

# blood

Prepublished online Sep 24, 2008;  
doi:10.1182/blood-2008-02-141770

## **Inherited and acquired variations in the hyaluronan synthase 1 (HAS1) gene may contribute to disease progression in multiple myeloma and Waldenstrom's macroglobulinemia**

Sophia Adamia, Amanda A Reichert, Hemalatha Kuppusamy, Jitra Kriangkum, Anirban Ghosh, Jennifer J Hodges, Patrick M Pilarski, Steven P Treon, Michael J Mant, Tony Reiman, Andrew R Belch and Linda M Pilarski

---

Information about reproducing this article in parts or in its entirety may be found online at:  
[http://bloodjournal.hematologylibrary.org/misc/rights.dtl#repub\\_requests](http://bloodjournal.hematologylibrary.org/misc/rights.dtl#repub_requests)

Information about ordering reprints may be found online at:  
<http://bloodjournal.hematologylibrary.org/misc/rights.dtl#reprints>

Information about subscriptions and ASH membership may be found online at:  
<http://bloodjournal.hematologylibrary.org/subscriptions/index.dtl>



Inherited and acquired variations in the hyaluronan synthase 1 (HAS1) gene may contribute to disease progression in multiple myeloma and Waldenstrom's macroglobulinemia

Sophia Adamia<sup>1</sup>, Amanda A. Reichert<sup>1</sup>, Hemalatha Kuppusamy<sup>1</sup>, Jitra Kriangkum<sup>1</sup>, Anirban Ghosh<sup>1</sup>, Jennifer J. Hodges<sup>1</sup>, Patrick M. Pilarski<sup>2</sup>, Steven P. Treon<sup>3</sup>, Michael J. Mant<sup>4</sup>, Tony Reiman<sup>1</sup>, Andrew R. Belch<sup>1</sup> and Linda M. Pilarski<sup>1</sup>

<sup>1</sup>Dept. Oncology, <sup>4</sup>Dept. Medicine, <sup>2</sup>Dept. Electrical and Computer Engineering, University of Alberta and Cross Cancer Institute, and Dept. Medical Oncology, <sup>3</sup>Dana-Farber Cancer Institute,

Corresponding author: Linda M. Pilarski

11560 University Ave.

Cross Cancer Institute

Edmonton, AB, Canada T6G 1Z2

Telephone: (780) 432 8925

Fax: (780) 432 8425

Email: [lpilarsk@ualberta.ca](mailto:lpilarsk@ualberta.ca)

**Running title:** Inherited and acquired genetic variations in HAS1

**Category:** Neoplasia

**Prior presentation:** Presented in Abstract format at the 49<sup>th</sup> annual meeting of the American Society of Hematology, Atlanta, Georgia, 2007.

**Abbreviations:** B-CLL, B chronic lymphocytic leukemia; BEC, buccal epithelial cells; BM, bone marrow; BMC, bone marrow cells; BP, branch point; ESE, exonic splicing enhancer; ESS, exonic splicing suppressors; gDNA, genomic DNA; GV, genetic

variation; HAS1, hyaluronan synthase 1; HD, healthy donors; HPC, hematopoietic progenitor cells; ISE, intronic splicing enhancer; ISS, intronic splicing suppressors; MGUS, monoclonal gammopathy of undetermined significance; MM, multiple myeloma; MT, mutated type; NCBI – National Center for Biological Information; PC, plasma cells; PB, peripheral blood; PBMC, peripheral blood mononuclear cells; PPT, polypyrimidine tract; SNP, single nucleotide polymorphism; SF, splicing factors; SR, splicing regulators; SS, splice site; WM, Waldenstrom's macroglobulinemia; WT, wild type.

## **Abstract**

To characterize genetic contributions towards aberrant splicing of the hyaluronan synthase 1 (HAS1) gene in multiple myeloma (MM) and Waldenstrom's macroglobulinemia (WM), we sequenced 3616 bp in HAS1 exons and introns involved in aberrant splicing, from 17 patients. We identified a total of 197 HAS1 genetic variations (GVs), a range of 3-24 GV/patient, including 87 somatic GV acquired in splicing regions of HAS1. Nearly all newly identified inherited and somatic GV in MM and/or WM were absent from B-CLL, non-malignant disease and healthy donors. Somatic HAS1 GV recurred in all hematopoietic cells tested, including normal CD34<sup>+</sup> hematopoietic progenitor cells and T cells, or as tumor-specific GV restricted to malignant B and plasma cells. An in vitro splicing assay confirmed that HAS1 GV direct aberrant HAS1 intronic splicing. Recurrent somatic GV may be enriched by strong mutational selection leading to MM and/or WM.

## Introduction

Splicing of pre-mRNA contributes to protein diversity in humans. Pre-mRNA splicing is regulated by cis- and trans-splicing elements, involving a complex repertoire of splicing factors, with spliceosome assembly directed by splicing motifs in the DNA template<sup>1</sup>. Evidence is accumulating for an association of defective gene splicing with susceptibility to and progression of cancer<sup>2;3</sup>. Genetic variations underlying cancer are linked to altered splicing<sup>4-6</sup>. Genetic changes in malignant cells can alter the genomic context of splice sites by activating otherwise weak or cryptic splice sites, leading to aberrant exon skipping and/or intron retention<sup>1</sup>. Inherited polymorphisms and somatic (acquired) mutations underlie aberrant splicing in cancer<sup>4-8</sup>. Most aberrantly spliced products involve loss of tumor suppressor activity<sup>3;6;7</sup>. However two genes, cyclin D1<sup>4;9</sup> and hyaluronan synthase 1 (HAS1)<sup>10-12</sup>, undergo aberrant splicing to generate proteins with new functions or localizations that may directly promote cancer.

Hyaluronan synthases have been implicated in malignant transformation<sup>13-16</sup>. HAS1 overexpression has been described in patients with multiple myeloma (MM)<sup>10</sup>, Waldenstrom's macroglobulinemia (WM)<sup>11</sup>, colon<sup>17</sup>, ovarian<sup>18;19</sup>, bladder<sup>20;21</sup> and endometrial carcinomas<sup>22</sup>. We identified a family of aberrant HAS1 splice variants, termed HAS1Va, HAS1Vb and HAS1Vc<sup>10</sup> in MM and WM that are undetectable in B-chronic lymphocytic leukemia (B-CLL) and healthy donors (HD)<sup>10;11</sup>. Aberrant intronic splicing of HAS1 pre-mRNA, spanning exons 3-5, correlates with significantly reduced survival in MM<sup>10</sup>. In WM, upregulation of HAS1 intronic splice variants occurs in a majority of CD20<sup>+</sup> B-lineage cells, detected by single cell RT-PCR<sup>11</sup>. HAS1Va has been detected in bladder cancer<sup>20</sup>.

Aberrant splicing results from genetic variations (GV) including substitutions, deletions and insertions detected in the sequence of classical splicing elements and within exons and introns<sup>23-30</sup>. These mutations include missense mutations, which change amino acids and consequently protein function, and nonsense or silent mutations leading to frame-shifting and novel protein production<sup>27;31-33</sup>. "Deep" intronic mutations may cause aberrant splicing of disease-related genes by creating or strengthening cryptic splicing

elements<sup>34;35</sup>. GVs promote aberrant splicing in genes encoding MLH1, MLH2<sup>8</sup>, CHEK2<sup>6</sup>, RB1<sup>36;37</sup>, p53<sup>5</sup>, NF1<sup>28</sup>, BRCA1<sup>38;39</sup>, PTEN<sup>40</sup>, and the cystic fibrosis transmembrane conductance regulator<sup>41</sup>.

Here, the contributions of germline origin and/or malignant-cell specific GVs to aberrant HAS1 splicing were determined by extensive sequencing of HAS1 gene segments from buccal epithelial cells (BEC), hematopoietic progenitor cells (HPC), T, B and plasma cells (PC) obtained from MM and WM patients. BEC represent the germline “host” genotype, while HPC, T, B and PC represent normal and malignant components of the hematopoietic lineage in MM and WM patients. We found a total of 197 GVs in 17 MM and WM patients. Nearly all of these GVs were absent from 23 control subjects, including 4 with B-CLL, 11 with monoclonal gammopathy of undetermined significance (MGUS) and 8 HD. In addition to 12 known HAS1 polymorphisms, in 17 patients we found 46 novel germline GVs, 87 somatic GVs and 52 unclassified GVs among all tested cell types. Among somatic GVs, defined as those that were absent from autologous BEC, 26 were found in presumptively normal HPC and T cells from MM/WM patients, and 61 tumor-specific GVs were found only in B and PCs of MM and/or WM patients. An in vitro splicing assay confirmed that a combination of germline and somatic GVs leads to aberrant HAS1 splicing.

## Materials and Methods

**Patients and Controls.** The present study includes samples from ten MM (age range=41-85 years, mean=64), seven WM patients (age range=44-76 years, mean=63), and as control subjects: four with B-CLL, eleven with MGUS (age range=61-84 years), and eight HDs (age range 50-69 years, with the exception of one HD who was 30 years of age). MM patients included 5 with IgG myeloma and 5 with IgA myeloma.; eight had detectable bone lesions. There are no known familial forms of disease, and no familial relationships among the MM and WM patients analyzed. The MM and WM patients expressed at least one HAS1 splice variant. About 60% of patients have such variants.<sup>10;11</sup> Patients were recruited independently at the Cross Cancer Institute and the University of

Alberta Hospital (Edmonton, AB), and the Dana-Farber Cancer Institute (Boston, MA). All patients were diagnosed according to the recommendations of the International Myeloma Working Group, and the consensus guidelines from the 2<sup>nd</sup> International Workshop on Waldenstrom's Macroglobulinemia.<sup>54,55</sup> Peripheral blood (PB), bone marrow (BM) and BEC samples were taken at the time of diagnosis or at follow up, after approval from the University of Alberta and the Alberta Cancer Board Institutional Review Boards or the Dana-Farber Institutional Review Board; informed consent according to the Helsinki Protocol was obtained from all subjects. All patients and HD included in this study were of Caucasian origin.

**Tissue and sample preparation.** Tissue and sample preparation, and cell sorting, were conducted as previously described<sup>42</sup>. gDNA samples from sorted cells and unfractionated PBMC or BMC were isolated using QIAamp DNA Blood mini kit (Qiagen, Mississauga, ON) according to the manufacturer's instructions. For 17 MM and WM patients sequencing of all indicated HAS1 gene segments and HAS1 minigenes was performed on populations of six sorted PB-B cells, three sorted BM-B cells, two sorted PCs, six sorted T cells, eight sorted CD34+HPCs, two unfractionated PBMCs, three unfractionated BMCs and eight BECs. Among the 23 control subjects, HAS1 gene segments were analyzed from purified B cells of four B-CLL (intron 4), unfractionated PBMC from eleven MGUS, sorted B and T cells from two healthy donors and unfractionated PBMC from six HDs.

**Cloning and sequencing.** Cloning and sequencing of HAS1 gene segments utilized gDNA isolated from BEC, purified cell subsets, BMC or PBMC obtained from MM, WM patients or HDs, PBMC from MGUS and purified B cells from B-CLL. MM and WM patients expressed HAS1 and its splice variants. The HAS1 genomic segments from exon 3 to exon 5 were amplified, cloned and sequenced using a series of primer sets (Figure 1 and Table 1) spanning the region involved in the observed HAS1 aberrant splicing (exons 3-4 and introns 3-4). The reverse and forward primers used in these PCR reactions were significantly overlapped, to evaluate the accuracy of PCR and sequencing in multiple sequencing reactions (Figure 1). PCR products were cloned and 3-19 subclones were sequenced in both directions for each HAS1 gene segment from each

subset of cells, using the ABI3130xl DNA capillary analysis system (Applied Biosystems, Foster City CA). For one small segment of intron 4 from one patient, only two subclones were sequenced. Due to primer overlap for HAS1 gene segments, we obtained up to 60 sequencing reactions for each sample, which included the ~40-100 bp range of exon-intron boundaries of the HAS1 gene.

For each subject, we also generated up to fifteen minigenes of 3503 bp (chromosomal location 56912272-56908769) in length which spanned gDNA of HAS1 exon 3 to exon 5. A minigene is the product of a gDNA PCR reaction. Usually, this segment includes exons and introns of any given gene, preserving the linkages among genetic variations on the DNA strand that was copied. In our case, HAS1 minigenes include the genomic sequence of HAS1 exons 3, 4, and 5 and introns 3 and 4. The main purpose for using minigenes in this study was to determine whether or not recurrent GVs are expressed as clusters and if they belong to the same allele. Each minigene was sequenced both directions using HAS1 gene primer sets shown in Figure 1. Sequencing of HAS1 minigenes allowed us to determine whether or not the recurrent GVs detected in patients occurred as a cluster of mutations. The regions of HAS1 that were sequenced included exons 3 and 4, introns 3 and 4 and part of exon 5.

Overall, over 4000 sequencing reactions were performed and analyzed. Sequencing runs included a defined patient sample to validate our sequencing analysis. Additionally, cloning and sequencing were conducted by 7 different individuals over the course of this study, and sequence analysis was performed by three different individuals, all with consistent results. As a control, some samples were blinded and sequenced more than once, with consistent results. All PCR reactions were performed using High Fidelity Taq polymerase which has proofreading capability. For randomly chosen reactions, direct sequencing was done to verify the subcloning analysis.

Conditions for gDNA PCR were as follows: 50ul PCR reaction mix contained 50ng gDNA, 5ul of 1XPCR buffer, 2mM MgSO<sub>4</sub>, 0.2mM dNTPs, 0.4mM HAS1 primer set (Figure 1), and 0.5U High Fidelity (HiFi) Platinum Taq (Invitrogen). The PCR cycling

parameters were: denaturation for 5min at 94°C, followed by denaturation for 30sec at 94°C, annealing for 40sec at 60°C, and extension at 68°C for 5min for 35 cycles, with a final extension period of 10min at 72°C. The HAS1 PCR products were cloned into the pCR2.1 TOPO TA cloning system (Invitrogen) and sequenced using BigDye V1.1 and V1.1 chemistry (ABI) according the manufacturer's instructions.

**Sequencing analysis.** Obtained sequences were analyzed using sequencing analysis software, while alignment was done using SeqScape software; both software packages were provided by ABI. Obtained sequences were compared to the HAS1 reference sequence reported in the NCBI database. For the analysis reported here, GVs from MM and WM patients were included if they were present on at least 20% of the subclones sequenced for that sample, or were identified in more than one patient. HAS1 GVs were identified as recurrent if they were detected in more than one patient. GVs were categorized as hematopoietic/germline origin if no BEC were available for sequencing. GVs were categorized as confirmed germline only if BEC sequences were available for comparison with hematopoietic cells. Other categories were as reported in the results section. HAS1 sequences from 23 control subjects were screened for the presence, or not, of GVs previously identified in MM and/or WM patients.

**In vitro splicing assay.** A HAS1 minigene, a segment of HAS1 gDNA, extending from exon 3 to exon 5 (NC\_000049) was amplified from gDNA of a WM patient who expressed HAS1Vb transcripts. HAS1 minigene includes sequence of the HAS1 exons 3, 4 and 5 and introns 3 and 4. The amplified HAS1 minigene was joined to the upstream sequence of HAS1 cDNA including exon 1 and 2 at SmaI site located within exon 3 to generate a hybrid HAS1 cassette comprised of cDNA linked to a gDNA fragment (Figure 3f). Four subclones of the hybrid fragments were cloned into pcDNA3 (Invitrogen, Burlington, ON), yielding pcDNA3HAS1-g3-4-5 constructs. RNA splicing was analyzed by transfecting the construct into HeLa cells using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. Cells were harvested 24 hours post-transfection for RT-PCR. We made a plasmid construct containing full length HAS1

cDNA, pcDNA3-HAS1-FL (Ghosh et al. in prep.), for use as a positive control in the transfection experiments.

Total RNA was isolated from transfected HeLa cells using a standard Trizol isolation method (Invitrogen). RT-PCR reactions were conducted following a standard protocol (Invitrogen). PCR primer sequences used in the RT-PCR reactions are as follows:

HAS1FL sense primer        5' GCGGTCCTCTAGGCCTATATAGGA 3'

HAS1FL antisense primer    5' CTGGAGGTGTACTTGGTAGCATAA 3'

HAS1Vb sense primer        5' GCGGTCCTCTAGAATCCTGCCAG 3'

HAS1Vb antisense primer    5' CTGGAGGTGTACCTGCACGGGGGC 3'

**Bioinformatic Analysis.** To evaluate recurrent GVs leading to aberrant HAS1 splicing in patients, we first identified classical splicing sites (5' and 3' SS, BP, PPT) and splicing elements (putative exonic and intronic enhancers and suppressors—ESE, ISE, ESS and ISS) located within the alternatively spliced exons and introns of this gene using “ESE finder” web interface (Release 2.0). Additionally, cis-splicing elements (exonic and intronic enhancers (ESE, ISE) and suppressors (ISE, ISS), splicing branch point (BP) and polypyrimidine tract (PPT) of wild type (WT) and mutated (MT) HAS1 gene segments were mapped and evaluated using the publicly available bioinformatic software-Splicing Signal Analysis tools (<http://www.ebi.ac.uk/asd-srv/wb.cgi>). Next, GVs detected through cloning and sequencing were mapped with the splicing elements identified by the above-mentioned methods. Each GV was evaluated alone and in combination to determine whether any of these variations or clusters of GVs had a predicted effect on the activation of cryptic splice sites in the HAS1 gene. Activation of cryptic splice sites was determined using the Splice Site Prediction tool (<http://www.cbs.dtu.dk/biolinks/pserve2.php>).

## Results

### Sequencing analysis of HAS1 gene segments in MM and WM

Because of their involvement in the aberrant splicing of HAS1 pre-mRNA, we sequenced genomic segments (exons and introns) and/or minigenes of HAS1 from exon 3 to exon 5

(3616bp or 3503bp, chromosomal location at 56912295bp–56908679bp or 56912272–56908769bp). We obtained samples from 10 MM and 7 WM patients expressing HAS1 splice variant transcripts and sequenced genomic HAS1 from defined cell subsets. To avoid false detection of GVs, 3-19 subclones or minigenes were sequenced in both directions for each cell subset. Only those GVs present in 20% or more of the subclones for a given sample are reported here.

GVs were defined as recurrent if they were detected in at least two patients. Sequencing analysis identified 50 recurrent and 147 unique HAS1 GVs that include substitutions, insertions and deletions (Table 2a-c, Supplementary Table 1). We identified 46 novel germline and 87 somatic HAS1 GVs (acquired in hematopoietic and/or tumor cells but absent from BEC), as well as, 52 unclassified novel GVs and 12 frequent NCBI single nucleotide polymorphisms (SNPs). A median of 23 GVs/patient were identified (range=3-24). The 50 recurrent HAS1 GVs included 28 inherited (16 germline origin GVs and 12 NCBI-SNPs), 10 acquired hematopoietic origin and 7 acquired tumor-specific GVs (Table 2a, b; classification of GVs is described below). We also detected 5 GVs that were classified as hematopoietic/germline origin, because no BEC were available to confirm germline or hematopoietic origin for the patients in whom these GVs were detected. Each category of recurrent HAS1 GVs was detected in 2 to 15 of the 17 MM and WM patients analyzed (Table 2c). Both inherited and acquired sets included GVs recurrent only in MM (7/50, ~14%), only in WM (10/50, ~20%) or shared by MM and WM (34/50, ~68%) (Table 2b).

The presence or not of GVs from the 17 MM/WM was determined for the NCBI HAS1 gene sequence and the HAS1 exons and introns from 23 control subjects (Table 2a). HAS1 segments were sequenced from 4 B-CLL, 11 MGUS and 8 HDs. In contrast to MM and WM, the HAS1 sequence from these control subjects had a higher degree of homology with the NCBI reference sequence (NC 000019). Most of the control subjects expressed the NCBI SNPs found in MM/WM, although four of the HDs expressed only major alleles. Four germline GVs were detected in B-CLL and a subset of MGUS, which were found in only minority of subclones or were absent from the HDs; these may be as

yet unreported polymorphisms. Of 147 unique GVs, 8 were detected sporadically in HDs or MGUS, in only minority of subclones per sample, and with one exception, each GV was found in only one of the 23 control subjects. Seven of the 8 unique GVs found in control subjects were classified as germline/hematopoietic origin (see below). This indicates that the majority of the HAS1 GVs reported here for MM and WM patients, including all of the recurrent acquired GVs, appear restricted to MM and WM.

### **Classification of GVs detected in MM and WM patients (Figure 2a)**

GVs were classified as follows: 1) GVs detected only in B and PC from patients were classified as “tumor-specific”. These GVs were absent from T, HPC and BEC, as well as from control subjects; the stage of disease during which these were acquired is unknown. 2) GVs identified in hematopoietic cell populations but absent from BEC were defined as being of “hematopoietic-origin”. They were found in all hematopoietic populations tested from MM and WM, including HPC and T cells (non-malignant) and B/PC (malignant). Their absence from BEC indicates that these are somatic GVs acquired by presumptively normal HPC and transmitted to their T and B-lineage progeny. Hematopoietic-origin GVs were absent from cell subsets of control subjects. 3) Newly identified HAS1 GVs identified in all cell populations, including BEC, were classified as “germline-origin”, four of which were detected in control subjects and may be unreported SNPs. In MM and WM, these substitutions were frequently homozygous, defined by their presence in every subclone sequenced for a given patient. We also identified a high frequency of the mutated alleles for 12 NCBI HAS1 SNPs. The mutated alleles of these SNPs were present in most MM and WM patients and were also detectable in control subjects (Table 2a).

### **GVs detected in MM and WM (Figure 2b)**

Tumor cells from MM and WM had four or more tumor-specific GVs in their B and/or PC, but the majority of these tumor-specific GVs were unique. In addition, 10/17 (59%) of MM and WM patients carried recurrent tumor-specific HAS1 GVs in B/PC, while 11 (65%) had recurrent hematopoietic-origin GVs and 16 (94%) had recurrent germline-

origin GVs (Table 2c). The type, genomic location and the distribution of recurrent GVs in MM and WM, are reported in Table 2b. None of the recurrent somatic GVs were found in control subjects.

Among the seven recurrent tumor-specific (TS) HAS1 GVs, four were detected only in WM and three were shared by WM and MM. Among these GVs two (#1TS and #2TS) are missense transversions and one, #5TS, is a deletion. Other recurrent GVs on intron 3 and intron 4 are transitions (Table 2b). Furthermore, groups of tumor-specific GVs are co-expressed in the same patients. For example, exon 3 #1TS, intron 3 #3TS and intron 4 #6TS are co-expressed together; a second co-expressed group included #2TS on exon 3 and #7TS on intron 4.

For 10 recurrent GVs of hematopoietic origin (HO), distributed across 65% patients, 6 (50%) recur in both MM and WM, while 3 HO GVs (40%) are specific to MM and one, #3HO (10%), is specific to WM (Table 2b). Acquired hematopoietic-origin GVs were detected in as many as 5 different patients.

Of 16 recurrent germline origin (GO) HAS1 GVs (defined by their presence in BEC), the majority are on exon 3 or intron 4. Most novel germline GVs (11/16) are shared by MM and WM patients, with three (#9GO, 10GO, 15GO) recurring only in MM and two (#6GO, 8GO) recurring only in WM (Table 2b). These may be candidate SNPs not yet reported. The sequencing analysis also detected mutated alleles of 12 NCBI-SNPs. These HAS1 SNPs are found in 88% of the MM and WM patients analyzed. Coding GVs detected on exon 3 and 4 (inserts in Figure 2b) were all shared by both MM and WM patients.

Five recurrent GVs were provisionally classified as “hematopoietic/germline-origin” because no BEC were available for the patients analyzed. Of these five GVs, two were shared by MM and WM, one was specific to MM and two were specific to WM (Table 2b).

As indicated above, three tumor-specific and six hematopoietic-origin HAS1 GVs were shared by MM and WM. The existence of these shared, somatically acquired GVs suggests that there may exist fundamental similarities in the events that underlie development and progression of MM and WM

### **Linked clusters of GVs are detected in HAS1 minigenes (Table 3)**

Our sequencing analysis demonstrated that recurrent GVs are distributed as linked clusters, as distinct from random locations. The linked clusters of GVs described in Table 3 are distributed in the vicinity of splicing elements (see below). HAS1 minigene sequencing identified three distinct clusters of GVs that include three NCBI-SNPs and germline/hematopoietic origin GVs that are detected within a specific sequence stretch of intron 4 (1<sup>st</sup>“T” stretch, 2<sup>nd</sup>“T” stretch and TTTA repeats-shown in Figure 2b), near splicing elements at the 3' end of intron 4, where partial intron retention leads to HAS1Vb transcripts. These GVs and NCBI-SNPs were present in every cluster detected in MM and WM patients, and will be referred to as a “common motif”. The common motif appears to cluster with other recurrent HAS1 GVs (Figure 2b).

The GV clusters in WM were compared to those in MM, referred to as “WM clusters” or “MM clusters”, respectively, both including the common motif. Interestingly, GVs comprising MM GV cluster #3 and WM GV cluster #3 are identical with the exception of a germline origin mutation in MM exon 3 (CH56912051) instead of the hematopoietic/germline GV in intron 4 (CH56909252) for WM cluster #3. Two other HAS1 GV clusters from MM or WM patients include the common motif plus other recurrent GVs (Table 3). MM and WM clusters appear to harbor abnormalities that may accompany early stages of malignancy and/or characterize progression events.

### **Recurrent mutations detected in intron 4 promote HAS1 gene aberrant splicing**

Bioinformatic analysis predicted that HAS1 GVs lead to splicing events that generate HAS1Vb, the variant most significantly correlated with poor outcome (Figure 3a-e; Adamia, in preparation.)<sup>10</sup>. We constructed a HAS1 minigene splicing cassette derived from gDNA of a WM patient whose cells expressed HAS1Vb transcripts. This HAS1

cassette was used in an in vitro splicing assay to verify that HAS1 GV from WM could direct the splicing of HAS1Vb. The splicing construct incorporating the HAS1 gDNA segment from exon 3 to exon 5 was transfected into HeLa cells which do not otherwise express full length HAS1 or the aberrant HAS1 splice variants (Figure 3f). Sequencing analysis identified two unique substitutions in exon 3 and two in intron 4, as well as the common motif on HAS1 minigene cassettes. After transfection of HeLa cells with the HAS1 minigene cassettes we detected aberrantly spliced HAS1Vb transcripts, and normally spliced full length HAS1 (Figure 3f). Sequencing of the HAS1Vb PCR product confirmed its identity as HAS1Vb. This indicates that HeLa cells conserve the trans-splicing elements (small ribonuclear proteins) required for aberrant splicing of HAS1 pre-mRNA, but the HAS1Vb transcript is spliced only when cis-elements of the WM HAS1 template are introduced via the splicing construct.

## Discussion

Sequencing of the HAS1 gene in patients with MM and WM evaluated genetic contributions to the aberrant intronic splicing of HAS1 pre-mRNA that correlates with significantly reduced overall survival<sup>10</sup>. We report the existence of novel inherited (germline) and acquired (somatic) GVs in regions of HAS1 gene involved in aberrant splicing events. Somatic HAS1 GVs appear to accumulate throughout hematopoietic development, thereby leaving a mutational “trace” in non-malignant HPC and T cells as well as in malignant B and PC, with the largest number of GVs occurring in malignant MM/WM cells. In MM and WM patients, we identified three categories of genetic change: inherited germline origin, acquired hematopoietic origin and acquired tumor-specific GVs. All types included recurrent HAS1 GVs, defined by their detection in two or more MM and/or WM patients. The majority of germline GVs and all recurrent acquired HAS1 GVs were absent from the 23 control subjects.

Of the 197 novel GVs reported here, 50 were recurrent and 147 were unique. We anticipate that when larger cohorts are analyzed, some of the GVs currently classified as unique will prove to be recurrent. Recurrent somatic GVs are restricted to MM, restricted to WM or shared by both MM and WM; none are detected in the 23 control subjects. The

majority of recurrent GVs (both inherited and acquired) including three recurrent tumor-specific GVs, are shared between MM and WM. Although global gene expression profiling suggests that WM has more in common with B-CLL than with MM<sup>43</sup>, genomic analysis of HAS1 indicates that MM and WM, but not B-CLL, share a very close genetic relationship. The acquisition of recurrent somatic changes in the HAS1 gene, particularly in the non-coding intron 4, suggests that the HAS1 gene undergoes hypermutation and that strong selective pressures enrich these GVs in MM/WM.

Somatic mutations with the potential to alter splicing are frequent in some cancers<sup>28;36;44;45</sup>. In MM and WM patients we detected 87 somatic GVs, including 61 tumor-specific and 26 of hematopoietic-origin on HAS1 exons and introns (Table 2a). MM and WM are B-lineage cancers. In malignant B cells, aberrant somatic hypermutation affects genes outside of the immunoglobulin variable region<sup>46-51</sup>, which may occur prior to neoplastic transformation<sup>48</sup>. The frequency of hypermutated genes such as BCL-6 and PAX5 ranges from ~0.2-0.6/100bp<sup>46-51</sup>, consistent with the degree of hypermutation detected within the HAS1 gene in MM and WM (~0.1-0.5/100bp).

Somatic GVs were found in all subsets of differentiated hematopoietic cells (PC, PB-B and -T cells) and in purified CD34<sup>+</sup> HPC. Somatic mutations were found in G-CSF mobilized blood autografts or bone marrow from all eight patients from whom HPC were available. This implies that somatic HAS1 GVs may be acquired at the earliest stages of hematopoietic development. HPC harbouring somatic HAS1 GVs have normal generative capabilities and are clearly non-malignant based on the fact that their non-malignant T cell progeny carry the same somatic GVs. Our previous work shows that HAS1Vb transcripts are found in MM and WM cells, but are undetectable in non-malignant MM HPC or MM T cells, and normal B cells<sup>10;12</sup>. Since at birth HPC must, by definition, have the same genotype as BEC, our work suggests that in individuals who are destined for MM and WM, an undefined mechanism may enrich those HPC that have acquired somatic HAS1 GVs.

The pattern of germline GVs suggests that MM and WM, but not B-CLL, inherit recurrent germline GVs that are necessary but not sufficient for progression to malignancy. Acquisition of recurrent, somatic HAS1 GVs in otherwise healthy HPC from MM and WM patients appears to further increase the risk of MM or WM. This idea is supported by our demonstration that transfection of a HAS1 splicing construct from a WM patient directs the aberrant splicing of HAS1Vb (in vitro splicing assay). The B cell stage, at which acquisition of critical tumor-specific GVs occurs, may determine whether an individual develops MM or WM. Our work suggests that, similar to leukemias, genetic changes leading to myelomagenesis may first accumulate in HPC during the non-malignant or pre-malignant stages of hematopoietic differentiation<sup>52-54</sup>. The presence of tumor-specific recurrent GVs restricted to MM B/PC, coupled with expression of clinically predictive aberrant HAS1Vb transcripts by B cells<sup>10</sup>, supports the concept that transforming events may occur at the B cell stage<sup>42;55</sup>, and that malignant B cells transfer mutated HAS1 alleles to their MM PC progeny.

Provocatively, clusters of recurrent GVs are localized at sites of the HAS1 gene that control pre-mRNA splicing, particularly in intron 4, as shown in Figures 2b and 3. NCBI HAS1 SNPs appear to be significant predisposing elements in oncogenesis, as evidenced by their presence in combination with the “common motif” detected on the HAS1 gene. In silico analysis predicted that this “common motif” contributes to aberrant HAS1 splicing. To support the idea that HAS1 GVs play a role in HAS1 aberrant splicing, we conducted an in vitro splicing assay using a HAS1 splicing cassette genetically engineered from gDNA of a WM patient. The successful in vitro splicing of HAS1Vb verified bioinformatic predictions that the GVs (particularly aberrations in the “common motif” region of 1<sup>st</sup>, 2<sup>nd</sup> “T” and “TTTA” stretches of the HAS1 cassette) would direct splicing of HAS1Vb after transfection with the WM HAS1 cassette of host cells that do not otherwise express HAS1 or HAS1Vb transcripts.

The HAS1 GVs described here and the aberrant splicing reported previously<sup>10</sup> are likely to be important in malignancy. The aberrant HAS1 splice variants synthesize HA in ex-vivo MM cells<sup>10</sup> and in transfectants (Ghosh et al., submitted). In ex-vivo MM cells from

patients, HAS1Va correlates with synthesis of extracellular HA and HAS1Vb is associated with synthesis of intracellular HA<sup>10</sup>, suggesting a contribution to, respectively, malignant spread<sup>56</sup> and altered mitosis<sup>57</sup>. In HAS1 transfectants, normally spliced HAS1 (HAS1-FL) has a short half life and localizes to the plasma membrane while the HAS1 splice variants have a prolonged half life and localize to the cytoplasm (Ghosh et al., submitted). The aberrant splice variants form hetero-multimers with HAS1-FL and each other, thereby prolonging the half life of HAS1-FL and providing a potential mechanism for promoting cancer (Ghosh et al., submitted). Finally, as a single agent an intronic splice variant, HAS1Vc, is aggressively transforming in vitro and forms tumors in vivo (Ghosh et al., submitted). It seems likely that HAS1 family members will act synergistically amongst themselves and with other molecules, e.g. RHAMM<sup>57-60</sup>, to confer malignant characteristics. Work is in progress to evaluate this.

If verified in large scale studies, the recurrent HAS1 GVs identified here may have the potential to identify individuals at risk of MM/WM. Individuals with germline HAS1 GVs are predicted to have some degree of risk for MM/WM. Individuals who in addition to inheriting HAS1 germline GVs have acquired HAS1 GVs, may have a greatly increased risk of developing MM (or WM), perhaps requiring closer monitoring. The possibility exists that early detection of acquired HAS1 mutations may identify cryptic early stages MM or WM. Overall, this work shows that inherited and acquired HAS1 GVs may contribute to development of overt disease and/or disease progression in MM and WM by directing the aberrant intronic splicing of HAS1.

**Acknowledgements.** We thank Carolyn McQuarrie, Erin Strachan, Juanita Wizniak and Sarah Motz for skilled technical assistance, and Daniel Santos for his help in obtaining WM samples. This work was funded by grants from the Canadian Institutes of Health Research, the Alberta Cancer Research Institute, the International Waldenstrom's Macroglobulinemia Foundation and the Alberta Cancer Board Research Initiatives Program. SA was supported by an AHFMR Studentship. TR is an AHFMR Clinical Investigator. LMP is the Canada Research Chair in Biomedical Nanotechnology and this work was supported in part by the Chairs Program.

**Authorship.** SA conceived the study, designed and executed experiments, performed data analysis and interpretation and wrote the manuscript; AAR, HK, AG and JJH acquired and analyzed data; PMP assisted in data analysis; JK designed and executed *in vitro* splicing experiments; TR, MJM, SPT and ARB assisted in experimental design and writing the manuscript; LMP directed the research and wrote the manuscript.

None of the authors have conflicts of interest to disclose.

### Reference List

1. Buratti E, Baralle M, Baralle FE. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res* 2006;34:3494-3510.
2. Kalnina Z, Zayakin P, Silina K, Line A. Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes.Cancer* 2005;42:342-357.
3. Venables JP. Aberrant and alternative splicing in cancer. *Cancer Res* 2004;64:7647-7654.
4. Knudsen KE, Diehl JA, Haiman CA, Knudsen ES. Cyclin D1: polymorphism, aberrant splicing and cancer risk. *Oncogene* 2006;25:1620-1628.
5. Lamolle G, Marin M, varez-Valin F. Silent mutations in the gene encoding the p53 protein are preferentially located in conserved amino acid positions and splicing enhancers. *Mutat.Res* 2006;600:102-112.
6. Staalesen V, Falck J, Geisler S et al. Alternative splicing and mutation status of CHEK2 in stage III breast cancer. *Oncogene* 2004;23:8535-8544.
7. Moon SD, Park JH, Kim EM et al. A Novel IVS2-1G>A mutation causes aberrant splicing of the HRPT2 gene in a family with hyperparathyroidism-jaw tumor syndrome. *J Clin Endocrinol.Metab* 2005;90:878-883.
8. Pagenstecher C, Wehner M, Friedl W et al. Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. *Hum Genet.* 2006;119:9-22.
9. Lu F, Gladden AB, Diehl JA. An alternatively spliced cyclin D1 isoform, cyclin D1b, is a nuclear oncogene. *Cancer Res* 2003;63:7056-7061.
10. Adamia S, Reiman T, Crainie M et al. Intronic splicing of hyaluronan synthase 1 (HAS1): a biologically relevant indicator of poor outcome in multiple myeloma. *Blood* 2005;105:4836-4844.

11. Adamia S, Crainie M, Kriangkum J et al. Abnormal expression of hyaluronan synthases in patients with Waldenstrom's macroglobulinemia. *Semin Oncol* 2003;30:165-8.
12. Adamia S, Treon SP, Reiman T et al. Single nucleotide polymorphism of hyaluronan synthase 1 gene and aberrant splicing in Waldenstrom's macroglobulinemia. *Clinical Lymphoma* 2005;5:253-256.
13. Simpson MA, Reiland J, Burger SR et al. Hyaluronan synthase elevation in metastatic prostate carcinoma cells correlates with hyaluronan surface retention, a prerequisite for rapid adhesion to bone marrow endothelial cells. *J.Biol.Chem.* 2001;276:17949-17957.
14. Liu N, Gao F, Han Z et al. Hyaluronan synthase 3 overexpression promotes the growth of TSU prostate cancer cells. *Cancer Res* 2001;61:5207-14.
15. Itano N, Sawai T, Atsumi F et al. Selective expression and functional characteristics of three mammalian hyaluronan synthases in oncogenic malignant transformation. *J Biol Chem* 2004;279:18679-87.
16. Itano N, Kimata K. Altered hyaluronan biosynthesis in cancer progression. *Semin.Cancer Biol.* 2008;18:268-274.
17. Yamada Y, Itano N, Narimatsu H et al. Elevated transcript level of hyaluronan synthase1 gene correlates with poor prognosis of human colon cancer. *Clin.Exp.Metastasis* 2004;21:57-63.
18. Yabushita H, Noguchi M, Kishida T et al. Hyaluronan synthase expression in ovarian cancer. *Oncol.Rep.* 2004;12:739-743.
19. Bourguignon LY, Gilad E, Peyrollier K. Heregulin-mediated ErbB2-ERK signaling activates hyaluronan synthases leading to CD44-dependent ovarian tumor cell growth and migration. *J Biol.Chem.* 2007;282:19426-19441.
20. Golshani R, Hautmann SH, Estrella V et al. HAS1 expression in bladder cancer and its relation to urinary HA test. *Int.J Cancer* 2007;120:1712-1720.
21. Golshani R, Lopez L, Estrella V et al. Hyaluronic acid synthase-1 expression regulates bladder cancer growth, invasion, and angiogenesis through CD44. *Cancer Res* 2008;68:483-491.
22. Yabushita H, Kishida T, Fusano K et al. Role of hyaluronan and hyaluronan synthase in endometrial cancer. *Oncol.Rep.* 2005;13:1101-1105.
23. Carbone MA, Applegarth DA, Robinson BH. Intron retention and frameshift mutations result in severe pyruvate carboxylase deficiency in two male siblings. *Hum Mutat.* 2002;20:48-56.

24. Chen LL, Sabripour M, Wu EF et al. A mutation-created novel intra-exonic pre-mRNA splice site causes constitutive activation of KIT in human gastrointestinal stromal tumors. *Oncogene* 2005;24:4271-4280.
25. Eng L, Coutinho G, Nahas S et al. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum Mutat.* 2004;23:67-76.
26. Lazzereschi D, Nardi F, Turco A et al. A complex pattern of mutations and abnormal splicing of Smad4 is present in thyroid tumours. *Oncogene* 2005;24:5344-5354.
27. Liu HX, Cartegni L, Zhang MQ, Krainer AR. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet.* 2001;27:55-58.
28. Serra E, Ars E, Ravella A et al. Somatic NF1 mutational spectrum in benign neurofibromas: mRNA splice defects are common among point mutations. *Hum Genet.* 2001;108:416-429.
29. Rutter JL, Goldstein AM, Davila MR, Tucker MA, Struwing JP. CDKN2A point mutations D153spl(c.457G>T) and IVS2+1G>T result in aberrant splice products affecting both p16INK4a and p14ARF. *Oncogene* 2003;22:4444-4448.
30. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev.Genet.* 2002;3:285-298.
31. Maquat LE. Defects in RNA splicing and the consequence of shortened translational reading frames. *Am J Hum Genet.* 1996;59:279-286.
32. Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* 1999;96:307-310.
33. Valentine CR. The association of nonsense codons with exon skipping. *Mutat.Res* 1998;411:87-117.
34. Bromidge T, Lowe C, Prentice A, Johnson S. p53 intronic point mutation, aberrant splicing and telomeric associations in a case of B-chronic lymphocytic leukaemia. *Br.J.Haematol.* 2000;111:223-229.
35. Coutinho G, Xie J, Du L et al. Functional significance of a deep intronic mutation in the ATM gene and evidence for an alternative exon 28a. *Hum Mutat.* 2005;25:118-124.
36. Nichols KE, Houseknecht MD, Godmilow L et al. Sensitive multistep clinical molecular screening of 180 unrelated individuals with retinoblastoma detects 36 novel mutations in the RB1 gene. *Hum Mutat.* 2005;25:566-574.

37. Gamez-Pozo A, Palacios I, Kontic M et al. Pathogenic validation of unique germline intronic variants of RB1 in retinoblastoma patients using minigenes. *Hum Mutat.* 2007;28:1245.
38. Sharp A, Pichert G, Lucassen A, Eccles D. RNA analysis reveals splicing mutations and loss of expression defects in MLH1 and BRCA1. *Hum Mutat.* 2004;24:272.
39. Hansen TV, Bisgaard ML, Jonson L et al. Novel de novo BRCA2 mutation in a patient with a family history of breast cancer. *BMC.Med Genet.* 2008;9:58.
40. Agrawal S, Pilarski R, Eng C. Different splicing defects lead to differential effects downstream of the lipid and protein phosphatase activities of PTEN. *Hum Mol.Genet.* 2005;14:2459-2468.
41. Hull J, Shackleton S, Harris A. Abnormal mRNA splicing resulting from three different mutations in the CFTR gene. *Hum Mol.Genet.* 1993;2:689-692.
42. Szczepek AJ, Seeberger K, Wizniak J et al. A high frequency of circulating B cells share clonotypic IgH VDJ rearrangements with autologous bone marrow plasma cells in multiple myeloma, as measured by single cell and in situ RT-PCR. *Blood* 1998;92:2844-2855.
43. Chng WJ, Schop RF, Price-Troska T et al. Gene-expression profiling of Waldenstrom macroglobulinemia reveals a phenotype more similar to chronic lymphocytic leukemia than multiple myeloma. *Blood* 2006;108:2755-2763.
44. Pros E, Larriba S, Lopez E et al. NF1 mutation rather than individual genetic variability is the main determinant of the NF1-transcriptional profile of mutations affecting splicing. *Hum Mutat.* 2006;27:1104-1114.
45. Schaffner C, Stilgenbauer S, Rappold GA, Dohner H, Lichter P. Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. *Blood* 1999;94:748-753.
46. Pasqualucci L, Neri A, Baldini L, la-Favera R, Migliazza A. BCL-6 mutations are associated with immunoglobulin variable heavy chain mutations in B-cell chronic lymphocytic leukemia. *Cancer Res* 2000;60:5644-5648.
47. Deutsch AJ, Aigelsreiter A, Staber PB et al. MALT lymphoma and extranodal diffuse large B-cell lymphoma are targeted by aberrant somatic hypermutation. *Blood* 2007;109:3500-3504.
48. Malpeli G, Barbi S, Moore PS et al. Primary mediastinal B-cell lymphoma: hypermutation of the BCL6 gene targets motifs different from those in diffuse large B-cell and follicular lymphomas. *Haematologica* 2004;89:1091-1099.

49. Gaidano G, Capello D, Cilia AM et al. Genetic characterization of HHV-8/KSHV-positive primary effusion lymphoma reveals frequent mutations of BCL6: implications for disease pathogenesis and histogenesis. *Genes Chromosomes.Cancer* 1999;24:16-23.
50. Pasqualucci L, Neumeister P, Goossens T et al. Hypermethylation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 2001;412:341-346.
51. Montesinos-Rongen M, Van RD, Schaller C, Wiestler OD, Deckert M. Primary diffuse large B-cell lymphomas of the central nervous system are targeted by aberrant somatic hypermutation. *Blood* 2004;103:1869-1875.
52. George AA, Franklin J, Kerkof K et al. Detection of leukemic cells in the CD34(+)CD38(-) bone marrow progenitor population in children with acute lymphoblastic leukemia. *Blood* 2001;97:3925-3930.
53. Mauro MJ, Druker BJ. Chronic myelogenous leukemia. *Curr.Opin.Oncol.* 2001;13:3-7.
54. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* 1997;3:730-737.
55. Pilarski LM, Baigorri E, Mant MJ et al. Multiple myeloma includes CD20+ B and plasma cells that persist in patients treated with rituximab. *Clinical Medicine: Oncology* 2008;2:275-285.
56. Masellis-Smith A, Belch AR, Mant MJ, Turley EA, Pilarski LM. Hyaluronan-dependent motility of B cells and leukemic plasma cells in blood, but not of bone marrow plasma cells, in multiple myeloma: alternate use of receptor for hyaluronan-mediated motility (RHAMM) and CD44. *Blood* 1996;87:1891-9.
57. Pilarski LM, Adamia S, Maxwell CA et al. Hyaluronan Synthases and RHAMM as Synergistic Mediators of Malignancy in B Lineage Cancers. In: Balazs EA, Hascall VC, eds. HA2003 Proceedings. 2004:
58. Maxwell CA, Keats JJ, Crainie M et al. RHAMM is a centrosomal protein that interacts with dynein and maintains spindle pole stability. *Mol.Biol.Cell* 2003;14:2262-2276.
59. Maxwell CA, Rasmussen E, Zhan F et al. RHAMM expression and isoform balance predicts aggressive disease and poor survival in multiple myeloma. *Blood* 2004;104:1151-1158.
60. Maxwell CA, Reiman T, Ye M, Belch AR, Pilarski LM. RHAMM overexpression: A potential mechanism to generate extensive centrosomal abnormalities in multiple myeloma. *Blood* 2003;102:35a.

**Table 1. The primer sets used for PCR and/or sequencing reactions**

HAS1 Primers	Sequence 5'-3'
5' exon 3	GGGGTCTGTGCTGATCCTGG
3' exon 3	GCTTCCAGTTTTATCCCATC
5' intron 3	CTTCCACTGTGTATCCTGCATC
3' intron 3	AACTGCTGCAAGAGGTTATTC
5' exon 4	TGGGGTTGGAAGTGGAGATG
3' exon 4	CATGCACACACGCTAGGATA
5' intron 4a	GCTCAGCATGGGTTATGCTA
3' intron 4a	GTATCCCCGCAGCTTAAACA
5' intron 4b	TTGGGATAATCCAGGGGAAT
3' intron 4b	CAAGATGGGTGTGGTTGCTA
5' intron 4c	GGTAGCAACCACACCCATCT
3' intron 4c	AGGAATGAGGGCATCATCG
5' exon 5a	CTCGCCCCCGTGCAGGTACA
3' exon 5a	AGGCCCCCAAGCAGCAGCAGCGC
<i>5' exon 3-exon 4</i>	<i>ATGGGATAGGCTTGGAGTCA</i>
<i>3' exon 3-exon 4</i>	<i>CCCATCCAAAACCCACTGCA</i>
<i>5' exon 4-intron 4 (a)</i>	<i>GGAGACCAAGGTAGCACAGT</i>
<i>3' exon 4-intron 4 (a)</i>	<i>GTCTCTTGCCCTTCCTACTT</i>
<i>5' exon 4-intron 4 (b)</i>	<i>CAAGGGTGGTGGATAGGAAGTT</i>
<i>3' exon 4-intron 4 (b)</i>	<i>CCTCAGGCACTCCACTTAACAC</i>
<i>5' intron 4-exon 5 (a)</i>	<i>GGGCACGATCATGGCTCACT</i>
<i>3' exon 4-intron 4 (a)</i>	<i>GAGGTAGGGGGATCACTTGA</i>
<i>5' intron 4-exon 5 (b)</i>	<i>CCCCAGGGAGCACGCGATGA</i>
<i>3' exon 4-intron 4 (b)</i>	<i>GCACGGGGGCGAGGAATGAG</i>
5' minigene	CTTCCACCTTACAGGTCTGTGACT
3' minigene	CCACTCTGGTTCATGGTGACTA
5'intron 4c (T)	GCAAATACTATCTACTGGTCCTAAGC
3'intron 4c (T)	TGTGTAGAATAGGGTGGATAATGGT
M13R	CAGGAAACAGCTATGAC
T7	TAATACGACTCACTATAGGG

**Legend to Table 1:**

Primer set “intron 4c (T)” was used to verify “common motif” sequences —1<sup>st</sup> “T”, 2<sup>nd</sup> “T”, and “TTTA” stretches. In the table primer sequences presented in italics were used in sequencing reactions only.

**Table 2. Mutational distributions of GVs detected on HAS1 gene from MM and WM patients.**

**Table 2a. Distribution of HAS1 GVs on HAS1 exons and introns.**

Classification of GVs	HAS1 gene							
	Exon 3		Intron 3		Exon 4		Intron 4	
	U	R	U	R	U	R	U	R
<b>MM and WM Patients (n=17)</b>								
<b>Tumor specific</b>	8	2	6	1	0	0	40	4
<b>Hematopoietic origin</b>	1	2	0	1	0	2	15	5
<b>Hematopoietic/ Germline origin</b>	9	1	11	2	1	0	26	2
<b>Germline origin</b>	4	5	9	1	0	0	17	10
<b>NCBI-SNP</b>	0	1	0	3	0	0	0	8
<b>Total</b>	22	11	26	8	1	2	98	29
<b>Controls (n=23)</b>								
<b>Tumor specific</b>	0	0	0	0	0	0	1	0
<b>Hematopoietic origin</b>	0	0	0	0	0	0	0	0
<b>Hematopoietic/ Germline origin</b>	2	0	0	0	0	0	5	0
<b>Germline origin</b>	0	0	0	0	0	0	0	4
<b>NCBI-SNP</b>	0	1	0	3	0	0	0	8
<b>Total</b>	2	1	0	3	0	0	6	12

**Table 2b. Recurrent GVs and their distribution in MM and WM patients.**

	Chromosomal location	Nucleotide Change	AA changes	Type of GVs	Recurrence of GVs in MM and WM
<b>Tumor specific GVs</b>					
Exon 3	56912068	a>T	Tyr>Phe	#1TS	2WM
Exon 3	56912051	t>A	Cys>Ser	#2TS	1MM & 1WM
Intron 3	56911759	a>T		#3TS	1WM
Intron 3	56911759	a>G		#3TS	1WM
Intron 4	56909899	a>G		#4TS	2WM
Intron 4	56909573	del.-c		#5TS	1MM & 1WM
Intron 4	56909521	t>C		#6TS	1MM & 1WM
Intron 4	56909482	t>C		#7TS	2WM
<b>Hematopoietic origin GVs</b>					
Exon 3	56912079	t>C	Ala>Ala	#1HO	2MM & 1WM
Exon 3	56912077	g>A	Cys>Try	#2HO	1MM & 1WM
Intron 3	56911668	a>G		#3HO	2WM
Exon 4	56911348	g>C	Arg>Pro	#4HO	1MM & 3WM
Exon 4	56911346	a>T	Met>Leu	#5HO	1MM & 3WM
Intron 4	56910198	a>C		#6HO	2MM
Intron 4	56910041	g>T		#7HO	1MM & 4WM
Intron 4	56909423	t>C		#8HO	2MM
Intron 4	56909270	a>G		#9HO	1MM & 1WM
Intron 4	56909186	a>G		#10HO	2MM
<b>Germline origin</b>					
Exon 3	56912080	c>T	Ala>Val	#1GO	1MM & 1WM
Exon 3	56912058	t>C	Cys>Cys	#2GO	1MM & 1WM
Exon 3	56912056	t>C	Val>Ala	#3GO	3MM & 1WM
Exon 3	56912051	t>C	Cys>Arg	#4GO	3MM & 1WM
Exon 3	56912041	g>A	Gly>Asp	#5GO	1MM & 1WM
Intron 3	56911526	g>A		#6GO	2WM
Intron 4	56910856	t>C		#7GO	1MM & 1WM

Intron 4	56910811	a>G		#8GO	2WM
Intron 4	56910410	a>G		#9GO	2MM
Intron 4	56910219	a>G		#10GO	2MM
Intron 4	56909762	ins(Ts)		#11GO	6MM & 7WM
Intron 4	56909764	ins (Ts)		#12GO	3MM & 2WM
Intron 4	56909589	del or ins (Ts)		#13GO	7MM & 7WM
Intron 4	56909447	del or ins (TTTA)s		#14GO	6MM & 5WM
Intron 4	56909315	c>T		#15GO	4MM
Intron 4	56909217	c>T		#16GO	4MM & 1WM
<b>Hematopoietic/Germline origin</b>					
Exon 3	56912197	a>G	Asp>Gly	#1H/G	1MM & 1WM
Intron 3	56911983	a>G		#2H/G	2MM
Intron 3	56911977	a>G		#3H/G	1MM & 1WM
Intron 4	56910790	a>C		#4H/G	2WM
Intron 4	56909253	g>A		#5H/G	2WM
<b>NCBI-SNPs</b>					
Exon 3	56912163	c>T	Asp>Asp	rs 11084111	7MM & 2WM
Intron 3	56911889	g>A		rs 11084110	8MM & 2WM
Intron 3	56911831	g>A		rs 11084109	8MM & 7WM
Intron 3	56911750	t>A		rs 11669079	8MM & 7WM
Intron 4	56910770	g>C		rs 11667974	3MM & 1WM
Intron 4	56910738	t>G		rs 11667949	8MM & 7WM
Intron 4	56910711	g>A		rs 7254072	8MM & 7WM
Intron 4	56910493	g>C		rs 4802850	4MM & 3WM
Intron 4	56910155	c>A		rs 4802849	5MM & 4WM
Intron 4	56910154	c>G		rs 4802848	5MM & 4WM
Intron 4	56909763	c>T		rs 8104157	8MM & 7WM
Intron 4	56909604	ins T		rs 11438660	8MM & 7WM

**Table 2c. Number of individuals in whom recurrent HAS1 GVs were detected.**

Type	Tumor	Hematopoietic	Germline	Germline/ Hematopoietic	NCBI- SNP
<b>MM and WM Patients (n=17)</b>					
<b>% of patients with the indicated type of HAS1 GVs</b>	<b>59% (10/17)</b>	<b>65% (11/17)</b>	<b>94% (16/17)</b>	<b>41% (7/17)</b>	<b>88% (15/17)</b>
<b>Number of patients in whom HAS1 GVs were recurrent</b>	<b>2-3</b>	<b>2 - 5</b>	<b>2- 14</b>	<b>2-4</b>	<b>4-15</b>
<b>Controls (n=23)</b>					
<b>% of controls with recurrent HAS1 GVs</b>	<b>0</b>	<b>0</b>	<b>4-39%</b>	<b>0</b>	<b>39- 83%</b>
<b>Number of controls in whom HAS1 GVs were recurrent</b>	<b>0</b>	<b>0</b>	<b>1-6</b>	<b>0</b>	<b>2-14</b>

**Legend to Table 2:**

a) This table represents the distribution of unique (U, gray rows) and recurrent (R, clear rows) of GVs on HAS1 gene exon 3-4 and introns 3-4 from MM and WM patients. We have detected total 196 unique and recurrent GVs in 17 MM and WM patients. Among these GVs 61 are tumor specific, 26 Hematopoietic origin, 52 hematopoietic/germline origin, 46 germline origin, and 11 NCBI SNPs. Some of the GVs detected in hematopoietic cells were classified as hematopoietic/germline origin because no BEC were available for these patients to confirm germline or hematopoietic origin. We also detected sporadic substitutions, each in only one subclone from one individual HD, that were not detected in MM or WM patients and are not reported in NCBI database.

b) This table includes details of recurrent GVs that were detected in MM and WM, in exons and introns. “TS” —tumor specific GVs, “HO”—Hematopoietic GVs, “GO”—germline origin GVs, H/G –hematopoietic/germline GVs. GVs are numbered according to their position on exons and introns. AA—amino acid. Ins = insertion. del = deletion. The numbers in column titled “Recurrence of GVs in MM and WM” represent the number of MM and/or WM patients in whom the GV was recurrent. A total of 49 chromosomal positions in HAS1 harbored GVs. For one position, 56911759, two different mutated alleles were detected, yielding aggregate of 50 recurrent GVs.

c) This table represents the % of patients or control subjects having a given type of GV and the numerical range of individuals in whom the indicated types of GVs were recurrent. A patient or control was counted as having the indicated type of GV if at least one such GV was detected in the samples that were sequenced.

Table 3. Clusters of GVs detected in HAS1 gene exons and introns from MM and WM patients.

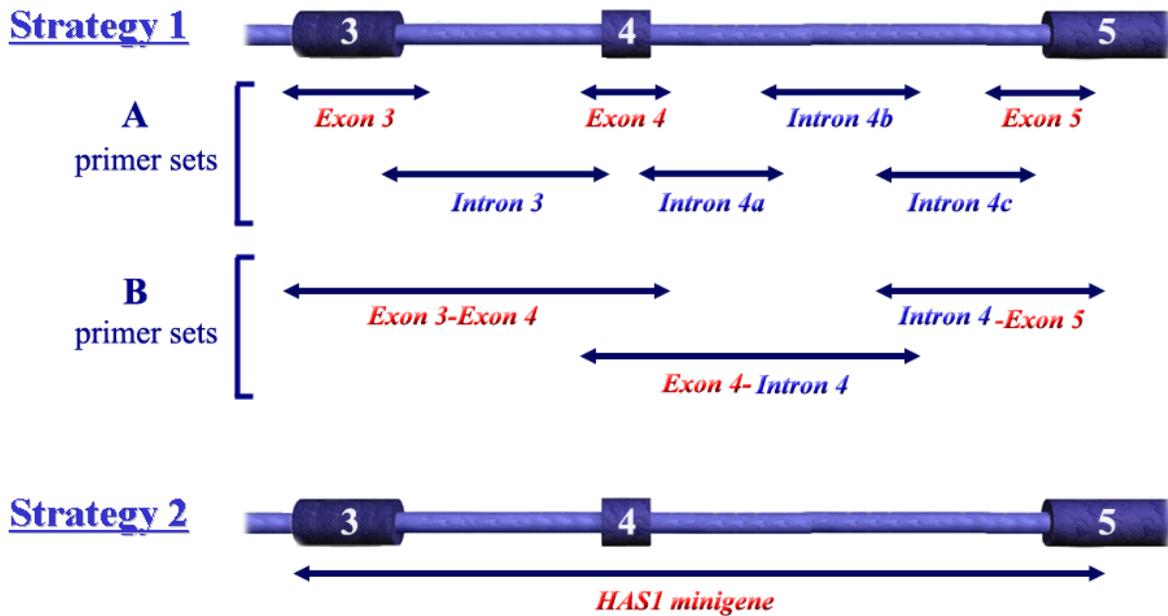
	Nucleotide changes	Chromosomal location	Type	Effects on the protein	Cluster
<b>GVs clusters detected in MM patients</b>					
Exon 3	c>T	56912163	NCBI rs 1108411		#1, 2, 3
Exon 3	t>C	56912056	Germline origin	Val>Ala	#1
Intron 4	g>C	56910770	NCBI rs 11667974		#1
Intron 4	c>T	56109315	Germline origin		#2
Intron 4	c>T	56109217	Germline origin		#2
Intron 4	inst T	56909764	Germline origin		#2
Exon 3	t>C	56912051	Germline origin	Cys>Arg	#3
Intron 4	g>C	56910493	NCBI rs 4802850		#3
Intron 4	c>A	56910155	NCBI rs 4802849		#3
Intron 4	c>G	56910154	NCBI rs 4802848		#3
<b>GVs clusters detected in WM patients</b>					
Exon 3	a>T	56912068	Tumor specific	Tyr>Phe	#1
Intron 3	a>G	56911668	Tumor specific		#2
Exon 4	a>T	56911346	Hematopoietic origin	Met>Leu	#2
Exon 4	g>C	56911348	Hematopoietic origin	Arg>Pro	#2
Intron 4	g>t	56910041	Hematopoietic origin		#2
Intron 4	t>C	56909252	Hematopoietic/germline origin		#3
Intron 4	c>G	56910154	SNP-NCBI rs 4802848		#3
Intron 4	c>A	56910155	SNP-NCBI rs 4802849		#3
Intron 4	g>C	56910493	SNP-NCBI rs 4802850		#3
<b>Common motifs</b>					
Intron 3	t>A	56911750	SNP-NCBI rs 11669079		
Intron 3	g>A	56911831	SNP-NCBI rs 11084109		
Intron 3	g>A	56911889	SNP-NCBI rs 11084110		
Intron 4	inst T	56909604	SNP-NCBI rs 11438660		
Intron 4	inst (TTTA)s	56909447	Germline origin		
Intron 4	del (TTTA)s	56909447	Germline origin		
Intron 4	inst/del (T)s	56909589	Germline origin		
Intron 4	inst (Ts)	56909762	Germline origin		
Intron 4	c>T	56909763	SNP-NCBI rs 8104157		
Intron 4	g>A	56910711	SNP-NCBI rs 7254072		
Intron 4	t>G	56910738	SNP-NCBI rs 11667949		

Legend to Table 3:

The first GV cluster of MM includes the common motif plus a recurrent hematopoietic origin missense mutation t>C (CH56912056) detected on exon 3 and NCBI-SNP rs 1667974. The second MM GV cluster is comprised of the common motif and two additional germline GVs, both on intron 4. The third MM GV cluster includes common motif with an additional one germline origin missense t>C substitution (Cys>Arg) on exon 3 and three NCBI-SNPs on intron 4. All MM patients were homozygous for mutated alleles of NCBI-SNPs included in the third GV cluster.

WM GV cluster #1 includes the common motif plus a recurrent tumor specific missense mutation a>T (Tyr>Phe, CH6912068) detected on exon 3. The second WM GV cluster is comprised of the common motif and four additional hematopoietic origin GVs, missense mutations, a>T (CH56911346) and g>C (CH56911348) in exon 4, that lead to amino acid changes Met>Leu and Arg >Pro, respectively, one tumor specific in intron 4 (CH56910041) and one recurrent tumor specific transition (CH5611668) in intron 3. The third WM GV cluster includes common motif with additional 4 GVs all detected in intron 4, one hematopoietic origin and three NCBI-SNPs. Similar to MM patients all WM patients were homozygous for mutated alleles of NCBI-SNPs included in any clusters.

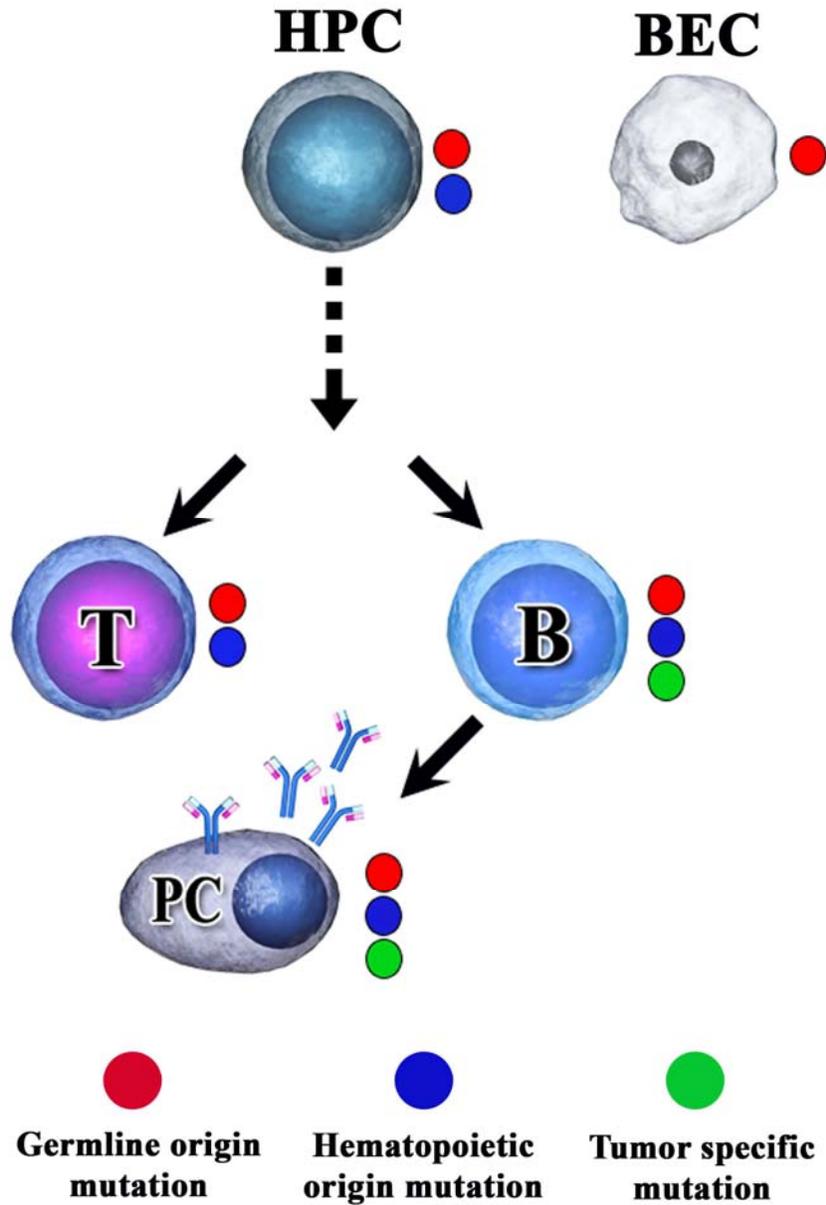
**Figure 1. Strategies for sequencing the HAS1 gene segments.**



This Figure describes two strategies used to amplify the HAS1 gene segments. Overlapping reverse and forward primers were designed to anneal with exons and introns of the HAS1 gene to identify genetic variations that may contribute to aberrant HAS1 splicing in WM patients. Minigene sequencing in strategy 2 was used to determine whether the recurrent mutations were clustered.

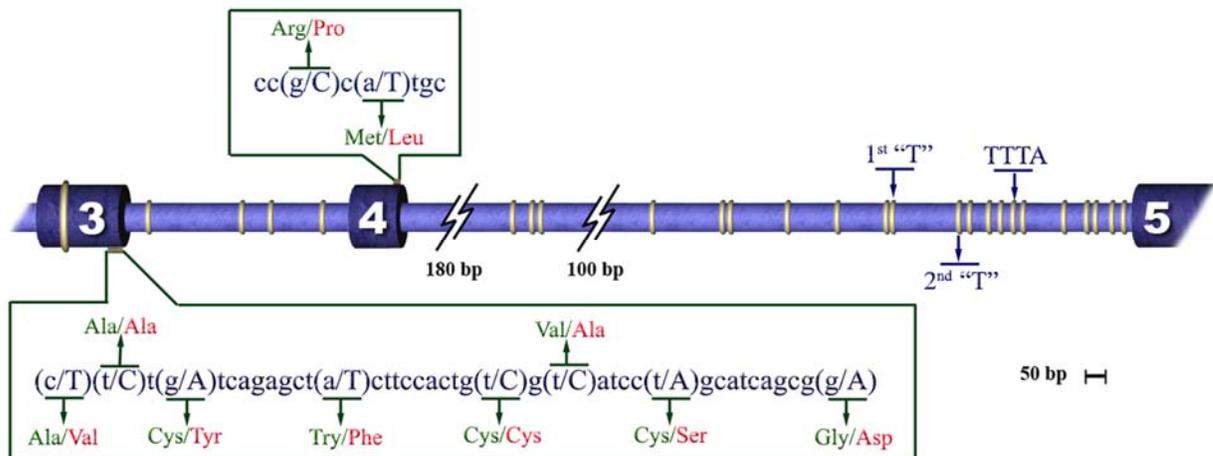
gDNA PCR and cloning were carried out as described in methods. We picked 24 to 48 subclones to screen and sequence inserts (gDNA PCR product of the HAS1 gene segment) in the TOPO TA plasmid using the appropriate primer sets for each segment. While using the first strategy of amplifying HAS1 gene segments from the patients or HDs, we cloned 7 or 3 segments from the exon 3 to exon 5 region of the HAS1 gene using A or B primer sets, respectively. For each segment, 3-10 positive subclones were selected and for each cell subset, more than 50 plasmids were isolated and sequenced both directions using M13 and T7 sequencing primers. Using Strategy 2, we cloned 30 HAS1 minigene plasmids from MM and WM, and 33 minigenes from B-CLL and MGUS. B-CLL minigenes encompassed only intron 4. MGUS minigenes encompassed exon 3 to exon 5. Each plasmid was sequenced using overlapping HAS1 gene specific A and B primer sets. Because we used overlapping primers either in gDNA PCR (Strategy 1) or for sequencing (Strategy 2), we analyzed 50-60 sequencing reactions for exon-intron spanning segments of the HAS1 gene. The HAS1 gene segments of two HDs (B and T cells) were sequenced using Strategy 1. The genetic variations identified in WM patients were assessed based on a total of 4119 sequencing reactions.

**Figure 2a. Cell type distribution of HAS1 GVs detected in MM and WM patients.**



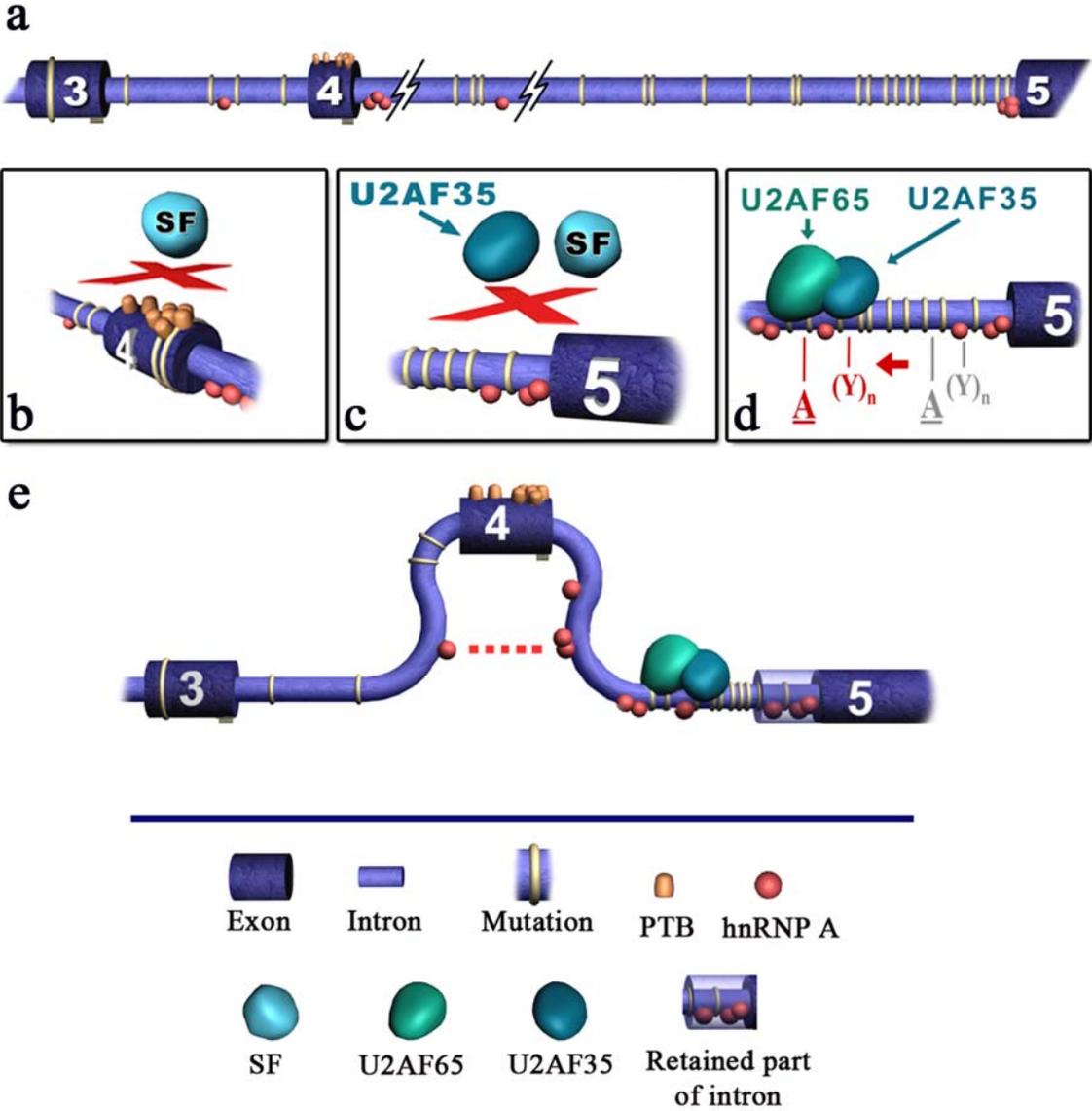
Mutations identified in various types of cells from MM patients were classified as tumor specific, hematopoietic and germline origin based on their occurrence in these cells. HPC—CD34<sup>+</sup>45<sup>low</sup> HPCs from mobilized blood of MM patients or bone marrow aspirates from WM patients; BEC—Buccal epithelial cell.

**Figure 2b. Distribution of GVs in genomic HAS1.**



This figure demonstrates relative distribution of GVs detected in MM and WM patients. Recurrent NCBI-SNPs are absent from this figure. The inserts detail sets of GVs located at the boundary of exon 3 and exon 4 respectively. GVs are represented by yellow rings. On the Figure, the 1<sup>st</sup> break on intron 4 represents 180 nucleotides, while the 2<sup>nd</sup> break represents 100 nucleotides. 1<sup>st</sup> "T", 2<sup>nd</sup> "T", and "TTTA" are the "common motif" detected in MM and WM patients. The spaces between mutations are arranged according to a scale of 50 bp=4mm.

**Figure 3. Bioinformatic Model - Clusters of recurrent GV mutations facilitate aberrant splicing of HAS1 gene in MM patients to create the intronic HAS1Vb splice variants.**



In this analysis we used the web based bioinformatic tool ESE finder V2. Results were evaluated using ESE V3 (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi>). For more detailed analysis we used ASD (The Alternative Splicing Database— workbench bioinformatics tools). Using these tools we evaluated the distribution of splicing elements in HAS1 exons 3, 4 and introns 3, 4 of wild type and mutated sequences. (a) demonstrates relative distribution of recurrent mutations detected in MM and WM patients and shows accumulation two important splicing co-factors, hnRNP I (PTB) and hnTNP A, in exon 4 and introns 3 and 4. (b), (c), and (d) represent the location on the HAS1 gene where the aberrations occur. (e) predicts the impact of recurrent GVs on HAS1 splicing. In (d), the red letters “A” and “Y” represent activated splicing branch point (BP) and polypyrimidine tract (PPT) of splicing respectively, while gray letters: “A” and “Y” represent native BP and PPT.

#### Description of the Model

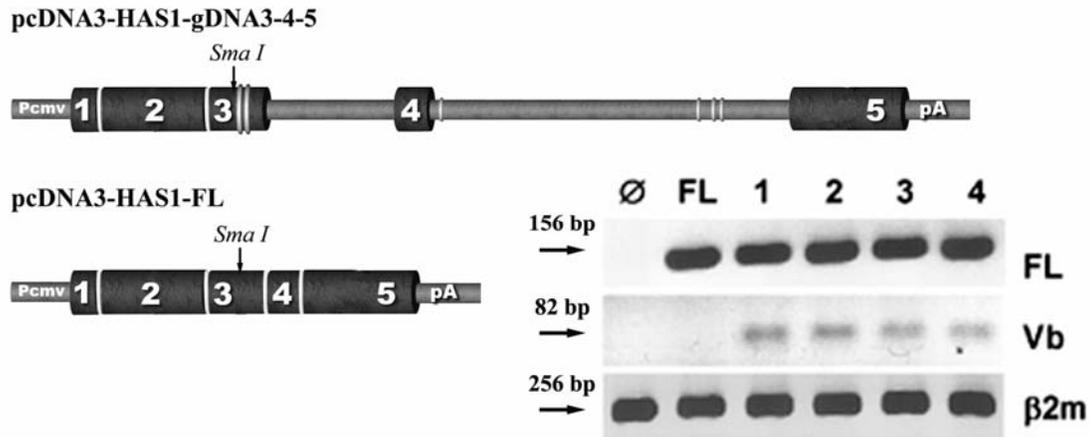
No differences were found between wild type and mutated exon 3 with respect to the accumulation of hnRNPs which bind mainly splicing suppressors and promote exon exclusion. However, in mutated exon 4, as compared to wild type exon 4 and in mutated exon 3, bioinformatic analysis predicted a massive accumulation of hnRNPs, including hnRNP I (PTB- polypyrimidine tract Binding protein), which is distributed across the entire mutated exon 4 (a, b). As suggested in the diagram, the binding of PTBs at several sites of an exon could cause a loopout of this exon, and subsequently these types of exons become inaccessible for the assembly of the spliceosome (b, c, e).

The analysis did not predict any significant differences between wild type and mutated intron 3 with respect to Serin/Arginine rich proteins (SRs) or the distribution of hnRNP binding motifs. Also, no significant difference was found when BP and PPT were mapped on wild type and mutated intron 3. However, for mutated intron 4, the existence of alternative splicing branch points were predicted. These alternative BPs are located upstream of the alternative PPT (d). Additionally, splicing element analysis of wild type and mutated intron 4 demonstrated an accumulation of a significant number of SR and hnRNP binding motifs in mutated intron 4. Among them, the most significant predicted

difference that contributes to intronic splicing of HAS1 is recruitment of U2AF65 protein by the alternative PPTs (d). These predicted PPT sequences overlap with the 1<sup>st</sup> and 2<sup>nd</sup> “T” stretches and TTTA repeats of mutated intron 4 (the “common motif”) where the MM clusters of GVs are located. The protein U2SF65 is known to be responsible for the recruitment of SFs to splicing BP. Subsequently, this protein acts as a “bridge” between BP and PPT and stabilizes the spliceosomal complex necessary for the first stage of the splicing reaction.

In addition, our analysis of wild type and mutated intron 4 predicted the loss of a significant number of binding motifs for hnRNP proteins from mutated intron 4 as compared to wild type. However, mutated intron 4 maintained ability to recruit hnRNP, a protein which most likely contributes to the exclusion of exon 4 through its ability to dimerise with other molecules of hnRNP A located within and on adjacent introns (e).

**Figure 3f. Splicing of aberrant HAS1 Vb transcripts in transfected HeLa cells.**



The diagram illustrates the expression cassettes for HAS1 minigene constructs. HAS1 sequences were flanked by a mammalian CMV promoter at the 5' end of HAS1 gene and the bovine growth hormone polyadenylation signal, poly A, at the 3' end. mRNA splicing was analyzed by transfecting HeLa cells with HAS1 minigene cassettes. RT-PCR was performed 24 hrs post-transfection using specific primers for HAS1 full length (FL), HAS1Vb or B<sub>2</sub>m (beta 2 microglobulin).

On the gel: Ø—result obtained from the cells transfected with cassette without HAS1 gene; FL—the result obtained from the cells transfected with pcDNA3-HAS1-FL cDNA construct, which is already spliced; Lanes 1-4 represent HeLa cells transfected with pcDNA3-HAS1-g3-4-5 construct. We tested four subclones of the HAS1 minigene cassette; transfection of all four subclones gave identical results. Product identity was confirmed by sequencing. For this experiment, Ø and FL used as controls to verify specificity of the *in vitro* splicing assay.