

# Genomic Organization, Physical Mapping, and Expression Analysis of the Human Protein Arginine Methyltransferase 1 Gene

Andreas Scorilas,<sup>\*,†</sup> Margot H. Black,<sup>\*</sup> Maroulio Talieri,<sup>‡</sup> and Eleftherios P. Diamandis<sup>\*,†,1</sup>

<sup>\*</sup>Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Canada; <sup>†</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada; and <sup>‡</sup>"G. Papanikolaou" Research Center of Oncology, "St. Savas" Hospital, 171 Alexandras Avenue, Athens, 115 22, Greece

Received October 13, 2000

**Protein arginine methyltransferases (PRMTs) regulate mRNA processing and maturation by modulating the activity of RNA-binding proteins through methylation. The cDNA for human PRMT1 (HRMT1L2) was recently identified. In this paper, we describe the complete genomic organization of the human PRMT1 gene (GenBank Accession No. AF222689), together with its precise chromosomal localization in relation to other neighboring genes. We have also examined its expression in a total RNA panel of 26 human tissues, the BT-474 breast carcinoma cell line, and 16 breast tumors. PRMT1, which spans 11.2 kb of genomic sequence on chromosome 19q13.3, is located in close proximity to the IRF3 and RRAS genes and is transcribed in the opposite direction. It is formed of 12 coding exons and 11 intervening introns, and shows structural similarity to other PRMT genes. Three PRMT1 isoforms exist as a result of alternative mRNA splicing. Amino acid sequence comparison of the splicing variants indicates that they are all enzymatically active methyl transferases, but with different N-terminal hydrophobic regions. PRMT1 expression was detected in a variety of tissues. We have shown that the relative prevalence of alternatively spliced forms of PRMT1 is different between normal and cancerous breast tissues. Although PRMT1 was not found**

**to be hormonally regulated by steroid hormones in breast cancer cells, our results suggest that two variants of PRMT1 are down regulated in breast cancer.** © 2000 Academic Press

**Key Words:** arginine methyltransferase; PRMT1; breast cancer; HRMT1L2; gene mapping; gene characterization; RRAS; IFR3; BT-474; SH3 domains.

The molecular mechanisms governing the transition from gene to protein have yet to be fully elucidated. Primary eukaryotic gene transcripts (pre-mRNAs) undergo a series of processing reactions to form functional, mature mRNA. RNA processing takes place in the nucleus through the action of a group of nuclear proteins, collectively referred to as the heterogeneous nuclear ribonucleoproteins (hnRNPs). hnRNPs play a regulatory role in mRNA maturation, stability, and export to the cytoplasm (1). hnRNP activity, in turn, is governed by posttranslational modifications such as glycosylation (2), phosphorylation (3), and methylation (4), which cooperatively regulate hnRNP binding and dissociation from RNA. Of particular interest is the protein arginine methylation event which produces  $N^G, N^G$ -dimethylarginine, an amino acid present in the majority of hnRNPs. hnRNPs contain ~65% of the total cellular  $N^G, N^G$ -dimethylarginine (5), indicating that this unusual modification plays an important role in regulating hnRNP activity.

The enzyme responsible for this methylation reaction is protein arginine methyltransferase (PRMT). PRMTs specifically methylate protein substrates that contain an RRG binding domain, a motif involved in RNA:RNA and RNA:protein interactions which consists of repeated glycine dipeptides interspersed with arginines and aromatic residues. PRMT1 methylates hnRNPs which contain an RRG box, such as hnRNP A1, in addition to other RNA-binding proteins with an

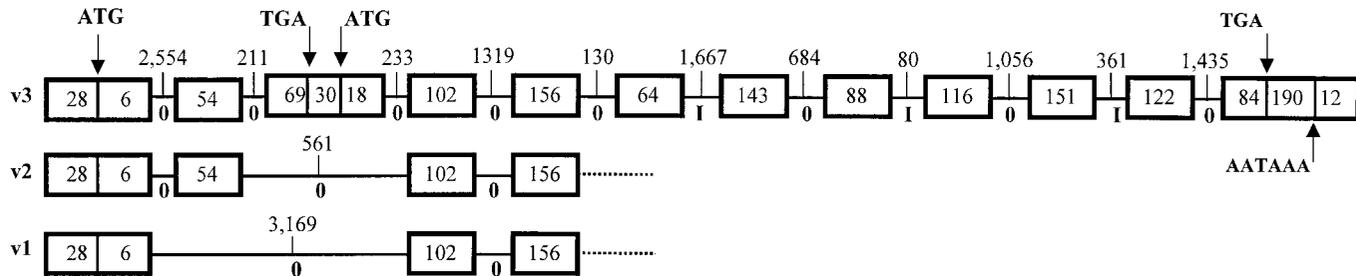
Abbreviations used: PRMT, protein arginine methyltransferase; PRMT1-v, protein arginine methyltransferase 1 splice variant; hnRNP, heterogeneous nuclear ribonucleoprotein; IRF3, interferon regulatory factor 3; PSA, prostate specific antigen; TIS21, 12-*O*-tetradecanoylphorbol-13-acetate (TPA)-inducible sequence gene 21; IFN, interferon; HRMT1L2, synonym to PRMT1 gene; HRMT1L1, synonym to PRMT2 gene; LLNL, Lawrence Livermore National Laboratory; BCM, Baylor College of Medicine; EST, expressed sequence tag; RT-PCR, reverse transcription-polymerase chain reaction; bp, base pair.

<sup>1</sup>To whom correspondence and reprints requests should be addressed at Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5. Fax: (416) 586-8628. E-mail: ediamandis@mtsinai.on.ca.



agcgggggtcgcggtagcgccggccgtggaccctctggtataagcgggtccccggggagtgaggagaaggggggtcttggcgccgg  
 aggaggagtaggtcgccggtgaagatGGCGGCAGCCGAGGCCGCGAACTGCATC **ATG** GAG **gtgagcgtt...**  
 ...*intron-1* .....tctctcctag **AAT TTT GTA GCC ACC TTG GCT AAT GGG ATG AGC CTC CAG**  
   N F V A T L A N G M S L Q  
**CCG CCT CTT GAA GAA** **glaaatatcc...*intron-2*...ctccacctcc **AAC CCC CCT TTG CCC TTC**  
                   P P L E E  
**CCT GTG TCT GCC CCC ATT TTC CTT CCC CTC CCC TCC CCA GCT GTG GGC** **TGA** **GCT**  
**AGA GAC GGG GTC AGA GAG ACT GGA GAG** **ATG** **GTA GGC GTG GCT GAG**  
   M V G V A E  
**gtatcgggggt...*intron-3*...cccctccc **GTG TCC TGT GGC CAG GCG GAA AGC AGT GAG AAG**  
   V S C G Q A E S S E K  
**CCC AAC GCT GAG GAC ATG ACA TCC AAA GAT TAC TAC TTT GAC TCC TAC GCA**  
                   P N A E D M T S K D Y Y F D S Y A  
**CAC TTT GGC ATC CAC GAG** **gtcagtgggg...*intron 4*...cttgccctag **GAG ATG CTG AAG GAC**  
                   H F G I H E  E M L K D  
**GAG GTG CGC ACC CTC ACT TAC CGC AAC TCC ATG TTT CAT AAC CGG CAC CTC TTC**  
                   E V R T L T Y R N S M F H N R H L F  
**AAG GAC AAG GTG CTG GAC GTG GGC TCG GGC ACC GGC ATC CTC TGC ATG**  
                   K D K V V L D V G S G T G I L C M  
**TTT GCT GCC AAG GCC GGG GCC CGC AAG GTC ATC GGG** **gtgagctcc...*intron-*  
                   F A A K A G A R K V I G  
**5...gggccctcag** **ATC GAG TGT TCC AGT ATC TCT GAT TAT GCG GTG AAG ATC GTC AAA**  
                   I E C S S I S D Y A V K I V K  
**GCC AAC AAG TTA GAC CAC G** **gtgagccag...*intron-6*...ccgcccgcag **TG GTG ACC ATC ATC**  
                   A N K L D H  V V T I I  
**AAG GGG AAG GTG GAG GAG GTG GAG CTC CCA GTG GAG AAG GTG GAC ATC ATC**  
                   K G K V E E V E L P V E K V D I I  
**ATC AGC GAG TGG ATG GGC TAC TGC CTC TTC TAC GAG TCC ATG CTC AAC ACC**  
                   I S E W M G Y C L F Y E S M L N T  
**GTG CTC TAT GCC CGG GAC AAG TGG CTG** **gtgagccccc...*intron-7*...ccctctgtag **GCG CCC**  
                   V L Y A R D K W L  A P  
**GAT GGC CTC ATC TTC CCA GAC CGG GCC ACG CTG TAT GTG ACG GCC ATC GAG**  
                   D G L I F P D R A T L Y V T A I E  
**GAC CGG CAG TAC AAA GAC TAC AAG ATC CAC T** **gtgagcgcgg...*intron-8*...gcctcccag **GG**  
                   D R Q Y K D Y K I H  W  
**TGG GAG AAC GTG TAT GGC TTC GAC ATG TCT TGC ATC AAA GAT GTG GCC ATT**  
                   W E N V Y G F D M S C I K D V A I  
**AAG GAG CCC CTA GTG GAT GTC GTG GAC CCC AAA CAG CTG GTC ACC AAC GCC**  
                   K E P L V D V V D P K Q L V T N A  
**TGC CTC ATA AAG** **gtgaggggt...*intron-9*...igtctgcag **GAG GTG GAC ATC TAT ACC GTC**  
                   C L I K  E V D I Y T V  
**AAG GTG GAA GAC CTG ACC TTC ACC TCC CCG TTC TGC CTG CAA GTG AAG CGG**  
                   K V E D L T F T S P F C L Q V K R  
**AAT GAC TAC GTG CAC GCC CTG GTG GCC TAC TTC AAC ATC GAG TTC ACA CGC**  
                   N D Y V H A L V A Y F N I E F T R  
**TGC CAC AAG AGG ACC GGC TTC TCC ACC A** **gtgagccggg...*intron-10*...gcgccccag **GC**  
                   C H K R T G F S T  S  
**CCC GAG TCC CCG TAC ACG CAC TGG AAG CAG ACG GTG TTC TAC ATG GAG GAC**  
                   P E S P Y T H W K Q T V F Y M E D  
**TAC CTG ACC GTG AAG ACG GGC GAG GAG ATC TTC GGC ACC ATC GGC ATG CGG**  
                   Y L T V K T G E E I F G T I G M R  
**CCC AAC GCC AAG AAC AAC** **gtgaggctcc...*intron-11*...ctccctgcag **CGG GAC CTG GAC TTC**  
                   P N A K N N  R D L D F  
**ACC ATC GAC CTG GAC TTC AAG GGC CAG CTG TGC GAG CTG TCC TGC TCC ACC**  
                   T I D L D F K G Q L C E L S C S T  
**GAC TAC CGG ATG CGC** **TGA** **GGCCCCGCTCTCCCGCCCTGCACGAGCCCAGGGGCTGAGC**  
                   D Y R M R  
**GTTCCCTAGGCGGTTTCGGGGCTCCCCCTCCTCCTCCCTCCCTCCCGCAGAAGGGGGTTTTAGGGG**  
**CCGGGCTGGGGGGATGGGGAGGGCACATCGTGACTGTGTTTTTCATAACTTATGTTTTTATAT**  
**GGTTGCATTTAGCCAATAAATCCTCAGCTGGGgtctgcttctgctgggggcaaaggaggttgggttctcttcg**  
**caggtactgcagtgccagcctcagcagtggtgggggaccggc**********************

**FIG. 1.** Genomic organization and partial genomic sequence of the PRMT1/HRMT1L2 gene. Intronic sequences are not shown except for short sequences around the splice junctions. Introns are shown with lower case letters and exons with capital letters. For full sequence, see GenBank Accession No. AF222689. The start and stop codons are encircled and the exon-intron junctions are in black boxes. The codons are in bold. Exons 2 and 3 that are responsible for the splicing variants of the PRMT1 gene are single and double underlined, respectively. A typical polyadenylation signal (AATAAA), located 181 bp downstream of the stop codon (TGA) is shown in bold. The in-frame stop codon in exon 2 is within box. One splice junction at the end of intron 2 is not conserved. For more details see text.



**FIG. 2.** Schematic diagram showing the comparison of the genomic structure of the three splicing variants of the PRMT1 gene (v1, v2, v3). Exons are shown as boxes and introns by the connecting lines. Arrows show the positions of the start codons, stop codons, and polyadenylation signal. Roman numerals indicate intron phases. The intron phase refers to the location of the intron within the codon; I denotes that the intron occurs after the first nucleotide of the codon, II that the intron occurs after the second nucleotide, and 0 that the intron occurs between codons. Numbers inside boxes indicate exon length in base pairs, and the vertical connecting lines show the intron lengths. Figure is not drawn to scale.

RRG binding motif, such as nucleolin, fibrillarin, and ribosomal protein S2 (5, 6). Arginine methylation is a cellular process which occurs over a wide range of eukaryotes, and PRMTs have been studied in several eukaryotic species. This enzyme was first isolated in 1988 from calf brain (7). Two homologous proteins, PRMT1 and PRMT2 (also called HMT1 and RMT2) and the corresponding genes were later identified in yeast (8, 9). Many recent studies have been performed on rodent and human PRMTs. Human PRMT1, also called HRMT1L2 or IR1B4, (10, 11) is 96% identical at the amino acid level to rat PRMT1 (6). Human PRMT2, also called HRMT1L1, is 33% identical in amino acid sequence to human PRMT1 (10, 12). Human PRMT3, was also recently cloned (13).

In this study, we describe the complete genomic organization and chromosomal location of PRMT1. We have examined the differential expression of the three PRMT1 splicing variants in a large number of normal human tissues, the BT-474 breast carcinoma cell line, and breast tumors. Our preliminary data suggest that two variants of PRMT1 are down regulated in breast cancer.

## MATERIALS AND METHODS

**Chromosome 19 sequence analysis.** Genomic sequences for chromosome 19 are available from the Lawrence Livermore National Laboratory database. We examined ~200 kb of nucleotide sequence (clone BC 42053), in the form of 90 separate contigs of variable lengths, spanning chromosome 19q13.3–q13.4. The chromosome 19 *EcoRI* restriction map (14) and long PCR strategies, were used to construct a contiguous area of the genomic area of interest. Bioinformatic approaches were used, as previously described (15–19), to predict the presence of new genes and a putative arginine methyltransferase was identified. Blast search (20) and *EcoRI* restriction digestion analysis of the sequence of one adjacent BAC clone (R31181) allowed us to identify the relative position of this methyltransferase gene and other previously identified genes; RRAS and IRF3, along the same chromosomal region (19q13.3).

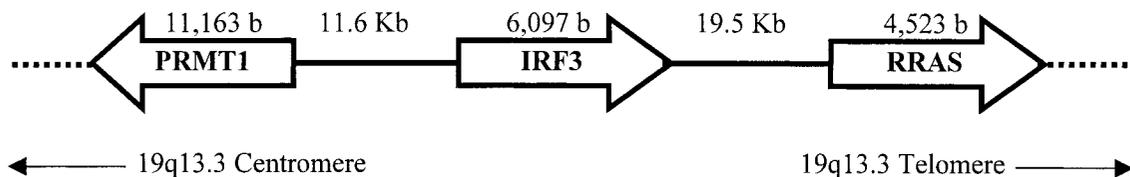
**Protein structural analysis.** The Baylor College of Medicine (BCM) search launcher was used to perform multiple alignment using Clustal X software (21), and hydrophobicity studies. Phyloge-

netic trees were constructed using the Phylip software package (22). Distance matrix analysis was performed using the "Neighbor-Joining/UPGMA" program, and parsimony analysis was done using the "Protpars" program (23). Signal peptide was predicted using the SignalP (24). Protein structure analysis was performed with the "SAPS" (structural analysis of protein sequence) program (25).

**Hormonal stimulation the BT-474 cell line.** The breast cancer cell line BT-474 was purchased from the American Type Culture Collection (Rockville, MD). Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/L), bovine insulin (10 mg/L), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were then aliquoted into 24-well tissue culture plates and cultured to 50% confluency. The culture media were changed to phenol red-free media containing 10% charcoal-stripped fetal bovine serum 24 h prior to stimulation. Hormonal stimulation was performed by the addition of various steroid hormones, dissolved in 100% ethanol, to the culture media at a final concentration of  $10^{-8}$  M. Cells stimulated with 100% ethanol were included as negative controls. The cells were cultured for 24 h before mRNA extraction.

**Human tissues.** Total RNA isolated from 26 different human tissues was purchased from Clontech (Palo Alto, CA). Noncancerous breast tissues were obtained from women undergoing reduction mammoplasties. Breast tumor tissues were obtained from leftover frozen specimens submitted for steroid receptor analysis. Tissues were immediately frozen in liquid nitrogen after surgical resection.

**RT-PCR analysis of PRMT1 expression.** Total RNA was extracted from cells and tissues using Trizol reagent (Gibco BRL) according to the manufacturer's specifications. RNA concentration was determined spectrophotometrically. Two micrograms of total RNA was reverse transcribed into first strand cDNA using the Superscript preamplification system (Gibco BRL). After aligning all known PRMT1 homologous genes, two primers (PF and PR) (Table 1; Fig. 1) were designed from areas with relatively low homology. The RT-PCR primers spanned at least 2 exons to avoid contamination by genomic DNA. Tissue cDNAs were amplified at various dilutions. PCR was performed in a 20  $\mu$ L reaction mixture containing 1  $\mu$ L of cDNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M dNTPs (deoxynucleoside triphosphates), 150 ng of primers and 2.5 units of AmpliTaq Gold DNA polymerase (Roche Molecular Systems, Branchburg, NJ) on a Perkin-Elmer 9600 thermal cycler. An initial incubation at 94°C for 10 min was followed by 43 cycles consisting of a 94°C denaturing step (30 s) and a 67°C annealing/extension step (1 min). A final extension step at 67°C was included for 10 min. PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining.



**FIG. 3.** Mapping of the PRMT1 gene. An approximate 60-kb region of contiguous genomic sequence around chromosome 19q13.3. Genes are represented by horizontal arrows denoting the direction of transcription. Figure is not drawn to scale. For full gene names see abbreviations footnote.

*Cloning and sequencing of PCR products.* To verify the identity of the PCR products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The inserts were sequenced from both directions, using vector-specific primers, with an automated DNA sequencer.

## RESULTS

### *Characterization of the PRMT1 Gene Sequence*

Computer analysis of the contiguous genomic sequence of interest around chromosome 19q13.3-q13.4 predicted a gene comprised of 12 exons. To verify the existence and identity of this gene, the putative exons were subjected to sequence homology search against the human expressed sequence tag database (dbEST), and 10 EST clones with >98% homology were identified. A similar search of the GenBank sequence database identified the cDNA from the human PRMT1/HRMT1L2 gene to be 100% homologous to the predicted exons. Thus, our genomic sequence was found to harbor the PRMT1 gene (please see also below).

### *Genomic Structure of the PRMT1 Gene*

By comparing the genomic sequences of interest with the cDNA sequences of the known three PRMT1 gene isoforms (variants; v) as described by Scott *et al.* (10), we were able to determine the genomic organization of the PRMT1 gene. Figure 1 demonstrates that the PRMT1 gene is made up of 12 (v3), 11 (v2) or 10 (v1) exons, separated by 11, 10, or 9 introns, respectively (all variants are included in our GenBank Sub-

mission No. AF222689). Interestingly, in variant 3, there is an in-frame stop codon in the middle of exon 3. A second downstream start codon resumes transcription. A schematic description of the variants is further shown in Fig. 2.

The gene spans 11,163 bp of genomic sequence on chromosome 19q13.3 and is processed to mRNA that is 1,433 (v3), 1316 (v2) or 1262 (v1) nucleotides in length, excluding the poly (A) tail. All splice junctions are in accordance with the conserved consensus sequence GT/AG (26), except for the end of intron 2, which contains CC. The lengths of the exons are 34, 54, 117, 102, 156, 64, 143, 88, 116, 151, 122, and 286 bp, respectively. In isoforms v1 and v2, the start codon (ATG) is present at position 29 of exon 1. In variant 3 there is a stop codon at position 67–69 of exon 3, while a second start codon is present downstream on the same exon at position 100. The flanking sequences of the two start codons match closely with the Kozak sequence (GCCA/GCCATGG) (27). The stop codon is located at position 82–84 of exon 12, followed by a 3' untranslated region, and a typical polyadenylation signal (AATAAA) is located 12 bp upstream of the poly-A tail. More details of the sequence have been described in our GenBank submission, Accession No. AF222689.

### *Mapping and Chromosomal Localization of the PRMT1 Gene*

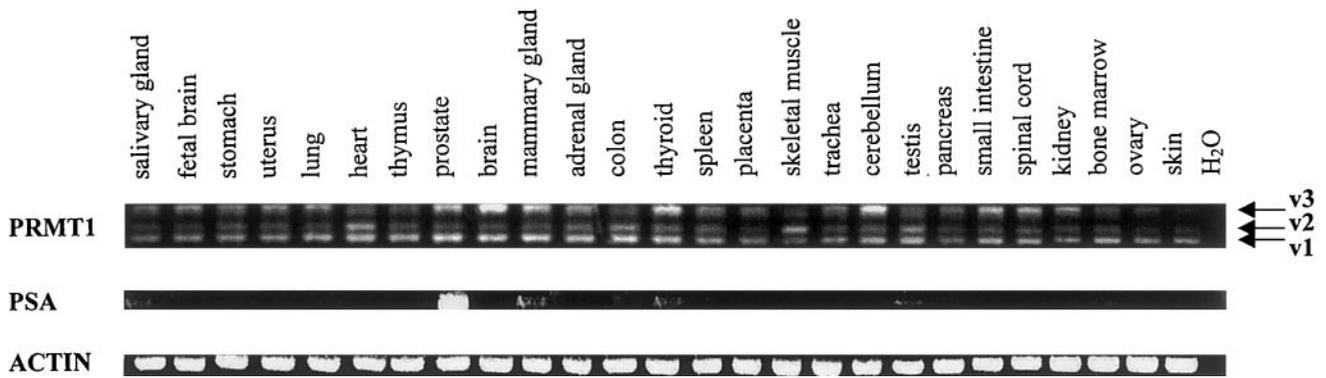
The knowledge of extensive genomic sequence on chromosome 19q13.3 enabled us to precisely localize

**TABLE 1**  
Primers Used for Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Analysis

Gene	Primer name	Sequence <sup>1</sup>	Product sizes (base pairs) <sup>2</sup>
PRMT1	PF	GAGGCCGCGAACTGCATCAT	502, 385, 331
	PR	TGGCTTTGACGATCTTCACC	
PSA	PSA-S	TGCGCAAAGTTCACCCTCA	754
	PSA-AS	CCCTCTCCTTACTTCATCC	
Actin	ACTIN-S	ACAATGAGCTGCGTGTGGCT	372
	ACTIN-AS	TCTCCTTAATGTACGCACGA	

<sup>1</sup> All nucleotide sequence are given in the 5' → 3' orientation.

<sup>2</sup> Primers PF and PR generate three bands, representing the three PRMT1 isoforms (variants).



**FIG. 4.** Tissue expression of the PRMT1 gene as determined by RT-PCR. The upper band (502 bp) is the variant 3 (v3), the middle band (385 bp) is the variant 2 (v2), and the lower band (331 bp) is the variant 1 (v1). For splice variant discussion see text. The primers used were PF and PR, as described in Table 1. For further discussion see text.

PRMT1 as well as two additional known genes, namely RRAS and IRF3, and to determine the direction of transcription, as shown in Fig. 3. RRAS is the most telomeric, followed by IRF3 and PRMT1, respectively. PRMT1 is transcribed in the opposite direction of RRAS and IRF3. The distance between the RRAS and IRF3 genes was calculated to be 19.5 kb. PRMT1 is located 11.6 kb centromeric to IRF3.

#### Tissue Expression of PRMT1 Splice Variants

Screening of cDNA from a panel of 26 different tissues using gene-specific primers for exons 1 and 6 (PF/PR; Table 1) (exon numbers refers to our GenBank Submission No. AF222689) revealed that this gene is expressed in many tissues. Three bands were observed, representing PCR products of 502, 385, and 331 bp, in the tissues examined (Fig. 4). The five tissues displaying the highest level of expression (cerebellum, mammary gland, prostate, brain, and thyroid) were cloned and sequenced. RT-PCR for actin and PSA (which exhibits tissue-restricted expression) were used as positive controls. The PSA gene was found to be highly expressed in the prostate, and to a lower extent in the mammary, thyroid, and salivary glands, as expected (28, 29).

The tissue expression profile of the PRMT1 variants is summarized in Fig. 4 and Table 2. Variant 1 is primarily expressed in the prostate, brain, heart, adrenal gland, mammary gland, colon, thyroid, and testicular tissues, and to a lower extent in many other tissues. The expression of the PRMT1-v2 is more predominant in heart, skeletal muscle, colon, and testicular tissues. PRMT-v3 is mainly expressed in the brain, cerebellum, thyroid, prostate, and mammary glands.

#### Differential Expression PRMT1 Splicing Variants in the BT-474 Breast Cancer Cell Lines

We used the steroid hormone receptor-positive breast carcinoma cell line BT-474 as a model system to

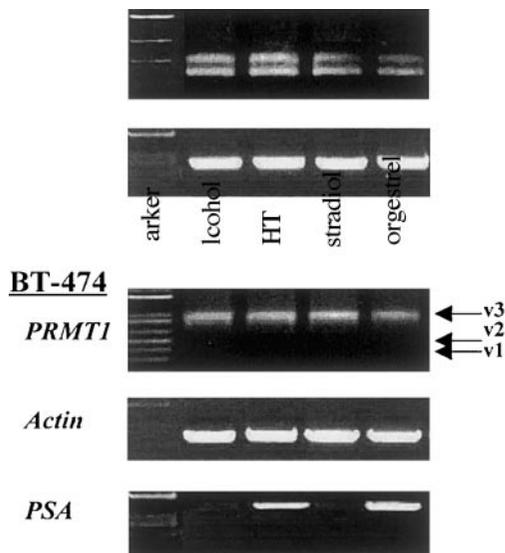
evaluate whether PRMT1 expression is under steroid hormone regulation. This gene does not seem to be regulated significantly by estrogen, androgens, or progestins (Fig. 5). However, only variant 3 was detected, in contrast to the data of Fig. 4, which shows expression of all three variants in normal mammary tissue.

#### PRMT1 Expression in Breast Cancer

Expression of the PRMT1 gene was examined in 16 breast tumors and three noncancerous breast tissue

**TABLE 2**  
Expression Profile of PRMT1 in Tissues and the BT-474 Breast Cancer Cell Line, by RT-PCR Analysis

Tissue/Cell line	PRMT1-v1	PRMT1-v2	PRMT1-v3
Salivary gland	+	+	+
Fetal brain	+	+	++
Stomach	+	+	++
Uterus	+	+	++
Lung	++	+	++
Heart	+++	++	+
Thymus	++	+	+
Prostate	+++	+	+++
Brain	+++	+	+++
Mammary gland	+++	+	+++
Adrenal gland	+++	+	++
Colon	+++	++	+
Thyroid	+++	+	+++
Spleen	+	+	++
Placenta	+	+	+
Skeletal muscle	+	++	+
Trachea	+	+	++
Cerebellum	+++	+	+++
Testis	++	++	+
Pancreas	+	+	+
Small intestine	++	+	++
Spinal cord	++	+	++
Kidney	++	+	++
Bone marrow	++	+	+
Ovary	+	+	+
Skin	+	+	+/-
BT-474	-	-	++



**FIG. 5.** PRMT1 expression in BT-474 breast carcinoma cell line stimulated with various steroids, as shown. Steroids were at  $10^{-8}$  M final concentration. Actin (not regulated by steroid hormones) and PSA (up-regulated by androgens and progestins), are control genes. This cell line produces only variant 3 and the expression is not significantly altered by steroid hormones. Norgestrel is a synthetic progestin. (DHT, dihydrotestosterone).

specimens using RT-PCR. All three PRMT1 splicing variants were expressed in the noncancerous tissues. The gene is not expressed at all in two tumors (Fig. 6, lanes 2 and 9). Three of the tumors expressed all the PRMT1 variants (lanes 3, 6, and 16). Expression of PRMT1-v2 was not observed in 12 tumors while PRMT1-v3 was not expressed in 13 tumors. PRMT1-v1 was expressed at variable levels in the majority of the tumors (14/16) (Fig. 6). We thus conclude that, in comparison to normal tissues, tumors have the following expression patterns: (a) very frequent downregulation of variants 2 and 3 and (b) less frequent downregulation of variant 1.

#### Analysis of the PRMT1 Proteins

The predicted protein-coding regions of the PRMT1 gene are comprised of 1,032 bp (v1), 1,086 bp (v2), or 1,044 bp (v3), and encode 3 deduced polypeptides with

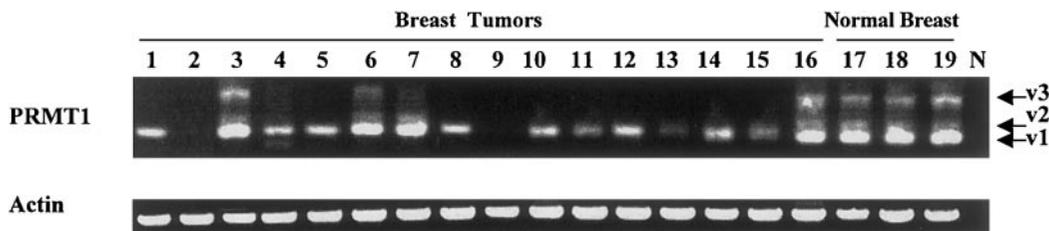
a predicted molecular weight of 39.6, 41.5, and 39.9 kDa, excluding any posttranslational modifications, with theoretical isoelectric points of 5.4, 5.2, and 5.4, respectively. PRMT1 contains 343 (v1), 361 (v2) or 347 (v3) amino acids. In Fig. 7, alignment of human PRMT proteins (for PRMT1 we used the longest variant, v2) and other non-human homologs is presented. Various areas of extensive protein sequence homology are evident, between the human PRMT members as well as the three rat orthologs and the yeast enzyme.

Furthermore, human PRMT1 protein has 96% amino acid sequence identity with the rat PRMT1, 45% identity with yeast PRMT1 and 34, 33, 29, and 28% identity with human PRMT3, rat PRMT3, human PRMT2, and rat PRMT2, respectively. To predict the phylogenetic relatedness of the human PRMT1 gene with other arginine methyltransferases, the amino acid sequences of these genes were aligned using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA method (Fig. 8). Phylogenetic analysis generated three clusters and grouped human PRMT1 with rat PRMT1, in accordance with previous studies (10, 13) and indicating that this group of genes probably arose from a common ancestral gene by duplication.

Examination of the hydrophobicity profile of the PRMT1 protein variants revealed three (v2 and v3) or two (v1) regions with long stretches of hydrophobic residues (Fig. 9). The first of these occurs at the N-terminus, suggesting the presence of a putative signal peptide but software analysis (24) predicts these proteins as nonsecretory.

#### DISCUSSION

Protein arginine methyltransferases (PRMTs) play a role in the control of mRNA processing and maturation by modulating the activity of RNA-binding proteins. To date, three human PRMTs have been cloned. This study reports, for the first time, the structural characterization and genomic organization of the full-length human PRMT1 gene. Furthermore, the genomic region comprising the PRMT1, RRAS, and IRF3 genes was elucidated. Three PRMT1 splicing variants (v1, v2, v3) were identified, in agreement with previously reported



**FIG. 6.** Expression of splicing variants of the PRMT1 gene in primary breast tumors (1-16) and normal breast tissues (17-19). N, negative control. Actin is a positive control. For further discussion see text.

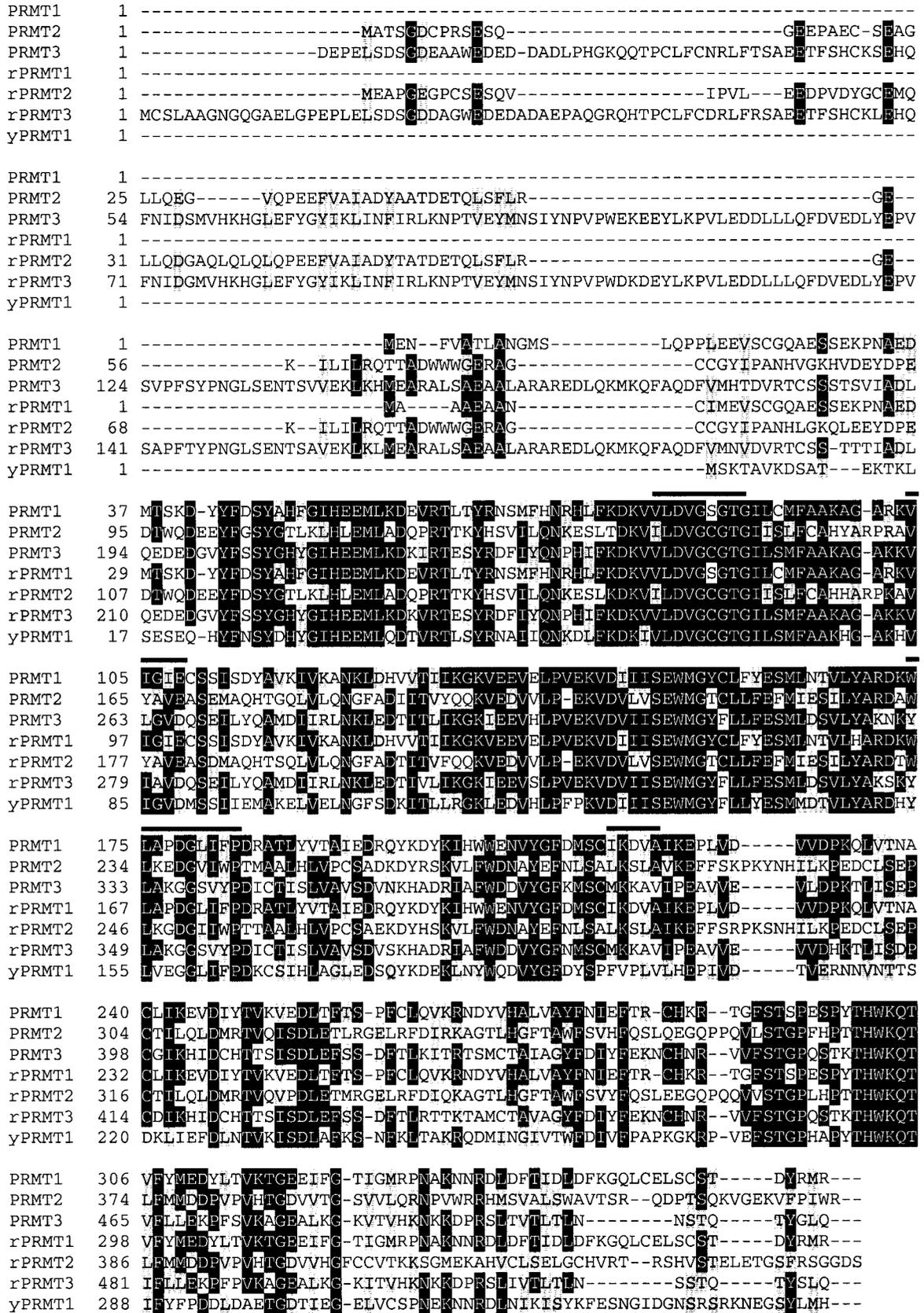
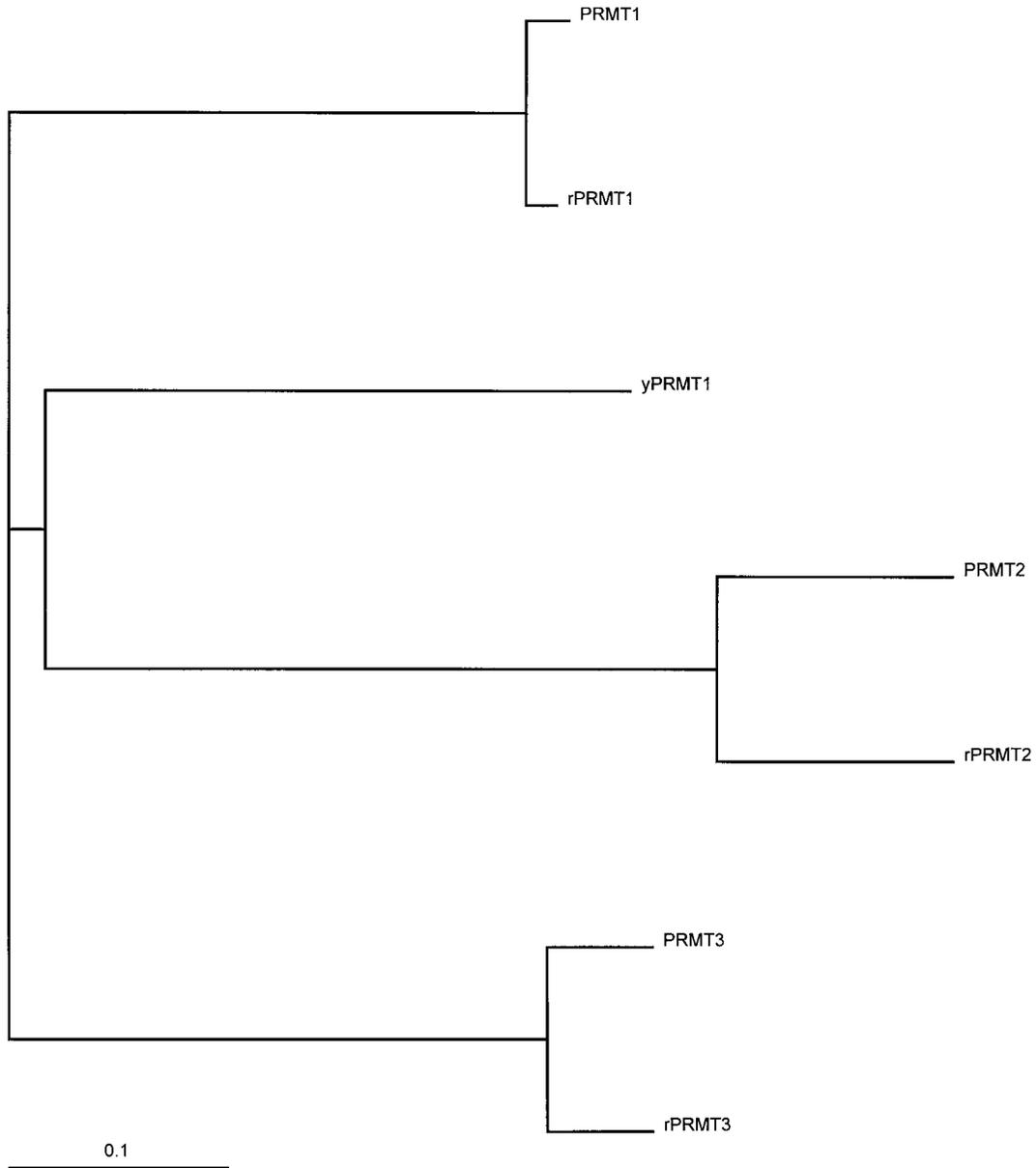


FIG. 7. Multiple alignment of the deduced amino acid sequence of PRMT1-1,2 (the longer of the three variants) with other arginine methyltransferases. r, rodent; y, yeast. Dashes represent gaps to bring the sequences to better alignment. Grey shading indicates similar amino acids, while black indicates identical amino acid residues. The predicted methyltransferase regions I, II, and III are shown by solid underline.

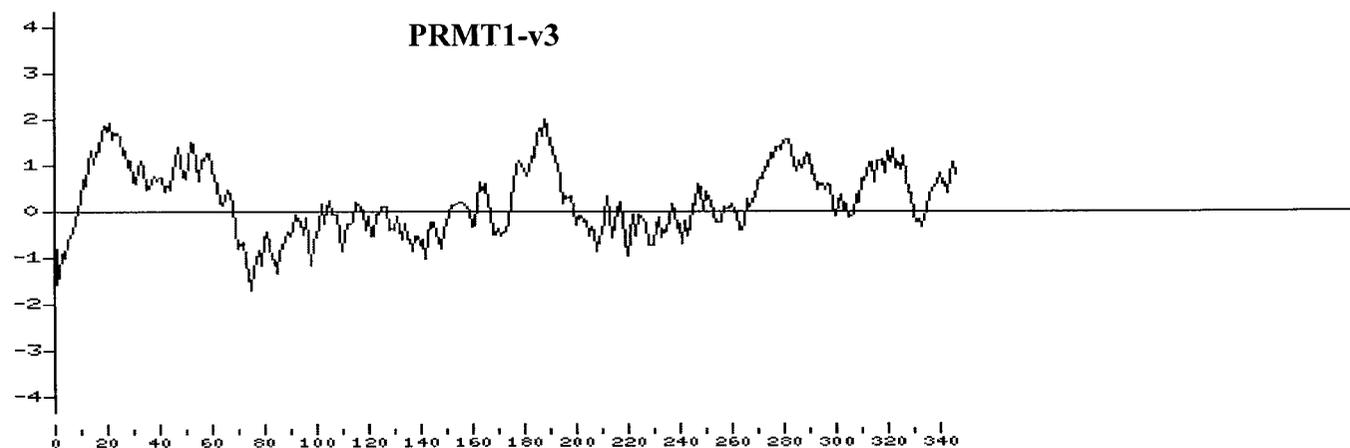
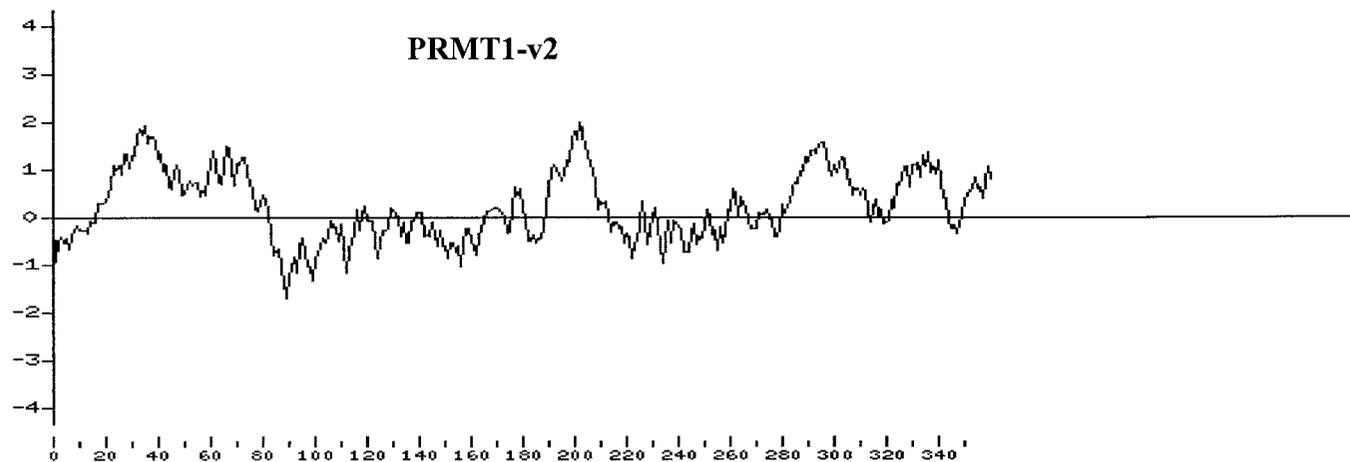
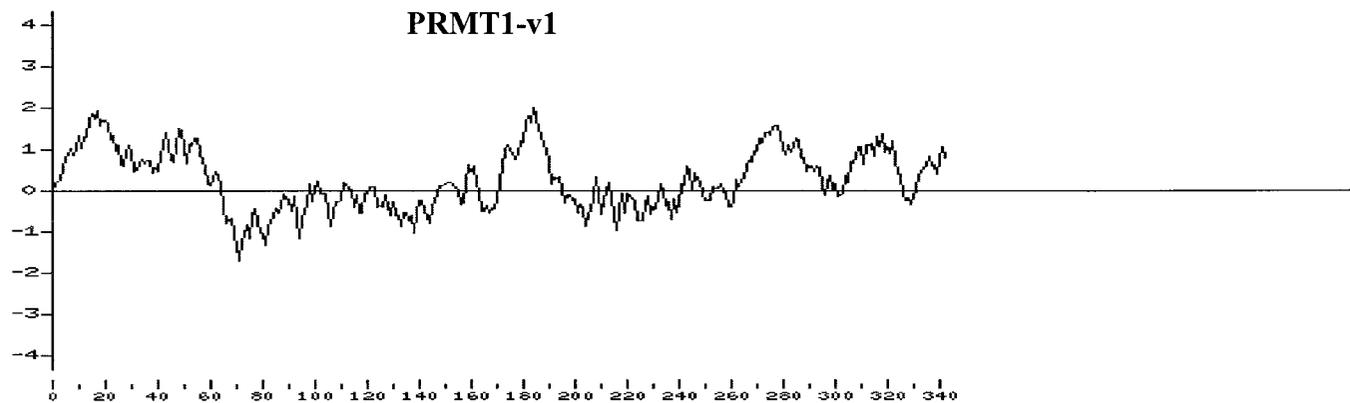


**FIG. 8.** Dendrogram of the predicted phylogenetic tree for the known arginine methyltransferase proteins. The neighbor-joining/UPGMA method was used to align PRMT1-v2 with other members of the arginine methyltransferase gene family. Three clusters were generated, one of which is comprised of the human and rat PRMT1 protein.

data (10). The PRMT1 gene is made up of 12 (v3), 11 (v2) or 10 (v1) exons, separated by 11, 10, or 9 introns, respectively. We examined the differential expression of the three PRMT1 isoforms in normal tissues, the BT-474 breast cancer cell line, and 16 breast tumors. PRMT1 was abundantly expressed in all samples, not surprisingly, given that methylation of RNA-binding proteins occurs in all cells. The relative predominance of the alternatively spliced forms of PRMT1, however, differed between tissues, as well as between normal breast tissue and breast cancers and BT-474 cells. PRMT1-v1 and PRMT1-v2 were significantly down

regulated in most of the breast tumors. This gene is not regulated by estrogen, androgens, or progestins in BT-474 cells.

The biological function of PRMTs is to methylate RNA-binding proteins, particularly hnRNPs, which, in turn, mediate mRNA maturation, splicing, localization, and metabolism. The overall physiological role of PRMT1, however, is not clear. It has been reported that PRMT1 may be involved in certain signal transduction events. PRMT1 is able to interact with the intracytoplasmic domain of the Type 1 interferon (IFN) receptor, upregulating the antiproliferative effect of



PRMT1-v1	1	-----MEVSCGQAESSEKPN AEDMTSKDYYFDSYAHFGIHEEMLKD .....	343
PRMT1-v2	1	MENFVATLANGMSLQPPLEEVSCGQAESSEKPN AEDMTSKDYYFDSYAHFGIHEEMLKD.....	361
PRMT1-v3	1	-----MVGVAEVSCGQAESSEKPN AEDMTSKDYYFDSYAHFGIHEEMLKD.....	347

**FIG. 9.** Hydrophobicity plots of PRMT1 splicing variants. The hydrophobic N-terminus in PRMT1-v2 and PRMT1-v3 may harbor a signal and/or activation peptide. For details see text. Alignment of the deduced N-terminal amino acid sequences of PRMT1-v1, PRMT1-v2, and PRMT1-v3 is presented at the bottom of the figure. Grey shading indicates similar amino acids, while black indicates identical amino acid residues. Sequence not shown ( . . . . . ) is identical between the three variants.

IFNs (10). Another study reported that PRMT1 associates with the product of the immediate-early response gene TIS21, which acts as a negative regulator of cell proliferation in response to external signals. TIS21 modulates PRMT1 enzymatic activity in terms of substrate specificity, and upregulates PRMT1 catalytic efficiency. The implication of PRMT1 as an external signal modulator is logical, as gene expression in response to external signals may be under transcriptional or posttranscriptional control. The latter may involve mRNA splicing, localization, or stability, all controlled by PRMT1. PRMT1 may also be of importance prior to mRNA processing. The recently identified coactivator-associated arginine methyltransferase 1 (CARM1), which is homologous to protein arginine methyltransferases, is incorporated into transcription initiation complexes, and methylates proteins in this complex in order to activate transcription (30). The above studies collectively suggest a role for PRMTs at the transcription and/or translation level during the cellular response to external signals.

No previous studies regarding PRMT1 expression and cancer have been reported. However, PRMT substrates, including hnRNPs (31), fibrillarlin (32), and ribosomal protein S2 (33) are overexpressed in some tumors. Furthermore, a dramatic increase in cellular arginine methylation has been reported in certain types of cancer (34–36). Additionally, the N-terminal region of the rat PRMT1 and PRMT2 proteins contains a Src homology 3 (SH3) domain (10, 37).

The development and progression of cancer may alter the expression of certain genes. Tumor-specific modulation of gene expression may occur at the post-transcriptional level, and include such processes as alternative splicing, yet the functional significance of PRMT1 splicing in breast cancer is unclear. Amino acid sequence comparison of the splicing variants indicates that they are all enzymatically active methyltransferases. The isoforms may, however, interact differently with members of signal transduction pathways or components of the transcriptional activation machinery. Here, we report the absence of PRMT1-v1 and 2 in the breast cancer cell line BT-474, and downregulation of variants 2 and 3 in breast tumors. It will be interesting to determine if the alterations are valuable for breast cancer prognosis or prediction of therapeutic response.

## REFERENCES

1. Pinol-Roma, S., and Dreyfuss, G. (1991) Transcription-dependent and transcription-independent nuclear transport of hnRNP proteins. *Science* **253**, 312–314.
2. Soulard, M., Veronique D. V., Siomi, M. C., Pinol-Roma, S., Codogno, P., Bauvy, C., Bellini M., Lacroix, J. C., Monod, G., Dreyfuss, G., and Larsen C. J. (1993) hnRNP G: Sequence and characterization of a glycosylated RNA-binding protein. *Nucleic Acids Res.* **21**, 4210–4217.
3. Cobianchi, F., Calvio, C., Stoppin, M., Buvoli, M., and Riva, S. (1990) Phosphorylation of human hnRNP protein A1 abrogates in vitro strand annealing activity. *Nucleic Acid Res.* **21**, 949–955.
4. Wilk, H. E., Wer, R. H., Friedrich, D., Kiltz, H. H., and Schafer, K. P. (1985) The core proteins of 35s hnRNP complexes: Characterization of nine different species. *Eur. J. Biochem.* **46**, 71–81.
5. Boffa, L. C., Karn, J., Vidali, G., and Allfrey, V. G. (1977) Distribution of  $N^G, N^G$ -dimethylarginine in nuclear protein fractions. *Biochem. Biophys. Res. Commun.* **74**, 969–976.
6. Lin, W. J., Gary, J. D., Yang, M. C., Clarke, S., and Herschman, H. R. (1996) The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine *N*-methyltransferase. *J. Biol. Chem.* **271**, 15034–15044.
7. Ghosh, S. K., Paik, W. K., and Kim, S. (1998) Purification and molecular identification of two protein methylases I from calf brain. Myelin basic protein and histone-specific enzyme. *J. Biol. Chem.* **271**, 12585–12594.
8. Henry, M. F., and Silver, P. A. (1996) A novel methyltransferase (Hmt1p) modifies poly(A)+-RNA-binding proteins. *Mol. Cell. Biol.* **16**, 3668–3678.
9. Niewmierzyccka, A., and Clarke, S. (1999) *S*-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J. Biol. Chem.* **274**, 814–824.
10. Scott, H. S., Antonarakis, S. E., Lalioti, M. D., Rossier, C., Siver, P. A., and Henry M. F. (1998) Identification and characterization of two putative human arginine methyltransferases (HRMT1L1 and HRMT1L2). *Genomics* **48**, 330–340.
11. Abramovich, C., Yakobson, B., Chebath, H., and Revel, M. (1997) A protein-arginine methyltransferase binds to the intracytoplasmic domain of the IFNAR1 chain in the type I interferon receptor. *EMBO J.* **16**, 260–266.
12. Katsanis, N., Yaspo, M. L., and Fisher, E. M. (1998) Identification and mapping of a novel human gene, HRMT1L1, homologous to the rat protein arginine *N*-methyltransferase 1 (PRMT1) gene. *Mamm. Genome* **8**, 526–529.
13. Tang, J., Gary, J. D., Clarke, S., and Herschman, H. R. (1998). PRMT 3, a type I protein arginine *N*-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. *J. Biol. Chem.* **273**, 16935–16945.
14. Ashworth, L. K., Batzer, M. A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J. A., Gordon, L. A., Lamerdin, J. E., Lennon, G., et al. (1995) An integrated metric physical map of human chromosome 19. *Nat. Genet.* **11**, 422–7.
15. Bonet, M., Steel, M., Warnow, T., and Yooseph, S. (1998) Better methods for solving parsimony and compatibility. *J. Comput. Biol.* **5**, 391–407.
16. Smith, D. K., and Xue, H. (1998) A major component approach to presenting consensus sequences. *Bioinformatics* **14**, 151–156.
17. Murakami, K., and Takagi, T. (1998) Gene recognition by combination of several gene-finding programs. *Bioinformatics* **14**, 665–675.
18. Yousef, G. M., Scorilas, A., and Diamandis, E. P. (2000) Genomic organization, mapping, tissue expression, and hormonal regulation of trypsin-like serine protease (TLSP PRSS20), a new member of the human kallikrein gene family. *Genomics* **63**, 88–96.
19. Yousef, G., Scorilas, A., Magklara, A., Soosaipillai, A., and Diamandis, E. P. (2000) The PRSS6 gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family-Genomic characterization, mapping, tissue expression and hormonal regulation. *Gene* **258**, 24–31.
20. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997) Gapped BLAST and PSI-

- BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
21. Aiyar, A. (2000) The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. *Methods Mol. Biol.* **132**, 221–241.
  22. Retief, J. D. (2000) Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.* **132**, 243–258.
  23. Takezaki, N. (1998) Tie trees generated by distance methods of phylogenetic reconstruction. *Mol. Biol. Evol.* **15**, 727–737.
  24. Nielsen, H., Brunak, S., and von Heijne, G. (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* **12**, 3–9.
  25. Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B. E., and Karlin, S. (1992) Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. USA* **89**, 2002–2006.
  26. Iida, Y. (1990) Quantification analysis of 5'-splice signal sequences in mRNA precursors. Mutations in 5'-splice signal sequence of human beta-globin gene and beta-thalassemia. *J. Theor. Biol.* **145**, 523–533.
  27. Kozak, M. (1991) An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell. Biol.* **115**, 887–892.
  28. Diamandis, E. P., Yu, H., and Sutherland, D. J. (1994) Detection of prostate-specific antigen immunoreactivity in breast tumours. *Br. Cancer Res. Treat.* **32**, 301–310.
  29. Scorilas, A., Yu, H., Soosaipillai, A. R., Gregorakis, A. K., and Diamandis, E. P. (2000) Comparison of the percent free prostate-specific antigen levels in healthy men and in men with recurrent prostate cancer after radical prostatectomy *Clin. Chim. Acta* **292**, 127–138.
  30. Chen, D., Ma, H., Hong, H., Koh, S. S., Huang, S.-M., Schurter, B. T., Aswad, D. W., and Stallcup, M. R. (1999). Regulation of transcription by a protein methyltransferase. *Science* **284**, 2174–2177.
  31. Sueoka, E., Goto, Y., Sueoka, N., Kai Y., Kozu, T., and Fujiki (1999) Heterogeneous nuclear ribonucleoprotein B1 as a new marker of early detection for human lung cancers. *Cancer Res.* **59**, 1404–1407.
  32. Ochs, R. L., Stein, T. W., Jr., and Tan, E. M. (1994) Coiled bodies in the nucleolus of breast cancer cells. *J. Cell Sci.* **107** (Pt. 2), 385–399.
  33. Chiao, P. J., Shin, D. M., Sacks, P. G., Hong, W. K., and Tainsky, M. A. (1992) Elevated expression of the ribosomal protein S2 gene in human tumors. *Mol. Carcinog.* **5**, 219–231.
  34. Shin, D. M., Chiao, P. J., Sacks, P. G., Shin, H. J., Hong, W. K., Hittelman, W. N., and Tainsky, M. A. (1993) Activation of ribosomal protein S2 gene expression in a hamster model of chemically induced oral carcinogenesis. *Carcinogenesis* **14**, 163–166.
  35. Li, C., Ai A. S., Lin, C. H., Hsieh, M., Li Y. C., and Li, S. Y. (1998) Protein N-arginine methylation in adenosine dialdehyde-treated lymphoblastoid cells. *Arch. Biochem. Biophys.* **351**, 53–59.
  36. Gu, H., Park, S. H., Park, G. H., Lin, I. K., Lee, H. W., Paik, W. K., and Kim, S. (1999) Identification of highly methylated arginine residues in an endogenous 20-kDa polypeptide in cancer cells. *Life Sci.* **65**, 737–745.
  37. Bonfini, L., Migliaccio, E., Pelicci, G., Lanfranccone, L., and Pelicci, P. G. (1996) Not all Shc's roads lead to Ras. *Trends Biochem. Sci.* **21**, 257–261.