have altered kinetics within the nucleoplasm. Alternatively, RE-IIBP is universally overexpressed and encodes the proper full-length protein in all t(4;14)P(OS) patients.

Materials and methods

Patient samples and cell lines

The study was approved by the University of Alberta/Capital Health Authority and Alberta Cancer Board research ethics boards. Bone marrow (BM) aspirates were obtained from 304 MM and 112 monoclonal gammopathy of undetermined significance (MGUS) patients after informed consent. Daudi, Raji, RPMI-8226, KMS-11, NCI-H9262, and JMM3 cell lines were maintained as previously described.12,13 The OPM-2 and LP-1 cell lines were purchased from the German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). The KMS-18 cell line was a kind gift from Takemi Otsuki (Kawasaki Medical School, Japan).26

Sample purification

Bone marrow mononuclear cells (BMMCs) were purified on Ficoll-Hypaque Plus (Amersham-Pharmacia Biotech, Uppsala, Sweden) density gradients using standard conditions. For bulk BM RNA samples, 2 to 10 million cells were suspended in Trizol Reagent (Invitrogen, Carlsbad, CA). Plasma cells were purified on an Automacs Magnetic Cell Separator (Miltenyi Biotec, Auburn, CA) using anti-CD138 microbeads. Alternatively, plasma cells with a CD38+ and CD138− phenotype were sorted on Miltenyi Biotec (Auburn, CA) using anti-CD138 microbeads. The purity of all sorted samples was verified to be more than 90% by morphologic examination.

Polymerase chain reactions

The reaction conditions and primers used to detect t(4;14) and FGFR3 were described previously.12 For qRT-PCR, 1 μg total RNA was first treated with 1 U DNase I (Sigma-Aldrich, St Louis, MO) for 15 minutes and then converted to cDNA using the TaqMan Reverse Transcription Reagents Kit (Applied Biosystems, Foster City, CA). qRT-PCR reactions were performed in a volume of 50 μL with 1x TaqMan Universal PCR Master Mix No AmpErase UNG (Applied Biosystems), 2 to 2.5 μL TaqMan Assays-by-Design or Assays-on-Demand (Applied Biosystems) primer and probe mixes, and 5 ng RNA converted to cDNA as template. Prior to test reactions, all primer and probe mixes were titrated with a paired glyceraldehydedehyde phosphate dehydrogenase (GAPDH) control to generate conditions that validated the ΔΔCT analysis method described in ABI User Bulletin no. 2,27 with a minimum PCR efficiency of 92%. All reactions were performed on an ABI Prism 7700 Sequence Detection System (Applied Biosystems). Primers used are as follows: Assays-on-Demand GAPDH (Hs99999905_m1), TACC3 (Hs00170571_m1), FGFR3 (Hs01179829_m1), LETM1 (Hs00360061_m1), MMSET Total (Hs00370212_m1), and WHSC2 (Hs00171805_m1); and Assays-by-Design (Table 1). Since the RE-IIBP assay could also amplify DNA, we confirmed that no detectable DNA template was present with no reverse transcription (RT) controls. No amplification was detectable in the absence of reverse-transcribed cDNA.

Expression vector construction

The open reading frames (ORFs) of MMSET I and MMSET II obtained from PLB were PCR amplified with High Fidelity Platinum Taq DNA polymerase (Invitrogen) using primers indicated in Table 1 and cloned into pcR4.0-TOPO (Invitrogen). The NdeI/XhoI and Ndel/Δstop-XhoI cloned fragments were transferred by site-directed cloning to pDRN-3 or pDRN-Dual, respectively, of the Creator Cloning System (BD Biosciences, San Jose, CA). The ORF of interest was then transferred from the donor vector into either pL-EGFP-C1 or pLPS-3’EGFP using Cre Recombinase (BD Biosciences). The alternatively spliced transcript, Exon 4a/MMSET III (The National Center for Biotechnology Information accession no. AY694128), that includes 5’ minor groove binder/nonfluorescence quencher (MGB/NFQ). The start and stop codons in the cloning primers are underlined; mutated and restriction endonuclease target sequences are indicated by lowercase letters. Any mutations incorporated within coding regions are silent mutations that maintain the proper amino acid sequence.

Transfections, immunoblotting, and microscopy

Lipofectamine 2000 (Invitrogen) was used to transfect HeLa cells. Protein expression was verified 24 hours after transfection by immunoblot. Cells were released from plates with 1x trypsin, washed 3 times with phosphate-buffered saline, and lysed at 5 × 10⁶ to 10⁷ cells/mL in 1% CHAPS buffer (GIBCO). The open reading frames of MMSET I and MMSET II obtained from PLB were PCR amplified with High Fidelity Platinum Taq DNA polymerase (Invitrogen) using primers indicated in Table 1 and cloned into pcR4.0-TOPO (Invitrogen). The NdeI/XhoI and Ndel/Δstop-XhoI cloned fragments were transferred by site-directed cloning to pDRN-3 or pDRN-Dual, respectively, of the Creator Cloning System (BD Biosciences, San Jose, CA). The ORF of interest was then transferred from the donor vector into either pL-EGFP-C1 or pLPS-3’EGFP using Cre Recombinase (BD Biosciences). The alternatively spliced transcript, Exon 4a/MMSET III (The National Center for Biotechnology Information accession no. AY694128), that includes 5’ minor groove binder/nonfluorescence quencher (MGB/NFQ). The start and stop codons in the cloning primers are underlined; mutated and restriction endonuclease target sequences are indicated by lowercase letters. Any mutations incorporated within coding regions are silent mutations that maintain the proper amino acid sequence.

Table 1. PCR primer and probe sequences

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’ RE-IIBP</td>
<td>TGATATCTCTGGATTTGTTCA</td>
</tr>
<tr>
<td>3’ RE-IIBP</td>
<td>GCCCCCAAAAATCCTCACCT</td>
</tr>
<tr>
<td>RE-IIBP probe</td>
<td>CCGTCCCTTCTACCATACCGTAC</td>
</tr>
<tr>
<td>5’ Exon 4a</td>
<td>CAGTCGAGATAAGGAG</td>
</tr>
<tr>
<td>3’ Exon 4a</td>
<td>GGGAGAGAGAGAGAG</td>
</tr>
<tr>
<td>Exon 4a probe</td>
<td>TTGTGGTGATCTCTGAC</td>
</tr>
<tr>
<td>5’ MMSET</td>
<td>CAGCTCTCTCTGACCTCTAACG</td>
</tr>
<tr>
<td>3’ MMSET</td>
<td>AGCCGACCTGGTACCATC</td>
</tr>
<tr>
<td>MMSET I probe</td>
<td>CTCCCCAAAACGACC</td>
</tr>
<tr>
<td>5’ MMSET II</td>
<td>TTAATGCTGAGAATCCGAACT</td>
</tr>
<tr>
<td>3’ MMSET II</td>
<td>TCCAAATGCTGATGTTCTCTCA</td>
</tr>
<tr>
<td>MMSET II probe</td>
<td>CCAGGACCAATCCTT</td>
</tr>
</tbody>
</table>

Taqman primer and probe pairs were designed by Applied Biosystems using the Assays-by-Design service. All probes contain the 5’ reporter dye 6-FAM (carboxyfluorescein) and the 3’ minor groove binder/nonfluorescence quencher (MGB/NFQ). The start and stop codons in the cloning primers are underlined; mutated and restriction endonuclease target sequences are indicated by lowercase letters. Any mutations incorporated within coding regions are silent mutations that maintain the proper amino acid sequence.
(3-[(3-cholamidopropyl)dimethylammonio]-1-propane-sulfonic acid) plus 10 μg/mL leupeptin, 10 μg/mL antipain, and 1 mM phenylmethylsulfonyl fluoride (Sigma-Aldrich). Lysate (50 μL) was run on a 5% stacking/8% separating sodium dodecyl sulfate–polyacrylamide gel electrophoresis gel. Fusion proteins were detected with a polyclonal anti-GFP serum (BD Biosciences). All imaging of GFP-tagged proteins was performed on live HeLa cells in conditioned culture media using an LSM 510 confocal microscope with a ×40/1.3 oil objective (Carl Zeiss, Thornwood, NY) and a Tempcontrol-mini (Carl Zeiss) objective warmer set at 37°C to minimize heat loss using the LSM 510/Version 3.0 SP3 software package (Carl Zeiss).

To determine the in vivo localization, the pinhole was set to collect a 2-μm optical slice and 2 μg/mL Hoechst 33342 (Molecular Probes, Eugene, OR) DNA stain was added to the media prior to imaging. Fluorescence recovery after photobleaching (FRAP) experiments were performed with the pinhole at max (1000 μm slice), a constant digital zoom of 4.5 (0.1 × 0.1 μm scaling), and a constant bleaching region of interest (ROI) height. The total imaging time and interval between image acquisitions was determined in pilot experiments for each construct. For each experiment, the signal intensity for 3 different ROIs was collected using the LSM 510/Version 3.0 SP3 software package (Carl Zeiss): ROI1, the bleached region; ROI2, the entire nuclear fluorescence; and ROI3, the background fluorescence. The relative fluorescence intensity (RFI) was determined by first subtracting ROI3 from both ROI1 and ROI2 and then calculated as (ROI1 time point/ROI1 prebleach)/(ROI2 time point/ROI2 prebleach). The first image acquired after bleaching was set to an RFI of 0, and the corrected prebleach image was set to an RFI of 1.

**Statistics**

Survival statistics were performed as previously described.12 Student t tests were calculated with Excel (Microsoft, Redmond, WA) using an unpaired 2-sided analysis.

**Results**

**Occurrence and significance of t(4;14)**

Patient samples were collected from September 1994 to February 2004 from the University of Alberta Hospital and Cross Cancer Institute. Our expanded cohort now includes 304 MM patients and 112 MGUS patients of whom, respectively, 43 (14.1%) and 2 (1.8%) are t(4;14)POS.12 These patients can be subgrouped into 3 different breakpoint clusters, MB4-1, MB4-2, and MB4-3, based on the size of the RT-PCR product produced.12,15 Of the 45 t(4;14)POS patients, 32 have the MB4-1 breakpoint, while 6 and 7 have the MB4-2 and MB4-3 breakpoints, respectively. Hybrid transcripts from MB4-1 patients encode the full-length wild-type MMSET protein, while hybrid transcripts from MB4-2 and MB4-3 patients lack the first or first and second translated exons of MMSET, respectively (Figure 1B).

The expression of FGFR3 was determined as previously described using a single-stage RT-PCR assay.12 Within the t(4;14)POS MM patients, FGFR3 expression was detectable in 31 (72%) of 43. Both t(4;14)POS MGUS patients expressed detectable levels of FGFR3. This frequency of FGFR3 expression is consistent with previous reports.12,14 Analysis of sequential BM samples, principally diagnosis and relapse, from 80 MM patients of whom 14 are t(4;14)POS and 66 are t(4;14)NEG, has not identified a change over time in the expression of FGFR3 or t(4;14) status.

The poor clinical outcome associated with t(4;14) is well established.2,10,12 However, the impact of FGFR3 expression and breakpoint type determined by RT-PCR needs further clarification. No significant difference in survival exists between t(4;14)POS/FGFR3POS and t(4;14)POSFGR3NEG patients (P = .425; HR = 0.723; 95% CI, 0.274-1.727) (Figure 2A). To determine if the expression of a full-length MMSET protein influences survival, we compared the survival difference between MB4-1 and MB4-2/MB4-3 patients and there was no significant difference (P = .421; HR = 0.743; 95% CI, 0.305-1.644) (Figure 2B). Therefore, neither the expression of FGFR3 nor the overexpression of wild-type or truncated forms of MMSET influences the poor prognosis associated with this translocation.

**Identification of potential t(4;14) target genes**

A number of genes flanking the genomic breakpoints at 4p16 have been suggested as potential t(4;14) target genes. Our working
A patient population was determined from CD138 analysis to include other genes. The expression of potential target genes in cell lines, respectively, did not fit our working hypothesis, as TACC3 and WHSC2, our most distant telomeric and centromeric genes, dysregulated as a result of t(4;14) translocation. We moved the study into patient samples and did not expand our analysis if our working hypothesis was that a t(4;14) target gene would be overexpressed or underexpressed at the mRNA level in all t(4;14) samples. As the effects of the mu and 3' alpha IgH enhancers may act over large distances, we identified all known genes within the 0.5-Mb region of 4p16.3 surrounding the translocation breakpoint sites (Figure 1A). We initiated the study with all previously proposed target genes with the intention of expanding our analysis if our most distant telomeric or centromeric genes fit the hypothesis.

The pilot experiments in a small panel of MM cell lines identified TACC3, FGFFR3, LETM1, MMSET total, and RE-IIBP as potential t(4;14) target genes dysregulated as a result of t(4;14) (Table 2). As TACC3 and WHSC2, our most distant telomeric and centromeric genes, respectively, did not fit our working hypothesis, we moved the study into patient samples and did not expand our analysis to other genes. The expression of potential target genes in a patient population was determined from CD138+ plasma cells isolated from diagnostic or relapse BM samples of 17 MM patients, of whom 6 are t(4;14)POS and 11 are t(4;14)NEG. For one of the t(4;14)POS patients, we purified both the diagnostic and relapse sample, bringing the total number of purified BM samples to 18. This cohort includes both diagnostic and relapse samples, and the t(4;14)POS patients included equal numbers of FGFR3 expressers and nonexpressers.

As pilot experiments identified MMSET as a potential target gene, we expanded the analysis of the patient samples to test the principle alternative splice variants of MMSET (ie, Exon 4a/MMSET III, MMSET I, and MMSET II) (Figure 1B). Within the patient samples, only transcripts originating from the MMSET locus (total MMSET, MMSET III, MMSET I, MMSET II, and RE-IIBP) were significantly dysregulated (Table 3). The expression of TACC3 was similar in both groups. Of interest, one t(4;14)POS patient lacking FGFR3 expression had an 18-fold increase in the expression of TACC3. This may reflect a translocation event where FGFR3 is deleted from der(14) but TACC3 is maintained and subsequently overexpressed. Within the panel of genes analyzed, only transcripts originating from the MMSET locus fit our working hypothesis, suggesting that the target gene of this translocation is likely the protein product of one of these transcripts.

Since RE-IIBP represents an overexpressed transcript that does not vary between breakpoint types, we confirmed its overexpression on a larger panel of unpurified MM BMMCs. This group of patients included 25 t(4;14)NEG and 21 t(4;14)POS patients. The median relative expression level of RE-IIBP in the t(4;14)NEG patients was 2.80 (range, 0.42-7.91), while in the t(4;14)POS samples it was 90.59 (range, 3.86-366.67), P less than .001 (Figure 3). The t(4;14)POS patient with the lowest expression was also tested as a purified sample and had an expression level of 124.78. Therefore, the low level is likely due to a low plasma cell percentage within the unpurified sample. Interestingly, the overexpression of RE-IIBP and other MMSET transcripts in ex vivo cells considerably exceeds that detected in MM cell lines.

Characterization of wild-type MMSET variants

In principle, the use of 2 different transcription initiation sites, and subsequent alternative splicing (Figure 1B), results in the production of 4 different wild-type protein products: MMSET I, MMSET II, RE-IIBP, and Exon 4a/MMSET III (Figure 1C). The localization of MMSET I, MMSET II, and RE-IIBP was determined by transient transfection of HeLa cells with both N- and C-terminal–tagged variants. As predicted, based on protein homology and the presence of nuclear localization signals (NLSs), the wild-type/MM4-1 MMSET I and II variants localized to the nucleus, however, they are excluded from nucleoli (Figure 4A-B). Some differences exist in the localization patterns between the N- and C-terminally tagged variants. The localization patterns are more pronounced in the C-terminal tags, which we believe reflects a lower level of expression, likely reflecting a difference in the translation efficiency of the MMSET and GFP start codons. Furthermore, the C-terminally tagged variants are not perfectly excluded from nucleoli like the N-terminal tags. The localization of RE-IIBP was the exact inverse of the wild-type MMSET variants. RE-IIBP

![Image](66x638 to 295x727)

**Figure 2. Kaplan-Meier survival plots.** (A) Survival comparison of the 43 t(4;14)POS patients expressing or not expressing FGFR3. 31 and 12 patients, respectively. (B) Survival comparison of MB4-1 versus MB4-2/MB4-3 combined t(4;14) POS patients, 31 and 12 patients, respectively. (P < .001).
resides in 2 different compartments: in foci within the cytoplasm and as a nuclear population almost exclusively localized to nucleoli (Figure 4A-B). This localization pattern is consistent in both N- and C-terminally tagged variants observed in live cells (Figure 4A-B), and also methanol-fixed and paraformaldehyde-fixed cells (data not shown).

Although localization does not necessarily reflect function, the localization pattern of MMSET II, which contains the predicted histone methyltransferase SET domain, is almost identical to the staining of DNA/chromatin by Hoechst dye (Figure 4C). Therefore, MMSET II may be a functional histone methyltransferase, as an interaction with chromatin is likely a necessary aspect of proteins within this family. Unlike MMSET II, the localization of MMSET I was often diffuse, though it may also interact with DNA/chromatin.

Characterization of novel t(4;14)-specific MMSET variants

In the t(4;14)POS MM patients with MB4-2 and MB4-3 breakpoints, translation of the hybrid transcripts and/or de novo transcripts from secondary translation initiation sites are predicted to produce truncated proteins lacking the N-terminus of MMSET, as the wild-type translation initiation site is lost (Figure 1B-C). To determine if hybrid transcripts from each breakpoint variant can produce protein products, we cloned the type I and II breakpoint variants from the beginning of exon 4 and 5 to reflect the hybrid transcripts present in MB4-2 and MB4-3 patients, respectively. These cloned fragments, lacking the proper translation of the hybrid transcripts and/or de novo transcripts from secondary translation initiation sites. For both the type I and II breakpoint variants, we detected protein products that reflected the predicted protein sizes if alternative translation initiation sites in
predicted size is produced from the alternative translation initiation site in exon 15, the predicted RE-IIBP translation initiation site, and therefore the localization pattern is likely representative of the endogenous protein.

The truncated MB4-2/MB4-3 breakpoint variants were localized using the C-terminal–tagged constructs. All 4 variants, both type I and II from each of the MB4-2 and MB4-3 genomic breakpoints, localized to the nucleus. However, unlike the wild-type/MB4-1 proteins that contain the N-terminus of MMSET, all 4 breakpoint variants are enriched in nucleoli (Figure 5B). Interestingly, in both MB4-2 and MB4-3 patients the only identified domain lost from MMSET is the N-terminal PWWP domain (Figure 1C). In MB4-3 patients the entire domain is lost, while in MB4-2 patients a part of the domain is maintained, however, the P-W-W-P motif is lost. Therefore, it appears that either the N-terminal PWWP domain or a yet-unidentified domain in the N-terminus of MMSET regulates its localization. Based on localization, it appears that MMSET variants from MB4-1 patients may function differently from those expressed in MB4-2 and MB4-3 patients.

Characterization of a novel localization domain within MMSET

The mechanism controlling the localization differences appears to be encoded by the N-terminus of MMSET. To determine if the N-terminus can independently regulate the localization of MMSET, we cloned the Exon 4a/MMSET III splice variant previously identified by Malgeri et al.15 from one of our MB4-1 patients. This novel splice variant contains an in-frame stop codon in the alternatively spliced exon 4a (Figure 1B), which leads to a truncated protein that largely represents the N-terminal portion deleted in the MB4-2 breakpoint (Figure 1C). The predicted protein shares 15 amino acids with the MB4-2 variants and though a portion of the N-terminal PWWP domain is lost, the P-W-W-P motif is maintained. The localization of Exon 4a/MMSET III, regardless of which terminus was tagged, was nuclear and excluded from nucleoli just like the wild-type/MB4-1 proteins (Figure 6A). Interestingly, no NLS has been identified in this variant, nor have we identified one using online prediction programs. However, due to its small size, 30.2 kDa, exon 4a/MMSET III may be capable of free diffusion across the nuclear membrane.

To determine if the N-terminus of MMSET characterized by Exon 4a/MMSET III could alter the localization of a nucleolar protein, we cloned MMSET III into a GFP-B23 expression vector.30 B23 is a nucleolar protein involved in ribosome biogenesis. Transient transfection of the GFP–MMSET III–B23 construct resulted in a mixed population of GFP-positive cells (Figure 6). A blinded quantization of 500 cells from random fields of view showed that 53.1% had an MMSET phenotype, 41.5% a B23 phenotype, and 5.4% an unclassifiable phenotype (diffuse nuclear staining in both compartments). Therefore the N-terminus of MMSET contains a domain that regulates its exclusion from nucleoli and can even prevent a nucleolar protein from localizing to the nucleolus.

Kinetics of MMSET variants

The localization differences between the wild-type/MB4-1 and MB4-2/MB4-3 breakpoint variants suggested a loss of function for MB4-2 and MB4-3 patients. Since the function of MMSET is unknown, we indirectly tested this hypothesis using a technique called fluorescence recovery after photobleaching (FRAP). FRAP involves the bleaching of a small region of fluorescence and recording over time the recovery of fluorescence into the bleached region, as bleached molecules are replaced by unbleached molecules, to determine the time to half recovery (t1/2). FRAP recovery
kinetics are mediated by size, cellular compartment, compartment viscosity, affinity of protein-protein interactions (binding/unbinding events), protein-protein/structure collisions, and temperature. The FRAP recovery kinetics of the N-terminal–tagged MMSET variants showed the predicted results, as the protein size increased (MMSET III/MMSET I/MMSET II) the t1/2 times increased, due in part, to diffusion kinetics of larger molecules (Table 4). However, the t1/2 of MMSET II, compared with MMSET I, was much slower (150 vs 4 seconds) than predicted by simple diffusion kinetics. These differential kinetics suggest that MMSET II, but not MMSET I, is binding to a nucleoplasmic structure with very high affinity.

The C-terminal–tagged MB4-2 and MB4-3 breakpoint variants were compared with the similarly tagged wild-type/MB4-1 MMSET variants. For both MMSET I and II constructs, the t1/2 times were faster for the C-terminal tags compared with the N-terminal tags, however, the difference was not statistically significant (Table 4). The recovery of the truncated MB4-2 and MB4-3 variants was substantially faster than their wild-type/MB4-1 counterparts ($P < .001$) (Figure 7A; Table 4). The loss of the N-terminus of MMSET drastically affects the mobility of MMSET variants, as the t1/2 of the type I constructs decreased to 26% and the type II constructs decreased to 14.6% and 9.2% of the wild-type/MB4-1 protein. The N-terminus appears to be a vital part of a synergistic interaction that involves multiple component domains of MMSET, as even the additive t1/2 of MMSET I and MB4-2 II, which represents the entire MMSET II protein, would not match the t1/2 of MMSET II (21.96 versus 130.00 seconds).

**Table 4. Fluorescence recovery after photobleaching results**

<table>
<thead>
<tr>
<th>Variant</th>
<th>N-terminal GFP t1/2 time, s</th>
<th>C-terminal GFP t1/2 time, s</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSET I</td>
<td>4.0 ± 0.07</td>
<td>2.957 ± 0.06</td>
</tr>
<tr>
<td>MMSET II</td>
<td>150.0 ± 0.08</td>
<td>130.0 ± 0.06</td>
</tr>
<tr>
<td>MMSET III</td>
<td>0.986 ± 0.12</td>
<td>—</td>
</tr>
<tr>
<td>MB4-2 I</td>
<td>—</td>
<td>0.772 ± 0.05</td>
</tr>
<tr>
<td>MB4-2 II</td>
<td>—</td>
<td>19.0 ± 0.07</td>
</tr>
<tr>
<td>MB4-3 I</td>
<td>—</td>
<td>0.772 ± 0.03</td>
</tr>
<tr>
<td>MB4-3 II</td>
<td>—</td>
<td>12.0 ± 0.05</td>
</tr>
</tbody>
</table>

The mean t1/2 recovery time for each construct is shown plus/minus the standard deviation. The bleached ROI, for each FRAP experiment from the different MMSET variants, was placed specifically in the nucleoplasm to measure the kinetic recovery of the nucleoplasmic pool of each variant. — indicates experiments were not conducted.

**Discussion**

The aim of this study was to identify gene products that are uniformly dysregulated in all (t(4;14))OS patients. Using qRT-PCR
Overexpression of MMSET transcripts should result in increased levels of the encoded protein products and potentially downstream myelomagenic effects. However, the dysregulated MMSET transcripts would not result in consistent protein products, as 30% of t(4;14)POS patients have genomic breakpoints that separate the first, or first and second, translated exons from the remaining translated exons. Alternative translation initiation sites have been identified in exon 4 and exon 6,1,3 which may be used by MB4-2 and MB4-3 breakpoint patients, respectively. Immunoblot analysis of type I and II C-terminal GFP-tagged MB4-2/MB4-3 constructs containing the entire exon 4 and exon 5, respectively, generated the predicted protein products. Therefore, the alternative translation initiation sites are probably functional.

Localization of MMSET I and II with GFP constructs confirmed the predicted nuclear localization. The constructs with N-terminal GFP tags were excluded from the nucleolus, while the constructs with C-terminal tags were largely nucleoplasmic with weak staining of the nucleoli. However, this discrepancy was easily explained once the localization of the MB4-2 and MB4-3 variants was identified as nuclear and enriched in nucleoli, as it appears that the wild-type/MB4-1 MMSET variants generally use the proper translation initiation site in exon 3. Although, on rare occasions the variants appear to use the alternative translation initiation sites in exons 4 and 6, which produce protein products that enrich in nucleoli. Therefore, an important domain for MMSET localization exists in the N-terminus, which is encoded by exons 3 and 4, and is lost in the MB4-2 and MB4-3 breakpoint variants. The natural MMSET splice variant, Exon 4a/MMSET III, contains exons 3, 4, and the alternatively spliced exon 4a. Exon 4a contains an in-frame stop codon leading to a truncated protein that predominantly represents the protein segment lost by the MB4-2 breakpoint variant. Confirming the presence of a regulatory domain, the Exon 4a/MMSET III variant localized to the nucleus and was excluded from the nucleolus. The PWWP domain19 is the only identified domain in this region and has recently been shown to bind DNA in vitro and chromatin in vivo.40-42 However, if the PWWP domain is essential, the question remains how the limited segment of the domain retained in Exon 4a/MMSET III is capable of mediating the localization. A fusion construct of Exon 4a/MMSET III with B23, a nucleolar protein, resulted in a mixed population of transiently transfected cells, the majority of which had a wild-type/MB4-1 MMSET phenotype of nucleolar exclusion. Therefore, this MMSET domain can override the nucleolar localization signal present in B23.

Since the predominant localization of all MMSET variants is nuclear, the localization of RE-IIBP to cytoplasmic foci with a nucleolar pool was unexpected. This localization pattern was consistent in transient transfections with both C- and N-terminal GFP tags and in stable transfectants with the N-terminal GFP tag. Because both the MB4-2 and MB4-3 type I breakpoint variants and RE-IIBP localize to the nucleolus but have no overlapping amino acid sequences, there appear to be 2 independent nucleolar localization mechanisms located within the N- and C-terminal regions.

To determine if the localization pattern differences between the wild-type/MB4-1 MMSET and MB4-2/MB4-3 breakpoint variant constructs reflected a functional difference, we determined the intracellular association kinetics of each variant using FRAP. As expected, the larger proteins had the slowest recovery kinetics. Interestingly, the recovery of MMSET II is very slow and may be similar to some histone H1 variants,43 though direct comparisons with identical conditions have not yet been performed. This slow recovery kinetic and the high degree of colocalization with DNA in
live cells suggest that MMSET II may be interacting with a chromatin component at a very high affinity that is not observed for other MMSET variants. The affinity of this interaction may explain the difficulties in obtaining strong bands by immunoblotting, as detergents used in the protein extraction step may not sufficiently solubilize this protein/chromatin complex. Also, if the expression level of a slow-moving protein is tightly regulated, a substantial loss of function might be expected in haplo-insufficient cells. Consistent with this hypothesis, haplo-insufficiency of the MMSET homolog NSD1 is associated with Soto syndrome and haplo-insufficiency of WHSC1/MMSET is likely a causative factor in WHS.22,24 Both the type I and II MB4-2/MB4-3 breakpoint variants have substantially reduced recovery kinetics compared with the wild-type/MB4-1 MMSET variants. Therefore, not only does the N-terminus control the localization of MMSET but it also influences the mobility of the protein. Furthermore, the contribution of the N- and C-terminus to the mobility of MMSET II must be synergistic, as MMSET I and MB4-2 type II would have an aggregate recovery comparable with MMSET II if the mobility was simply additive.

RE-IIBP was initially identified, as it bound response element II of the interleukin-5 (IL-5) promoter in vitro and regulated IL-5 transcription in vivo.25 However, our localization data would suggest this is a minor activity given that very little RE-IIBP localized to the nucleoplasm where this function is predicted to occur. Furthermore, the PWWP domain present in RE-IIBP is similar to the PWWP domain of DNA methyltransferase 3b (Dnmt3b), which binds DNA unspecifically.42 Thus the function of RE-IIBP remains elusive and requires further study. Interestingly, the overexpressed RE-IIBP transcripts detected in t(4;14)-positive patients appear to be de novo transcription events originating from the RE-IIBP promoter and not IgH hybrid transcripts.

The mechanism by which t(4;14) contributes to myelomagenesis remains unclear, however, 2 different possibilities exist based on our observations. Based on a classic model of simple, direct gain of function, RE-IIBP is a likely candidate, as it is the only RE-IIBP transcript that is not functionally disrupted. Alternatively, RE-IIBP or the much more abundant MMSET I, and particularly MMSET II, may act through indirect mechanisms (eg, interfering or abrogating the normal function of MMSET or its important protein/chromatin partners). Support for this second hypothesis comes from the related gene MLL, which is dysregulated by more than 40 different reciprocal fusion proteins resulting in proteins that have either nuclear or cytoplasmic localization, with no shared unifying properties.45,46 Recently it has been shown that all oncogenic forms of MLL retain an ability to interact with Menin, a product of the MEN1 tumor suppressor locus, whose loss phenocopies loss of MLL.47 Future studies are required to determine the protein products of the MMSET locus and the mechanism by which they contribute to myelomagenesis.

Acknowledgments

We are grateful for the skilled technical assistance of Jennifer Szylowsk, Dr. and Tara Tiffler with the screening assays and Gerry Barron and Dr. Xuejun Sun for assistance with the microscopy. We would like to thank Dr. Takemi Otsuki for his continued generous contribution of MM cell lines.

References


42. Chen T, Tsujimoto N, Li E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol. 2004;24:9048-9058.


Overexpression of transcripts originating from the MMSET locus characterizes all t(4;14)(p16;q32)-positive multiple myeloma patients

Jonathan J. Keats, Christopher A. Maxwell, Brian J. Taylor, Michael J. Hendzel, Marta Chesi, P. Leif Bergsagel, Loree M. Larratt, Michael J. Mant, Tony Reiman, Andrew R. Belch and Linda M. Pilarski